



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Multivariate random-effects meta-analysis for sparse data using `smvmeta`

Christopher James Rose
Center for Epidemic Interventions Research
and
Cluster for Reviews and Health Technology Assessments
Division for Health Services
Norwegian Institute of Public Health
Oslo, Norway
cjro@fhi.no

Abstract. Multivariate meta-analysis is used to synthesize estimates of multiple quantities (“effect sizes”), such as risk factors or treatment effects, accounting for correlation and typically also heterogeneity. In the most general case, estimation can be intractable if data are sparse (for example, many risk factors but few studies) because the number of model parameters that must be estimated scales quadratically with the number of effect sizes. This article presents a new command, `smvmeta`, that makes estimation tractable by modeling correlation and heterogeneity in a low-dimensional space via random projection. This reduces the number of model parameters to be linear in the number of effect sizes. `smvmeta` is demonstrated in a meta-analysis of 23 risk factors for pain after total knee arthroplasty. Validation experiments show that, compared with meta-regression (a reasonable alternative model that could be used when data are sparse), `smvmeta` can provide substantially more precise estimates (that is, narrower confidence intervals) at little cost in bias.

Keywords: `st0749`, `smvmeta`, multivariate meta-analysis, sparse data, dimensionality reduction, penalized maximum likelihood, risk factors

1 Introduction

Meta-analysis is used to synthesize multiple estimates, most often of one quantity. This is called univariate or pairwise meta-analysis (after pairs of treatments that are compared to estimate treatment effect). Multivariate meta-analysis is used to synthesize one or more estimates of each of *multiple* quantities. This article is about multivariate meta-analysis when data are sparse, which easily occurs when the number of quantities of interest is even moderately large. For consistency with the nomenclature used in `meta` (see [META] `meta`) and much of the meta-analysis literature, the remainder of this article uses “effect size” to refer to the value or an estimate of a quantity of interest, as should be clear from context.

The canonical application of multivariate meta-analysis is arguably diagnostic test accuracy, in which sensitivity and specificity are of interest (Riley, Thompson, and Abrams 2008; Riley 2009). Stata 16 introduced an excellent suite of built-in meta-

analysis commands. Support for multivariate meta-analysis was added in Stata 17 (see [META] **meta** and [META] **meta mvregress**). Stata's meta-analysis commands were preceded by third-party add-on commands, notably **mvmeta** (White 2009, 2011), which does multivariate meta-analysis.

Naïvely, multivariate meta-analysis can be posed as multiple independent univariate meta-analyses or meta-regressions. However, these approaches fail to model potentially important correlations between effect sizes or heterogeneity (differences in the included studies' estimation targets). Such approaches are expected to yield excessively biased and imprecise estimates (Riley 2009). Multivariate models are recommended in meta-analyses of diagnostic test accuracy because sensitivity and specificity are interdependent, and it is therefore beneficial to account for the correlation between them.

Higher-dimensional problems, such as questions on risk factors, can be challenging because data are often sparse (informally, the number of effect sizes published by studies is less than the number of model parameters that must be estimated). A more detailed description of the sparsity problem is given in section 4. Existing multivariate meta-analysis methods can be applied when data are sparse, but results may be untrustworthy (White 2009) and performance may be poor (White 2011). Sparse multivariate meta-analysis has been addressed previously, for example, by Lin and Chu (2018).

This article describes a new command, **smvmeta**, that implements a multivariate random-effects meta-analysis model for sparse data. It is a frequentist version of the Bayesian model described by Rose et al. (2021), which in turn was motivated by the model presented by Lin and Chu (2018). Section 2 presents the syntax of **smvmeta**. Section 3 shows how to use the command, using a meta-analysis of risk factors for pain after total knee arthroplasty as an example. Section 4 provides mathematical details on the model and describes its implementation in **smvmeta**. Section 5 presents simulation-based validation experiments. The article closes with a discussion.

Figure 1 illustrates how **smvmeta** makes multivariate meta-analysis tractable when data are sparse. In the most general case in which correlations and heterogeneity are unknown for all effect sizes, the number of model parameters that must be estimated in existing models scales quadratically with the number of effect sizes because correlation and heterogeneity must be estimated for each unique pairing of effect sizes. **smvmeta** models correlation and heterogeneity in a low-dimensional space, such that the number of effect sizes to be estimated is the dominant term in the model's complexity and the number of model parameters scales linearly instead of quadratically. This means that fewer studies and study estimates are needed to support estimation. This comes at a price: while the random-effects variance and covariance components are *modeled* (that is, accounted for), they are not directly *estimated*. This price may be worth paying because simulations suggest that **smvmeta**'s estimates are on average substantially more precise compared with those obtained using meta-regression (section 5).

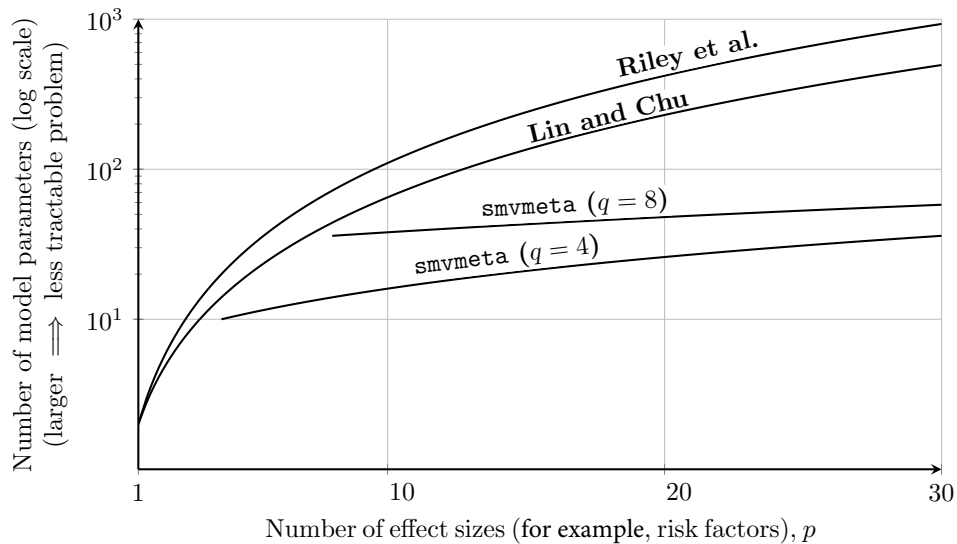


Figure 1. Multivariate random-effects meta-analysis can be intractable when there are many effect sizes (for example risk factors) but few studies. In the most general case, the number of model parameters that must be estimated scales quadratically with the number of effect sizes (the models of Riley, Thompson, and Abrams [2008] and Lin and Chu [2018]). `smvmeta` makes estimation tractable by modeling correlation and heterogeneity in a space of low dimension, q , such that the number of model parameters scales linearly with the number of effect sizes.

2 The `smvmeta` command

2.1 Syntax

2.1.1 Specify generic effect sizes, standard errors, and a factor variable identifying them

Specify the variables containing generic effect sizes, their standard errors, and a factor variable that identifies the distinct effect sizes to be estimated (for example, risk factors) using `smvmeta set`:

```
smvmeta set esvar sevar fvar
```

where data are arranged with one estimate per row. `esvar` and `sevar` correspond to variables containing point estimates of effect sizes and their standard errors, respectively. `fvar` specifies a factor variable identifying the distinct effect sizes to be estimated. The factor variable operator is not allowed.

esvar (and therefore *sevar*) must be specified in the metric closest to normality, such as log odds-ratios instead of odds ratios.

It is not necessary to specify which study provides a given estimate, and there is no `studylabel()` option as with Stata's `meta set`. For the moment, assume that each study reports at most one estimate for a given effect size.

Data must be `smvmeta set`. You are free to subsequently change the data without needing to `smvmeta set` it again, provided you do not change the names of *esvar*, *sevar*, or *fvar*. If you do change those names, you must `smvmeta set` the data again.

2.1.2 Sparse multivariate random-effects meta-analysis as declared with `smvmeta set`

Meta-analysis is performed using the following syntax:

```
smvmeta estimate [if] [in] [, est_opts rep_opts]
```

The options *est_opts* and *rep_opts* are explained in section 2.2. `by` and therefore `bysort` are allowed for `smvmeta estimate`.

