MULTIVARIATE META-ANALYSIS: USE AND APPLICATIONS

LEARNING OBJECTIVES

- Understand the concept of MVMA and why it can be useful
- Understand data requirements for MVMA
- Understand limitations associated with MVMA
INTRODUCTION TO MULTIVARIATE META-ANALYSIS (MVMA)

META-ANALYSIS: THE STATISTICAL AGGREGATION OF DATA FROM MULTIPLE SOURCES

- Confirmatory data analysis
  - Estimate an average effect even when studies have conflicting results
  - Make predictions of effects in new studies
- Potential for greater ability to extrapolate to general population
- Can examine sources of heterogeneity among study results
- Considered to be high in the evidence hierarchy
Multiple outcomes or treatment effects within studies
- E.g., systolic and diastolic blood pressure
- Multiple time points
- All outcomes measured within the same patients are (usually) correlated
- Yet, we meta-analyze each endpoint separately in separate (univariate) meta-analyses

FREQUENTLY MULTIPLE OUTCOMES ARE OF INTEREST

UNIVARIATE META-ANALYSIS (UMA)
INDIVIDUAL ASSESSMENTS OF DIFFERENT OUTCOMES

Mean treatment effect of Outcome 1

Mean treatment effect of Outcome 2
MULTIVARIATE META-ANALYSIS (MVMA) MULTIPLE OUTCOMES ASSESSED CONCURRENTLY *

Mean treatment effect of Outcome 1

Mean treatment effect of Outcome 2

*Fine print: additional data required

ADVANTAGES OF MVMA

- Utilizes information of correlated outcomes
- Borrows information about missing outcomes from other studies
- Describes the multivariate relationship between endpoints
- Obtains pooled estimates with better statistical properties
- Generates joint confidence regions
- Can model, test and make predictions based on the joint association of endpoints
- Estimates some function of the pooled endpoints
## There are several reasons MVMA is not more common

<table>
<thead>
<tr>
<th>Reason</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tradition</td>
<td>We do it this way, because that’s the way it’s done...</td>
</tr>
<tr>
<td>Increased complexity of multivariate approach</td>
<td>Parsimony is my friend...</td>
</tr>
<tr>
<td>Need for specialized statistical software or knowledge</td>
<td>If it can’t be done in SAS, it can’t be done...</td>
</tr>
<tr>
<td>Lack of appreciation for the consequences of ignoring correlations between endpoints</td>
<td>This workshop can help with that...</td>
</tr>
</tbody>
</table>

---

## A closer look at multivariate meta-analysis

Joseph C. Cappelleri  
Pfizer Inc
Correlation: a mutual relationship or connection between two or more variables

- Conceptually, this means that the plausible range of values of one outcome is dependent on the value of the other

**WHY DO CORRELATIONS MATTER?**

**Correlation:**

- **Observed values span entire range for each variable**
- **Observed values span entire 2D space**

---

**Correlation:**

- **When correlated, each variable spans the entire range**
- **Observed values are related and clustered**
**CORRELATION EXAMPLE**

**DISTRIBUTION OF BMI AND CHOLESTEROL LEVELS IN THE UNITED STATES (1988-1994)**

*Schwartz and Woloshin 1999 Eff Clin Pract (1st 3 figures)*

When BMI and TC are independent, equal probability of having 120 mg/dL and 240 mg/dL

Assume BMI=31

At given BMI, probability distribution of TC is changed

**CORRELATION TYPE 1:** WITHIN-STUDY CORRELATIONS

- Indicates the