2.1.3 Make a forest plot showing estimates obtained using `smvmeta estimate`

Stored estimation results can be displayed as a forest plot using the following syntax:

```
smvmeta forestplot [columns] [, rep_opts forest_opts]
```

Options *rep_opts* are not “remembered” from previous `smvmeta` commands (see section 2.2). The optional *columns* may be one or more of `_fv`, `_plot`, `_es`, `_ci`, `_lb`, `_ub`, `_esci`, `_p`, `_k`, `_Pscore`, and `_I2`:

| <i>columns</i> | Description |
|----------------------|---|
| <code>_fv</code> | <i>fvar</i> label |
| <code>_plot</code> | forest graph |
| <code>_es</code> | effect size |
| <code>_ci</code> | confidence interval (CI) |
| <code>_lb</code> | lower bound on CI |
| <code>_ub</code> | upper bound on CI |
| <code>_esci</code> | effect size and CI |
| <code>_p</code> | <i>p</i> -value; see sections 2.2 and 4.3.3 |
| <code>_k</code> | number of observations (estimates) included |
| <code>_Pscore</code> | <i>P</i> -score superiority statistic; see sections 2.2 and 4.4 |
| <code>_I2</code> | <i>I</i> ² heterogeneity statistic |

The default columns are `_fv`, `_plot`, `_es`, `_ci`, `_p`, `_k`, `_Pscore`, and `_I2`.

2.1.4 Redisplay stored estimation results

Stored estimation results can be redisplayed using the following syntax:

```
smvmeta[, rep_opts]
```

Options *rep_opts* are not “remembered” from previous `smvmeta` commands (see section 2.2).

2.1.5 Display coefficient legend after estimation

The coefficient legend (how to specify coefficients in an expression) can be displayed after estimation using the following syntax:

```
smvmeta, coeflegend
```

2.2 Options

2.2.1 est_opts specifies how estimates are computed and stored

`dimension(#)` sets the number of dimensions in which to model correlation and heterogeneity. This corresponds to the symbol q in section 4, which explains the role of this option in detail. An error will result unless $2 \leq q < p$, where p is the number of distinct effect sizes to be estimated, as identified by variable *fvar*, and is present in the sample selected by `if`, `in`, `by`, etc. By default, `smvmeta` will search for a suitable value of q . This option is provided because a priori domain knowledge of the underlying dimensionality of the problem may facilitate better modeling.

`[no]log` displays or suppresses the penalized log likelihood and other maximization information at the start of each iteration. Logging is enabled by default.

2.2.2 rep_opts specifies how stored estimates are reported

`eform` is a synonym for `transform(exp)`.

`transform(transf_name)` reports transformed estimates. By default, results are displayed in the metric in which the problem was posed. *transf_name* affects how results are displayed, not how they are estimated and stored. *transf_name* is `corr`, `efficacy`, `exp`, `invlogit`, or `tanh`:

`corr` is a synonym for `tanh`.

`efficacy` transforms the effect sizes and CIs using the $1 - \exp()$ function (or more precisely, the `-expm1()` function). This transformation is used, for example, when the effect sizes are log risk-ratios so that the transformed effect sizes can be interpreted as treatment efficacies, $1 -$ risk ratios.

exp exponentiates effect sizes and CIs. This transformation is used, for example, when the effect sizes are log risk-ratios, log odds-ratios, and log hazard-ratios so that the transformed effect sizes can be interpreted as risk ratios, odds ratios, and hazard ratios.

invlogit transforms effect sizes and CIs with the inverse-logit function, `invlogit()`. This transformation is used, for example, when the effect sizes are logit of proportions so that the transformed effect sizes can be interpreted as proportions.

tanh applies the hyperbolic tangent transformation, `tanh()`, to the effect sizes and CIs. This transformation is used, for example, when the effect sizes are Fisher's z -values (compare Z -scores) so that the transformed effect sizes can be interpreted as correlations.

superior(*supspec*) computes P -scores using the definition of superiority specified by *supspec*. P -scores are explained in section 4.4. *supspec* is **+inf** (the default), **-inf**, **big**, or **small**:

+inf specifies that, of two effect sizes, the one closest to $+\infty$ is superior. **-inf** specifies the opposite.

big specifies that, of two effect sizes, the one with the largest magnitude is superior. **small** specifies the opposite.

P -scores are computed in the metric of *esvar*; this is relevant to the **efficacy** transform, which reverses estimates' signs.

sort(*column*[, **ascending**|**descending**]) sorts transformed results by *column*. By default, results are sorted in ascending order. *column* is one of **_fv**, **_es**, **_p**, **_k**, **_Pscore**, or **_I2**:

_fv sorts by the levels of *fvvar* and is the default.

_es sorts by estimated effect size.

_p sorts by p -value.

_k sorts by number of observations (estimates) included.

_Pscore sorts by P -score.

_I2 sorts by I^2 .

level(#) specifies the confidence level as a percentage.

smvmeta does not “remember” *rep_opts* from previous **smvmeta** commands, so it is essential to specify how results should be reported when displaying stored estimation results (for example, when using **smvmeta** to redisplay stored results or using **smvmeta forestplot** to produce figures).

2.2.3 forest_opts facilitate some control over forest plots

`if(exp)` restricts the results included in the forest plot to those identified by the expression. This option uses the stored estimates; it does not result in reestimation. The same variable names may be used as for the `sort()` option.

`cformat(%fmt)` and `pformat(%fmt)` specify the formats of coefficients and their CIs and *p*-values, respectively. By default, `c(cformat)` and `c(pformat)` are used.

`forest(meta_forest_opts)` passes `meta_forest_opts` as they are as options to `meta forestplot` (see [META] **meta forestplot**), which provides some control over the forest plot. For example, if you want to title the effect-size column **Correlation**, use the option `forest(columnopts(_es, title("Correlation")))`. Similarly, the option `forest(note(""))` removes the note stating the model used. `forest()` cannot be repeated, so all options must be passed in `meta_forest_opts`. Other options provided by `meta forestplot` may be used, but `smvmeta forestplot` and `meta forestplot` may not always play nicely with one another. The help file for `smvmeta` provides more details on how forest plots are constructed, particularly how the CIs are constructed for graphical display.

2.3 Stored results

`smvmeta set` stores the following characteristics:

| | |
|--------------------------|--------------|
| <code>_dta[z]</code> | <i>esvar</i> |
| <code>_dta[z_se]</code> | <i>sevar</i> |
| <code>_dta[fvvar]</code> | <i>fvvar</i> |

`smvmeta estimate` stores the following in `e()`:

Scalars

| | |
|---------------------------|--|
| <code>e(N)</code> | number of observations (estimates) included |
| <code>e(p)</code> | number of levels of <i>fvvar</i> included |
| <code>e(q)</code> | number of dimensions used to model correlation and heterogeneity |
| <code>e(converged)</code> | 1 if converged or 0 otherwise |

Macros

| | |
|----------------------------|----------------------|
| <code>e(cmd)</code> | <code>smvmeta</code> |
| <code>e(cmdline)</code> | command as typed |
| <code>e(depvar)</code> | <i>esvar</i> |
| <code>e(fvvar)</code> | <i>fvvar</i> |
| <code>e(properties)</code> | <code>b V</code> |

Matrices

| | |
|------------------------|---|
| <code>e(b)</code> | coefficient vector (not transformed) |
| <code>e(V)</code> | variance–covariance matrix of the estimators |
| <code>e(k)</code> | numbers of observations (estimates) included |
| <code>e(P)</code> | $p \times 4$ matrix of <i>P</i> -scores (one column per supported <code>superior()</code> option) |
| <code>e(I2)</code> | heterogeneity statistics |
| <code>e(levels)</code> | levels of <i>fvvar</i> included |
| <code>e(pvals)</code> | <i>p</i> -values |
| <code>e(eff_se)</code> | effective standard errors |

Functions

| | |
|------------------------|-------------------------|
| <code>e(sample)</code> | marks estimation sample |
|------------------------|-------------------------|

In addition to the above, the following are stored in `r()`:

| | |
|---------------------------|--|
| Scalars | |
| <code>r(level)</code> | confidence level |
| Macros | |
| <code>r(transform)</code> | canonical form of <i>transf_name</i> |
| Matrices | |
| <code>r(table)</code> | matrix containing transformed effect-size estimates and <code>r(level)% CIs</code> |

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any `r`-class command is run after the estimation command.

3 Example: Risk factors for pain after knee replacement

3.1 Background

The development of `smvmeta` was motivated by a systematic review of pain and function after total knee arthroplasty for osteoarthritis (Olsen et al. 2020, 2022, 2023). Around 20% of patients experience pain and poor function after surgery. The identification of risk factors for pain and function could lead to better patient outcomes.

The review identified and extracted estimates of association between risk factors measured at baseline prior to surgery and pain measured 12 months after surgery. Because it is generally not possible to randomize patients to risk factors, the review was based on observational studies. To facilitate meta-analysis of estimates of association reported using various metrics (for example, risk ratios, odds ratios, mean differences, and correlations) and between risk factors and pain assessments measured using a mix of binomial and continuous variables, correlation coefficients were extracted or imputed and then Fisher z -transformed. The following example is illustrative and may differ from the published systematic review.

3.2 Overview of pain12.dta

We begin by loading `pain12.dta` and listing its first 10 rows:

```
. use pain12
. list in 1/10
```

| | study | factor | z | z_se |
|-----|-------------------|-------------|-----------|----------|
| 1. | Lingard 2007 | Men. Health | .0242904 | .038518 |
| 2. | Papakostidou 2012 | Pain | .1760553 | .0713318 |
| 3. | Papakostidou 2012 | Older Age | -.0094856 | .0745588 |
| 4. | Papakostidou 2012 | Male | -.0968132 | .0735721 |
| 5. | Papakostidou 2012 | BMI | -.0395895 | .0744007 |
| 6. | Papakostidou 2012 | Support | .136985 | .0725892 |
| 7. | Papakostidou 2012 | Education | -.1422099 | .0724378 |
| 8. | Papakostidou 2012 | Urban/Semi | -.1482175 | .0722573 |
| 9. | Sullivan 2011 | Male | -.0012088 | .0945578 |
| 10. | Sullivan 2011 | Older Age | -.0012088 | .0945578 |

There are four variables with one effect-size estimate per row, as required by `smvmeta`:

- **study**: the studies that contribute estimates.
- **factor**: the risk factors. The first few risk factors are better mental health, pain, and older age (all measured at baseline).
- **z**: Fisher z -transformed correlation coefficients between presence of the factor (for binomial factors) or higher values of the factor (for risk factors measured on a continuum) and pain 12 months after surgery.
- **z_se**: specifies standard errors on **z**.

Preprocessing would have been required had the data not been arranged in `smvmeta`'s preferred form. For example, if the data were arranged in wide rather than long form—as used by Stata's `meta mvregress` (see [META] `meta mvregress`) and `mvmeta`—it would be necessary to use Stata's `reshape` (see [D] `reshape`) command.

3.3 Specify generic effect sizes, standard errors, and risk factors

The meta-analysis is specified using `smvmeta set`:

```
. smvmeta set z z_se factor
Meta-analysis setting information
Data
  No. observations: 51
      Sparse: Yes
Effect size
      Type: Generic
      Label: Correlation with pain (Fisher z-transformed)
      Variable: z
      Missing: 0
Precision
      Type: Standard error
      Label: SE on correlation
      Variable: z_se
      Missing: 0
Factor variable
      Label: Risk Factor
      Variable: factor
      Levels: 23
      Missing: 0
Model and method
      Model: Sparse multivariate random-effects meta-analysis (smvmeta)
      Method: Penalized maximum likelihood
```

Similarly to Stata's `meta set` (see [META] `meta set`), `smvmeta set` displays a summary of the data that will be used by subsequent `smvmeta` commands. We see that there are 51 observations (estimates of correlations with pain) and that the data are considered sparse (that is, amenable to analysis by `smvmeta`). If the sample is subsequently restricted for estimation (for example, using `if` or `by`), the restricted sample may not be considered sparse, but `smvmeta` will not report this. A summary of each of the three variables `smvmeta` needs to perform meta-analysis is displayed. The “types” of `esvar` and `sevar` are shown (“Generic” for effect sizes and “Standard error” for precisions). Types are not inferred from the data but displayed to tell you how `smvmeta` understands the data. Variable labels are displayed if they exist, along with variable names. The number of levels of the `fvar` variable is shown, which tells us there are 23 risk factors. The number of missing values is shown for each variable. Missing values may exist in the dataset but cannot be included in the estimation sample. Finally, the model and estimation method are stated. Meta-analysis can now be performed.

3.4 Meta-analysis using `smvmeta estimate`

Before running the meta-analysis, we first set the random-number generator's seed to ensure that we can reproduce results and define a local macro, `rep_opts`, containing reporting options that we will reuse in subsequent commands. We are estimating mean correlations, so we use `transform(corr)`. We consider risk factors that are more correlated with pain to be superior (irrespective of the direction of correlation), so we define superiority using `superior(big)` and specify a `sort()` option that will present results

in descending order by P -score (that is, the risk factor with greatest mean extent of superiority will appear at the top of the results).

```
. set seed 1234
. local rep_opts transform(corr) superior(big) sort(_Pscore, descending)
```

We then use `smvmeta estimate` to meta-analyze the data, modeling correlations and heterogeneity between effect sizes in three dimensions. To eliminate superfluous information from this article, we specify the `nolog` option. If logging is not disabled, then information about the penalized log likelihood will be displayed for each iteration. If logging is not disabled and the `dimension()` option is not specified, `smvmeta` will search for a dimensionality in which the data can be modeled. This may result in a long report that may include scary-looking messages from Stata's optimizer about "flat or discontinuous" regions of the parameter space. Such messages are generally not concerning unless the entire optimization fails (check `e(converged)`).

```
. smvmeta estimate, dim(3) nolog `rep_opts'
Sparse multivariate meta-analysis (smvmeta)      Number of obs   =      51
Factor variable label : Risk Factor              Num. factors    =      23
Optimization          : Penalized ML             Dimensions (q)  =       3
```

| | Coef. | Eff. SE | P>u | [95% Conf. Int.] | k | P-score | I2 |
|-------------|--------|---------|-------|------------------|---|---------|------|
| ROCF Recall | 0.342 | 0.126 | 0.005 | 0.089 0.554 | 1 | 92.4 | . |
| Catastroph. | 0.278 | 0.074 | 0.000 | 0.118 0.424 | 2 | 89.0 | 59.9 |
| Temp. Summ. | 0.211 | 0.076 | 0.005 | 0.050 0.361 | 2 | 77.1 | 0.0 |
| K-L Grade | -0.154 | 0.036 | 0.000 | -0.230 -0.075 | 3 | 65.3 | 0.0 |
| S. Joints | 0.153 | 0.036 | 0.000 | 0.073 0.230 | 2 | 64.9 | 0.0 |
| Heat Thold | 0.159 | 0.117 | 0.173 | -0.052 0.355 | 1 | 61.9 | . |
| Urban/Semi | -0.147 | 0.097 | 0.125 | -0.318 0.033 | 1 | 60.1 | . |
| C. Retain. | -0.140 | 0.054 | 0.009 | -0.252 -0.024 | 1 | 60.0 | . |
| Education | -0.141 | 0.100 | 0.156 | -0.317 0.044 | 1 | 58.4 | . |
| Support | 0.136 | 0.100 | 0.172 | -0.045 0.308 | 1 | 57.1 | . |
| Expectation | -0.142 | 0.150 | 0.341 | -0.365 0.096 | 1 | 56.9 | . |
| Pain | 0.111 | 0.032 | 0.000 | 0.042 0.178 | 8 | 50.6 | 97.3 |
| Warm Thold | 0.081 | 0.180 | 0.654 | -0.136 0.289 | 1 | 44.5 | . |
| Pain S.-E. | -0.029 | 0.268 | 0.913 | -0.198 0.141 | 1 | 38.9 | . |
| BMI | 0.062 | 0.047 | 0.182 | -0.027 0.150 | 4 | 35.3 | 70.9 |
| Men. Health | -0.048 | 0.105 | 0.646 | -0.174 0.079 | 5 | 34.9 | 79.6 |
| Function | -0.016 | 0.215 | 0.942 | -0.131 0.100 | 2 | 34.9 | 90.1 |
| Pat. Resur. | -0.047 | 0.091 | 0.603 | -0.165 0.072 | 1 | 33.8 | . |
| Male | -0.038 | 0.058 | 0.508 | -0.125 0.049 | 4 | 29.3 | 0.0 |
| Comorbidity | 0.032 | 0.077 | 0.676 | -0.063 0.126 | 3 | 29.1 | 0.0 |
| Older Age | -0.035 | 0.062 | 0.577 | -0.121 0.053 | 4 | 28.6 | 0.0 |
| Kinesophob. | 0.001 | 0.095 | 1.000 | -0.229 0.232 | 1 | 23.6 | . |
| Surg. Dur. | -0.001 | 0.095 | 1.000 | -0.232 0.229 | 1 | 23.6 | . |

Estimates have been tanh-transformed.

Untransformed effect sizes with larger magnitudes are superior.

The header section of the results table summarizes the analysis, providing information such as the number of observations (effect-size estimates) included in the analysis and the number of dimensions used to model correlation and heterogeneity. Each row of the table shows results for a particular risk factor. The first result is for Rey-Osterrieth

complex figure (ROCF) recall, which assesses functional decline on multiple cognitive dimensions. Working left to right, the mean correlation between this variable and pain is estimated to be 0.342 with an effective standard error of 0.126 (section 4.3.3). The p -value testing the hypothesis that the mean correlation is zero is 0.005 (section 4.3.3). The 95% CI for mean correlation is [0.089 to 0.554] (section 4.3.3). Only one observation (study result) of correlation was used in the meta-analysis (column *k*). The P -score assessing the mean extent of certainty that ROCF recall is superior to (has mean correlation with larger magnitude than) all the other risk factors is 92.4% (section 4.4). The I^2 heterogeneity statistic cannot be computed, because only one estimate (study result) of correlation was available. The footer tells us if and how the estimates have been transformed and what definition of superiority was used (see sections 2.2 and 4.4.2).

In principle, we can now use almost all of Stata's postestimation commands. For example, we could `test` (see [R] `test`) if the correlation between urban or semiurban residency and pain is zero. To do this, we need to know how this variable is coded. We could use `smvmeta`, `coeflegend` to determine the code, but using a local macro is perhaps more readable:

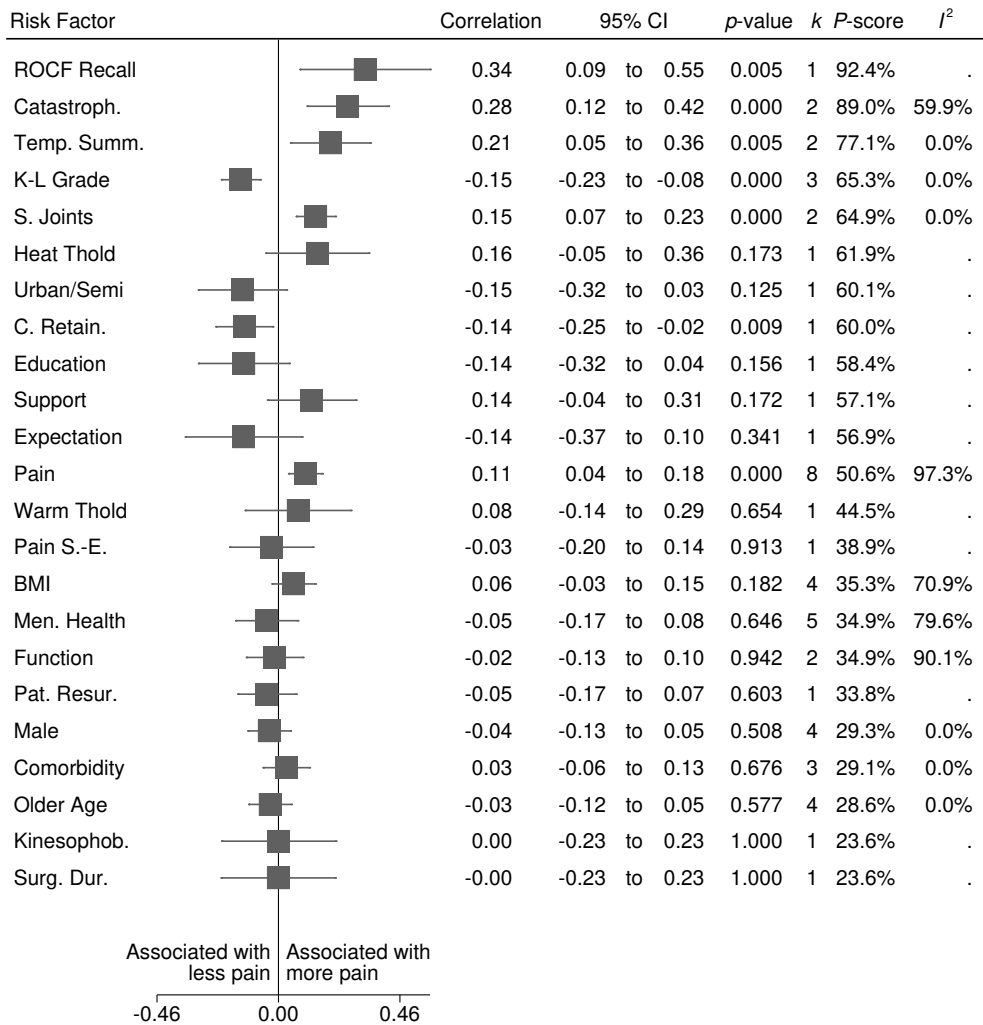
```
. local semi_urban = "Urban/Semi":`value label factor'
. test `semi_urban'.factor
( 1) 23.factor = 0
      chi2( 1) =    4.02
      Prob > chi2 =   0.0450
```

This test gives a p -value of 0.045, suggesting that studies like those included in the meta-analysis generally estimate nonzero correlations between urban or semiurban residency and pain. However, the p -value provided by `smvmeta` ($p = 0.125$) is different and does not reject the hypothesis using the conventional $p < 0.05$ criterion. Section 4.3.3 explains the discrepancy. Let this example encourage you to be careful when using postestimation commands after `smvmeta`.

3.5 Present estimates graphically using `smvmeta forestplot`

The command below uses `smvmeta forestplot` to plot the stored estimates. `smvmeta` does not “remember” `rep_opts`, so one must specify how results should be presented, hence the definition of the `rep_opts` macro previously. For the same reason, we now define the local macro `forest_opts`, which specifies how estimates and p -values should be formatted. The `forest()` options specify a title for the column of effect-size estimates and what the directions of correlation mean.

```
. local forest_opts cformat(%5.2f) pformat(%5.3f)
. smvmeta forestplot, `rep_opts' `forest_opts'
> forest(columnopts(_es, title("Correlation")))
> nullrefline(favorsleft("Associated with" "less pain")
> favorsright("Associated with" "more pain"))
```



Sparse multivariate random-effects meta-analysis model

Figure 2. Forest plot showing estimates of mean correlation between risk factors and pain 12 months after total knee arthroplasty, the number of studies (*k*) contributing estimates, *P*-scores that measure the mean extent of certainty of superiority, and *I*² values that measure the percentage of heterogeneity not explained by sampling error

The forest plot is shown in figure 2. The three risk factors estimated to have the greatest mean extent of certainty of superiority (that is, the highest P -scores; see section 4.4) are cognitive in nature:

- ROCF recall (see above; mean correlation 0.34; 95% CI [0.09 to 0.55]; P -score 92.4%). However, this finding is supported only by one study.
- Pain catastrophizing (0.28; 95% CI [0.12 to 0.42]; P -score 89%; 2 studies; I^2 59.9%).
- Temporal summation, which—informally—assesses pain experienced with respect to the frequency of a controlled pain stimulus (0.21; 95% CI [0.05 to 0.36]; P -score 77.1%; 2 studies; I^2 0.0%).

While it is useful to show estimates for all risk factors and all quantities that we can report, we might want to produce a simpler forest plot, for example, showing only a subset of the available columns, for “interesting” results. This can be done by specifying the forest plot columns of interest and an expression that restricts the results that should be presented (but does not change the estimates). We specify that we want to show columns for the names of the risk factors, the forest graph, effect-size point estimates (correlations, as transformed via the `rep_opts` macro), and the number of estimates supporting each risk factor. We use the `if()` option to restrict the results to estimates with (transformed) point estimates at least 0.15 in magnitude and that have p -values less than 0.05:

```

. smvmeta forestplot _fv _plot _es _k, `rep_opts' `forest_opts'
> forest(columnopts(_es, title("Correlation")))
> columnopts(_k, title("Num. studies"))
> nullrefline(favorsleft("Associated with" "less pain")
> favorsright("Associated with" "more pain"))
> note("")
> if(abs(_es) >= 0.15 & _p < 0.05)

```

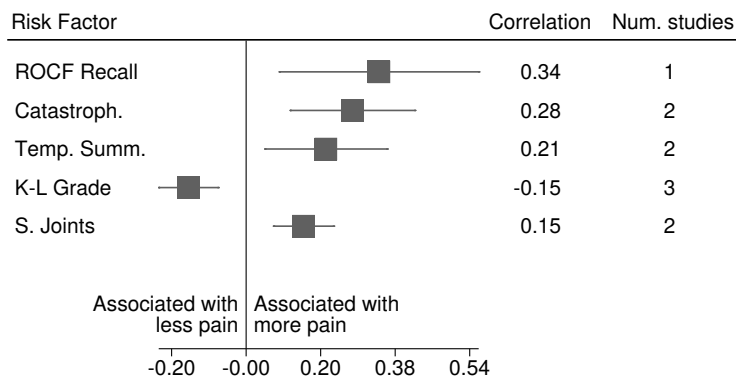


Figure 3. A simplified presentation of the results, restricted to correlations with point estimates of at least 0.15 in magnitude and p -values less than 0.05

The forest plot is shown in figure 3. In addition to the three cognitive risk factors discussed above, the plot shows estimates for Kellgren–Lawrence grade (a tool for assessing radiological evidence of osteoarthritis) and number of symptomatic joints.

4 The estimator and its implementation

4.1 Assumptions and nomenclature

Let \mathbf{y} be an n -vector of effect-size estimates (that is, *esvar*) and $\mathbf{\Lambda}$ be an $n \times n$ diagonal matrix of sampling variances (that is, the squares of *sevar*), such that $\Lambda_{i,i}$ is the sampling variance for y_i . Let \mathbf{X} be an $n \times p$ design matrix whose (i, j) th element is unity if effect size i corresponds to factor level j and is zero otherwise. We make the following assumptions:

1. The n estimates of the p effect sizes are exchangeable across study.
2. There are no nonrandom study-level effects.
3. There is nonnegligible correlation and heterogeneity between effect sizes and studies that is well approximated in a low-dimensional space using a $q \times q$ covariance matrix $\mathbf{\Sigma}$, where $2 \leq q < p$.
4. No further information is available about the structure of the correlations and heterogeneity (for example, studies do not publish correlations between estimates).

Assumptions 1 and 2 explain why it is not necessary to specify to `smvmeta` the studies from which estimates arise. If more is known about the structure of the correlations and heterogeneity, then the problem may not in fact be sparse, and `smvmeta` may not be a good model choice. Obvious alternatives are Stata's multivariate meta-analysis commands and `mvmeta`.

4.2 Sparsity and smvmeta's multivariate random-effects model

In the most general case, there may be correlation and heterogeneity between each pair of effect sizes. This gives p mean effect sizes, $p(p-1)/2$ correlations, and $p(p+1)/2$ covariances that model heterogeneity for a total of $p+p^2$ model parameters. If p is even moderately large, then $p+p^2$ is likely to be substantially larger than n , the number of estimates available to support estimation. `smvmeta` considers data to be sparse if $n < p+p^2$.

`smvmeta` makes the simplifying assumption that correlations and heterogeneity do not need to be modeled separately, so effect-size estimates \mathbf{y} can be modeled as multivariate normal with mean $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\mathbf{\Lambda} + \mathbf{X}\boldsymbol{\Psi}\mathbf{X}^\top$, where $\mathbf{\Lambda}$ and \mathbf{X} are as defined in section 4.1, $\boldsymbol{\beta}$ is a p -vector of mean effect sizes, and $\boldsymbol{\Psi}$ is a $p \times p$ covariance matrix that models correlation and heterogeneity.

Assuming unstructured $\boldsymbol{\Psi}$, it appears necessary to estimate $p+p(p+1)/2$ model parameters (that is, the p parameters of $\boldsymbol{\beta}$ plus one of the triangles of $\boldsymbol{\Psi}$). This is still quadratic in p . However, if we can assume that $\boldsymbol{\Psi}$ is anisotropic, then it may be possible to approximate $\boldsymbol{\Psi}$ in a space of dimension $q < p$. This can be achieved by defining a random projection \mathbf{R} between \mathbb{R}^p and \mathbb{R}^q and estimating a $q \times q$ covariance matrix $\mathbf{\Sigma}$,

where $\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^\top \approx \boldsymbol{\Psi}$. This reduces the number of parameters that must be estimated to $p + q(q + 1)/2$, which is linear in p and quadratic in q rather than p . This gives the following model:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Lambda} + \mathbf{X}\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^\top\mathbf{X}^\top) \quad (1)$$

A related dimensionality-reduction problem is principal component analysis (Pearson 1901; Jolliffe 2005). In principal component analysis, a linear transformation into a low-dimensional space can be defined via eigendecomposition of a known or estimated covariance matrix. A matrix of the eigenvectors with largest eigenvalues defines a transform that preserves at least a chosen proportion of total variation. This is possible if the covariance matrix is known or can be estimated but cannot be used to choose q because $\boldsymbol{\Psi}$ is unknown.

The term “random projection” is typically used in the context of mapping point sets from high- to low-dimensional spaces in a way that approximately preserves some useful aspect of the data, particularly relative distances between points. `smvmeta` uses random projection to establish an arbitrary low-dimensional orthonormal basis in which an approximation to $\boldsymbol{\Psi}$ can be estimated.

For some meta-analyses, it may be possible to use a priori domain knowledge to choose q . For example, perhaps it is reasonable to consider some effect sizes as facets of a common construct for which distinct estimates are desired, preventing pooling at the level of the construct, but facilitating dimensionality reduction. In general, however, domain knowledge may not be available to choose q . If q is not specified via the `dimension()` option, `smvmeta` will attempt to find a suitable value. Section 4.3 describes how this is done. As for principal components analysis, good results can be achieved with surprisingly small q .

4.3 Estimation

4.3.1 Penalized maximum likelihood

The primary estimation target is $\boldsymbol{\beta}$, though $\boldsymbol{\Sigma}$ is unknown and must also be addressed. White (2009) summarizes estimation methods applicable to the multivariate setting and chooses likelihood-based methods for their generality and optimality properties. There is little support for estimating $\boldsymbol{\Sigma}$ when data are sparse. The log-likelihood function corresponding to (1) can be trivially maximized by setting $\boldsymbol{\Sigma}$ close to the zero matrix, which violates assumption 3. This is addressed using penalized maximum likelihood (Cole, Chu, and Greenland 2014).

The log density corresponding to $\boldsymbol{\Lambda} + \mathbf{X}\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^\top\mathbf{X}^\top \sim \mathcal{W}^{-1}(\mathbf{I}, p)$, an inverse-Wishart distribution, is added to the log likelihood for (1) to penalize values of $\boldsymbol{\Sigma}$ that are incompatible with assumption 3 while otherwise assuming as little as possible. This results in the following penalized log-likelihood problem. `smvmeta` uses Mata’s [M-5] `optimize()` to find $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\hat{\boldsymbol{\theta}})$ with

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \frac{1}{4} \left[-2\mathbf{z}(\boldsymbol{\theta})^\top \mathbf{Q}(\boldsymbol{\theta})^{-1} \mathbf{z}(\boldsymbol{\theta}) - 2 \log \text{tr}(\mathbf{Q}(\boldsymbol{\theta})^{-1}) \right. \\ \left. - 4(p+1) \log |\mathbf{Q}(\boldsymbol{\theta})| + 4 \log \Gamma_p \left(\frac{p}{2} \right) \right. \\ \left. - 2n \log 2\pi + p^2 \log \pi - p^2 \log 4 - p \log \pi \right]$$

where $\mathbf{Q}(\boldsymbol{\theta}) = \boldsymbol{\Lambda} + \mathbf{X}\mathbf{R}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{R}^\top\mathbf{X}^\top$; $\mathbf{z}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\beta}(\boldsymbol{\theta})$; $\boldsymbol{\beta}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ form the p -dimensional mean vector and q -dimensional covariance matrix, respectively; and Γ_p is the multivariate gamma function. Matrix products $\mathbf{X}\mathbf{R}$ and $\mathbf{R}^\top\mathbf{X}^\top$ are independent of $\boldsymbol{\theta}$ and can be precomputed. Symmetry and positive semidefiniteness of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is ensured by constructing the covariance matrix from Cholesky factors. Random projection matrix \mathbf{R} is formed by sampling elements independently and identically distributed from $N(0, 1)$ and then applying a singular value decomposition to ensure orthonormality.

4.3.2 Optimization strategy

The optimization problem described above can be nontrivial. I found that Stata's default optimization algorithm (modified Newton–Raphson) sometimes failed to converge and other algorithms often failed to start in a fruitful direction. Optimization is performed by initializing the elements of $\boldsymbol{\theta}$ that correspond to the effect sizes with univariate meta-analysis estimates, running 10 modified Newton–Raphson iterations, and then switching to Broyden–Fletcher–Goldfarb–Shanno. If q is not specified, `smvmeta` will search for the largest value of q satisfying $2 \leq p + q(q+1)/2 < n$. However, admissible values of q do not exist for all possible values of p and n . `smvmeta` will issue an error if an admissible value of q does not exist.

Optimization can fail if q is too high. `smvmeta` detects lack of convergence and will restart optimization with the next smallest admissible value of q (and a new random projection) until no such values remain. The actual value of q used is stored in `e(V)` and may differ from that specified by `dimension()`.

4.3.3 Constructing CIs and p -values

We seek a test statistic and its sampling distribution to construct $100(1 - \alpha)\%$ CIs and p -values on the elements of effect-size vector $\boldsymbol{\beta}$. I found that Wald-type CIs constructed using a normal approximation of the sampling distribution of $\hat{\boldsymbol{\theta}}$ do not always provide nominal coverage (and similarly for p -values). Thus, be careful if you use `e(V)` for postestimation, as is done in commands such as Stata's `test` and `margins` (see [R] `test` and [R] `margins`). Instead, `smvmeta` uses profile penalized likelihood.

Cole, Chu, and Greenland (2014) provide a useful introduction to likelihood ratio-based CIs. Briefly, in the nonpenalized case, CIs can be constructed for β_j (the j th effect size) on the basis that

$$2 \left\{ f(\hat{\boldsymbol{\theta}}) - f_j(\beta) \right\} \sim \chi_1^2 \quad (2)$$

where f is the log-likelihood function and $f_j(\beta) = \max_{\theta|\beta_j=\beta} f(\theta)$ is the maximum value of the log-likelihood function with the element of θ corresponding to the j th element of β constrained to be the scalar value β . This allows a critical value corresponding to a specified confidence level to be found and used to compute the lower and upper limits, β_L and β_U , of a CI on β_j . This section explains how the nonpenalized profile likelihood method can be modified to account for penalization.

Let g and g_j be log-penalty equivalents of f and f_j . To account for penalization, we are interested in the distribution of $\{f(\hat{\theta}) - f_j(\beta)\} + \{g(\hat{\theta}) - g_j(\beta)\}$. Recall that if $X \sim \chi_\nu^2$, then $kX \sim \Gamma(\nu/2, 2k)$, a gamma distribution in the shape-scale parameterization. We can therefore restate (2) as $F_j = f(\hat{\theta}) - f_j(\beta) \sim \Gamma(1/2, 1)$ and, by the same reasoning, define $G_j = g(\hat{\theta}) - g_j(\beta) \sim \Gamma(1/2, 1)$. We now seek the distribution of $U_j = F_j + G_j$, the sum of two gamma variates.

The sum of an arbitrary number of gamma variates has moment-generating function $M(s) = s^{-1} \prod_i (1 - \lambda_i s)^{-a}$, where a is a shape parameter common to the gamma variates and λ_i is the i th eigenvalue of a matrix constructed from the scale parameters and the correlation coefficients between the variates (Alouini, Abdi, and Kaveh 2001). In the case of the sum of two gamma variates with correlation ρ , this simplifies to

$$M(s) = \frac{1}{s\sqrt{1+s} - s\sqrt{\rho}\sqrt{1+s+s\sqrt{\rho}}} \quad (3)$$

The inverse Laplace transform of (3) can be used to obtain the cumulative distribution function of U_j for a particular value of ρ and hence a sampling distribution that can be used to compute critical values and p -values.

`smvmeta` reports CIs constructed using $\rho = 0$, which corresponds to $U_j \sim \text{Exp}(1)$. This gives the shortest intervals if $100(1 - \alpha)\% \gtrsim 78.5\%$ ¹ and the longest ones otherwise. Validation experiments (section 5) show that this provides CIs with empirical coverage close to the nominal level but that are shorter compared with those provided by meta-regressions that do not model correlation and heterogeneity between effect sizes. In summary, approximate CIs are constructed on the basis that

$$\left\{f(\hat{\theta}) - f_j(\beta)\right\} + \left\{g(\hat{\theta}) - g_j(\beta)\right\} \sim \text{Exp}(1) \quad (4)$$

Equation (4) can be used to test the hypothesis $H_j : \beta_j = 0$. The test statistic is $u_j = \{f(\hat{\theta}) - f_j(0)\} + \{g(\hat{\theta}) - g_j(0)\}$, and the p -value is $P(U_j \geq u_j | H_j) = e^{-u_j}$. Recall from section 3.4 that Stata's `test` reported a p -value that is different from that reported by `smvmeta`. The discrepancy is explained by the different test statistics and their distributions (`test` uses a χ_1^2 statistic).

Having identified the distribution from which a critical value can be computed, we can find bounds β_L and β_U for β_j efficiently via a quadratic approximation of ℓ around $\hat{\theta}$.

1. I could not persuade Mathematica to provide a closed-form solution and suspect it does not exist.

The presentation below is based on that for the LOGISTIC procedure of SAS/STAT (SAS Institute, Inc. 2016). `smvmeta` searches for β_L and β_U such that

$$\ell(\hat{\boldsymbol{\theta}}) - \ell_j(\beta_L) = \ell(\hat{\boldsymbol{\theta}}) - \ell_j(\beta_U) = \log \frac{1}{1-\alpha}$$

where $\ell(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + g(\boldsymbol{\theta})$, $\ell_j(\beta) = f_j(\beta) + g_j(\beta)$, and the $100(1-\alpha)$ th percentile of $\text{Exp}(1)$ is given by $\log 1/(1-\alpha)$. Figure 4 illustrates the construction of u_j , β_L , and β_U .

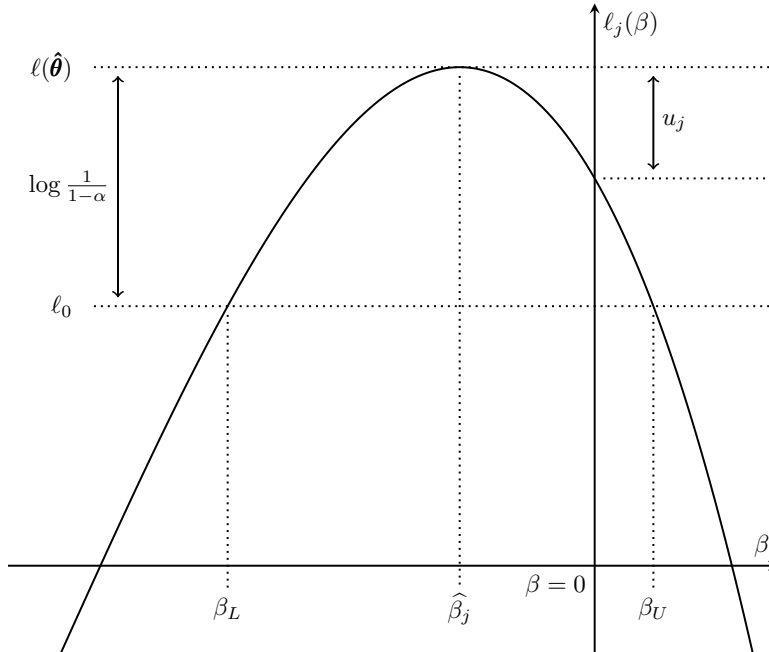


Figure 4. Constructing the test statistic, u_j for $H_j : \beta_j = 0$, and approximate CI bounds β_L and β_U for β_j , via profile penalized likelihood

The quadratic approximation

$$\tilde{\ell}(\boldsymbol{\theta} + \boldsymbol{\delta}) = \ell(\boldsymbol{\theta}) + \boldsymbol{\delta}^\top \nabla \ell(\boldsymbol{\theta}) + \frac{1}{2} \boldsymbol{\delta}^\top \mathbf{V}(\boldsymbol{\theta}) \boldsymbol{\delta}$$

is used to find a value of $\boldsymbol{\delta}$ that can be used in the update rule $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + \boldsymbol{\delta}$ to minimize $|\ell(\boldsymbol{\theta}) - \ell_0|$, where $\ell_0 = \ell(\hat{\boldsymbol{\theta}}) - \log 1/(1-\alpha)$ and $\nabla \ell(\boldsymbol{\theta})$ and $\mathbf{V}(\boldsymbol{\theta})$ are the gradient vector and Hessian matrix of ℓ at $\boldsymbol{\theta}$. The next value of $\boldsymbol{\delta}$ is found by solving

$$\frac{d}{d\boldsymbol{\delta}} \tilde{\ell}(\boldsymbol{\theta} + \boldsymbol{\delta}) + \frac{d}{d\boldsymbol{\delta}} \eta(\mathbf{e}_j^\top \boldsymbol{\delta} - \beta) = \mathbf{0}$$

where η is a Lagrange multiplier and \mathbf{e}_j is a unit vector that indicates the dimension of $\boldsymbol{\theta}$ corresponding to β_j , giving

$$\boldsymbol{\delta} = -\mathbf{V}(\boldsymbol{\theta})^{-1}\{\nabla\ell(\boldsymbol{\theta}) + \eta\mathbf{e}_j\} \text{ with } \eta = \pm \left[\frac{2\{\ell_0 - \ell(\boldsymbol{\theta}) + \frac{1}{2}\nabla\ell(\boldsymbol{\theta})^\top\mathbf{V}(\boldsymbol{\theta})^{-1}\nabla\ell(\boldsymbol{\theta})\}}{\mathbf{e}_j^\top\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{e}_j} \right]^{\frac{1}{2}} \quad (5)$$

Negative and positive values of η yield lower and upper bounds β_L and β_U , respectively. The quantities in (5) are iteratively updated, starting at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, until $|\ell(\boldsymbol{\theta}) - \ell_0| \leq \epsilon$ and $\{\nabla\ell(\boldsymbol{\theta}) + \eta\mathbf{e}_j\}^\top\mathbf{V}(\boldsymbol{\theta})^{-1}\{\nabla\ell(\boldsymbol{\theta}) + \eta\mathbf{e}_j\} \leq \epsilon$ with $\epsilon = 1 \times 10^{-4}$.

Recall that Wald-type CIs that are constructed from standard errors obtained from the variance–covariance matrix of the estimators do not necessarily have nominal coverage. `smvmeta` therefore reports *effective* standard errors imputed under the assumption that p -values computed as above arise from a test statistic that is normally distributed. This is reported in the results table as `Eff. SE`, stored as `e(eff_se)`, and used to compute P -scores to assess superiority.

4.3.4 Alternative approaches to penalization

Penalized maximum likelihood is used by `smvmeta` to prevent trivial solutions that collapse $\boldsymbol{\Sigma}$ (section 4.3). The penalty used arises from an inverse Wishart distribution over the approximation of $\boldsymbol{\Psi}$. In the Bayesian framework, this distribution is the conjugate prior for the covariance matrix of the multivariate normal. This prior has been criticized within Bayesian statistics (for example, see Schuurman, Grasman, and Hamaker [2016]). Some criticism concerns the distribution’s inflexibility: it is not possible to construct an inverse Wishart prior that expresses that some covariances are known a priori with more precision than others. However, this is not relevant to `smvmeta`. Other criticism concerns the low density of covariances close to zero, but this is exactly why it yields a useful penalty in `smvmeta`. One criticism that may have merit is that larger variances are associated with correlations with absolute values close to unity (Tokuda et al. 2011). Alternative distributions are used within Bayesian statistics to construct priors for covariance matrices, such as the Lewandowski–Kurowicka–Joe distribution over correlation matrices.

The model used by `smvmeta` does not attempt to estimate or disambiguate the full correlational and heterogeneity structure, which is approximated in a low-dimensional space. While most research questions are likely focused on mean effect sizes, correlation and heterogeneity may also be of interest. However, there is very little information about these $p \times p$ matrices in the setting addressed herein. While $\boldsymbol{\Sigma}$ is estimated, it is defined in an arbitrary low-dimensional space that precludes interpretation. It is therefore not stored. While the approximation of $\boldsymbol{\Psi}$ is interpretable in principle, it is subject to run-to-run variance arising from the random projection and spans at most a q -dimensional subspace of \mathbb{R}^p . This matrix is not stored to prevent over interpretation.

4.4 Assessing superiority

Multivariate meta-analyses are often motivated by a need to know which effect size is “superior” or to rank them. For example, it is natural to ask which risk factor is most important or which of several treatments is most effective. `smvmeta` facilitates this by reporting P -scores (Rücker and Schwarzer 2015). Originating in frequentist network meta-analysis (which can be posed as a multivariate meta-analysis), a P -score is a statistic that measures the mean extent of certainty that a given effect size is superior to all the others.² A P -score reflects not only the estimated effect size relative to those of the others but also the precision of the estimates. This section explains how P -scores are computed and how various definitions of superiority are implemented in `smvmeta`. It ends with a discussion of the advantages and disadvantages of P -scores over alternatives.

4.4.1 Computing P -scores

The following presentation differs slightly from that of Rücker and Schwarzer (2015) but describes the method concisely in a way that is hopefully relatively intuitive. Assume momentarily that, of two effect sizes, the one closest to $+\infty$ is superior. Let $\pi_{i,j}$ be a one-sided p -value for the null hypothesis $\beta_i \leq \beta_j$. Then define the $p \times p$ matrix \mathbf{P} with elements

$$P_{i,j} = \begin{cases} 1 - \pi_{i,j} & \text{if } \hat{\beta}_i \leq \hat{\beta}_j \\ \pi_{i,j} & \text{otherwise} \end{cases} \quad \forall i, j \in \{1, \dots, p\}$$

The elements of this matrix are simply 1 minus the one-sided p -values, such that $P_{i,j}$ will be close to unity if there is strong evidence that $\beta_i > \beta_j$. `smvmeta` calculates p -values using a normal approximation and effective standard errors. A vector of P -scores is then computed as

$$\mathbf{p} = \frac{1}{p-1} (\mathbf{P} \mathbf{1}_p - \text{diag } \mathbf{P}) \times 100\%$$

where $\mathbf{1}_p$ is a $p \times 1$ vector of ones. The i th element of \mathbf{p} is the mean of the 1 minus p -values over all but the i th effect size, expressed as a percentage. The one-minus in the definition of matrix \mathbf{P} is used so that P -scores closer to 100%, rather than 0%, can be interpreted as superior. P -scores are expressed as percentages but do not generally sum to 100%.

4.4.2 Defining superiority

So far we have assumed that, of two effect sizes, the one closest to $+\infty$ is superior. This corresponds to the `superior(+inf)` option and is the default. This could be applicable in a meta-analysis of risk factors in which relative risks *greater* than unity are associated with the outcome. Risk factors with P -scores closer to 100% would be interpreted as superior to those closer to 0%.

2. P -scores are often interpreted as estimating the probability that an effect size is superior to all others. This explanation is probably useful to laypeople and as a shorthand, but from the frequentist perspective, an effect size (that is, its true value) is or is not superior to another effect size, so the only admissible probabilities are 0 and 1.

If relative risks less than unity were associated with the outcome, it may be useful to reverse the definition of superiority so that a P -score of 100% would have the interpretation of superiority rather than inferiority. Similarly, in meta-analyses of correlation coefficients, correlations with large magnitudes may be considered superior to small correlations. Section 2.2.2 describes four options for `superior()`. These are implemented by temporarily transforming $\hat{\beta}$ in the computation of P -scores:

$$\hat{\beta} \rightarrow \begin{cases} +\hat{\beta} & \text{if } \text{superior}(+\text{inf}) \\ -\hat{\beta} & \text{if } \text{superior}(-\text{inf}) \\ +|\hat{\beta}| & \text{if } \text{superior}(\text{big}) \\ -|\hat{\beta}| & \text{if } \text{superior}(\text{small}) \end{cases}$$

`smvmeta` computes and stores P -scores using all four definitions, allowing results to be extracted or redisplayed using alternatives as desired. P -scores are computed in the metric in which the meta-analysis was posed. This is relevant to `transform(efficacy)`, which reverses estimates' signs.

4.4.3 Alternative methods for assessing superiority

P -scores and their Bayesian equivalent, surface under the cumulative ranking curve (SUCRA) values (Salanti, Ades, and Ioannidis 2011), have been argued to be preferable to the simpler approach of estimating a Bayesian posterior probability that a given effect size is superior to all others, for example, as implemented by the `pbest()` option of `mvmeta` (White 2011).

A weakness of estimating posterior probabilities of superiority is that an imprecisely estimated effect size may have substantial probability mass for effect sizes that are implausibly large in magnitude. This phenomenon is most apparent in meta-analyses of correlations in which the magnitude of a coefficient, rather than its magnitude and direction, determines superiority. The probability mass in the tails of a posterior distribution that extends far in the negative and positive directions “doubles up” after taking the magnitude of the effect size. Consequently, the posterior probability that the effect size is indeed superior (or inferior) may also be large.

There is nothing wrong with such a probability: there really is a high posterior probability of superiority.³ However, such a result can lack face validity and cause nonstatisticians to doubt the entire analysis. This is not unreasonable. Few people should accept that an imprecisely estimated quantity is highly likely to be superior to other quantities that are supported by much stronger evidence. P -scores and SUCRA values are quite robust to this problem but are admittedly more challenging to interpret.

An additional benefit of P -scores is that they are trivial to compute. In particular, it is not necessary to use sampling-based estimation methods, as is the case with SUCRA values and posterior probabilities of superiority. This means that results can be obtained quickly.

3. Perhaps “high posterior *possibility* of superiority” would be clearer.

5 Simulation-based validation

I ran three simulation experiments to validate a prerelease version of `smvmeta`, using simulated effect sizes similar to the pain data (section 3). The numbers of effect sizes, p , ranged from 10 to 30. Standard errors on the simulated effect-size estimates and effect-size heterogeneity were generated to approximately match the distributions of standard errors and I^2 values in the pain data.

In the first two experiments, effect-size estimates were missing within each simulated study either completely at random (MCAR) or not at random (MNAR). The MNAR scenario used almost the same data as the MCAR scenario, but I discarded effect-size estimates—the values “published” by simulated studies—if their CIs included ± 0.05 . This models a publication bias scenario in which studies do not publish effect-size estimates if they are small or imprecisely estimated. I assumed no “domain knowledge” was available to choose q , the dimension in which correlation and heterogeneity are approximated. In these scenarios, `smvmeta` chooses q automatically.

The third experiment was an MCAR scenario in which a priori estimates of q were assumed to be available. Values of q were obtained by applying principal components analysis to the known correlation matrices used to generate the simulated data. In each of these meta-analyses, I set `smvmeta`’s `dimension()` option to the minimum number of components that accounted for at least 90% of total variation. However, recall that `smvmeta` will try to use a different value of q if the specified model does not converge.

I generated synthetic data for 1,000 meta-analyses for each experiment. I used the known mean effect sizes to quantify and compare bias and empirical coverage of 95% CIs with results of random-effects meta-regressions in which mean effect size is estimated via a categorical covariate. This is a simple alternative model choice that can be applied to high-dimensional sparse data. It does not account for correlations between effect sizes and assumes heterogeneity is the same for all effect sizes.

Results for the three scenarios were very similar, suggesting that `smvmeta` has some robustness to the MNAR scenario and performs well even if nothing can be assumed about q . Because the results were so similar, I will report the MCAR results unless stated otherwise. Overall bias is about 1.03 (95% CI [1.02 to 1.04]) times larger for `smvmeta` compared with meta-regression, but this was negligible in absolute terms for most meta-analyses. Overall empirical coverage of 95% CIs is 94.4% (95% CI [94.1% to 94.7%]). If p is small (close to 10, say), there is some evidence that bias may be very high and CIs may be excessively wide (empirical coverage is unaffected). There is also some evidence that empirical coverage may drop slightly below the nominal level if p is close to 30, but there is considerable uncertainty on this finding.

These findings are perhaps unsurprising given that random projection works less well in lower dimensions and that it will not always be possible to model correlation and heterogeneity between many effect sizes in a low-dimensional space. The experiments suggest that `smvmeta` can probably be used safely if there are at least 15–20 effect sizes. For $p > 20$ effect sizes, `smvmeta`'s CIs are on average about 90.3% (95% CI [88.0% to 92.7%]) of the length of those provided by meta-regression, and the results suggest they become shorter as p increases.

In the third experiment, the prespecified value of q was used in about 87% of meta-analyses. Around 10% of meta-analyses used a value of q 1 less than the prespecified value. The median number of dimensions used was six. Fewer than 1% of analyses used a value of q less than 4, and fewer than 1% used a value greater than 7. These findings suggest that $4 \leq q \leq 7$ is probably reasonable if you want to specify q and the number of effect sizes is not more than 30, but `smvmeta` does a good job of finding a sensible value if you do not specify q .

6 Conclusions

This article has presented `smvmeta`, a new command for sparse multivariate random-effects meta-analysis. The underlying model makes estimation tractable in the sparse setting by modeling correlation and heterogeneity in a low-dimensional space via random projection. `smvmeta`'s syntax is modeled on Stata's univariate meta-analysis commands (see [META] `meta`) with the aim of making the command feel familiar and easy to use.

The command was demonstrated in an example multivariate meta-analysis of pain after total knee arthroplasty. Mathematical and computational details of the model and P -scores were then presented. Simulation-based validation experiments show that, in the higher-dimensional sparse setting for which `smvmeta` was developed, the method is expected to provide substantially more precise estimates (that is, narrower CIs) at little cost in bias compared with random-effects meta-regression.

The fundamental limitations of `smvmeta` are due to the simplifying assumptions that facilitate random-effects meta-analysis in the challenging sparse multivariate setting. In particular, modeling correlation and heterogeneity in a low-dimensional space using one covariance matrix combines two distinct concepts, thus precluding interpretation of variance components and focusing estimation on mean effect sizes. While estimating means is almost always the main purpose of random-effects meta-analysis, it is nevertheless also important to estimate and report variance in effect sizes (IntHout et al. 2016). Future work could explore alternative approximations that could be used in the sparse setting to facilitate such estimation. However, despite the approximations used, the simulation-based validation provides reassurance that trustworthy and precise estimates can be expected if the number of effect sizes p is large enough. This holds regardless of whether sufficient domain knowledge is available to specify q , the number of dimensions in which to model correlation and heterogeneity, and also if missing effect-size estimates are not MCAR but instead are missing because of a typical MNAR mechanism by which small or imprecise (“nonsignificant”) estimates go unreported.

7 Acknowledgments

The author thanks Unni Olsen, Maren Falch Lindberg, and Anners Lerdal (University of Oslo and Lovisenberg Diaconal Hospital, Norway), Eva Marie-Louise Denison (Norwegian Institute of Public Health), and Arild Aamodt (Lovisenberg Diaconal Hospital) for their work on the systematic reviews that motivated development of `smvmeta`. He also thanks Stephen Jenkins (London School of Economics and Political Science, United Kingdom) and an anonymous reviewer, whose detailed comments improved this article and the code.

8 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 24-2
. net install st0749      (to install program files, if available)
. net get st0749         (to install ancillary files, if available)
```

9 References

- Alouini, M.-S., A. Abdi, and M. Kaveh. 2001. Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels. *IEEE Transactions on Vehicular Technology* 50: 1471–1480. <https://doi.org/10.1109/25.966578>.
- Cole, S. R., H. Chu, and S. Greenland. 2014. Maximum likelihood, profile likelihood, and penalized likelihood: A primer. *American Journal of Epidemiology* 179: 252–260. <https://doi.org/10.1093/aje/kwt245>.
- IntHout, J., J. P. A. Ioannidis, M. M. Rovers, and J. J. Goeman. 2016. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 6: e010247. <https://doi.org/10.1136/bmjopen-2015-010247>.
- Jolliffe, I. 2005. Principal component analysis. In Vol. 3 of *Encyclopedia of Statistics in Behavioral Science*, ed. B. S. Everitt and D. C. Howell. Chichester, U.K.: Wiley. <https://doi.org/10.1002/0470013192.bsa501>.
- Lin, L., and H. Chu. 2018. Bayesian multivariate meta-analysis of multiple factors. *Research Synthesis Methods* 9: 261–272. <https://doi.org/10.1002/jrsm.1293>.
- Olsen, U., M. F. Lindberg, E. M.-L. Denison, C. J. Rose, C. L. Gay, A. Aamodt, J. I. Brox, et al. 2020. Predictors of chronic pain and level of physical function in total knee arthroplasty: A protocol for a systematic review and meta-analysis. *BMJ Open* 10: e037674. <https://doi.org/10.1136/bmjopen-2020-037674>.
- Olsen, U., M. F. Lindberg, C. Rose, E. Denison, C. Gay, A. Aamodt, J. I. Brox, et al. 2022. Factors correlated with physical function 1 year after total knee arthroplasty

- in patients with knee osteoarthritis: A systematic review and meta-analysis. *JAMA Network Open* 5: e2219636. <https://doi.org/10.1001/jamanetworkopen.2022.19636>.
- . 2023. Factors correlated with pain after total knee arthroplasty: A systematic review and meta-analysis. *PLOS ONE* 18: e0283446. <https://doi.org/10.1371/journal.pone.0283446>.
- Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 6th ser., 2: 559–572. <https://doi.org/10.1080/14786440109462720>.
- Riley, R. D. 2009. Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society, A ser.*, 172: 789–811. <https://doi.org/10.1111/j.1467-985X.2008.00593.x>.
- Riley, R. D., J. R. Thompson, and K. R. Abrams. 2008. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 9: 172–186. <https://doi.org/10.1093/biostatistics/kxm023>.
- Rose, C. J., U. Olsen, M. F. Lindberg, E. M.-L. Denison, A. Aamodt, and A. Lerdal. 2021. A new multivariate meta-analysis model for many variates and few studies. [arXiv:2009.11808 \[stat.ME\]](https://arxiv.org/abs/2009.11808). <https://doi.org/10.48550/arXiv.2009.11808>.
- Rücker, G., and G. Schwarzer. 2015. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology* 15: Article 58. <https://doi.org/10.1186/s12874-015-0060-8>.
- Salanti, G., A. E. Ades, and J. P. A. Ioannidis. 2011. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *Journal of Clinical Epidemiology* 64: 163–171. <https://doi.org/10.1016/j.jclinepi.2010.03.016>.
- Schuurman, N. K., R. P. P. P. Grasman, and E. L. Hamaker. 2016. A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research* 51: 185–206. <https://doi.org/10.1080/00273171.2015.1065398>.
- SAS Institute, Inc. 2016. *SAS/Stat 14.2 User's Guide: The LOGISTIC Procedure*. Cary, NC: SAS Institute.
- Tokuda, T., B. Goodrich, I. Van Mechelen, A. Gelman, and F. Tuerlinckx. 2011. Visualizing distributions of covariance matrices. Technical report, Columbia University, New York.
- White, I. R. 2009. Multivariate random-effects meta-analysis. *Stata Journal* 9: 40–56. <https://doi.org/10.1177/1536867X09000900103>.
- . 2011. Multivariate random-effects meta-regression: Updates to mvmeta. *Stata Journal* 11: 255–270. <https://doi.org/10.1177/1536867X1101100206>.

About the author

Chris Rose is a statistician at the Norwegian Institute of Public Health. His meta-analysis work supports decision making in the Norwegian healthcare and welfare services. He also works on the design and analysis of RCTs of nonpharmaceutical infection control interventions, educational interventions to improve health choices, and digital interventions for maternal and child health. He offers consulting and training within the institute, nationally, and internationally. Previously, he worked on methods for analyzing imaging-based early phase trials of novel cancer therapies and on neoantigen prediction for personalized oncoimmunotherapy.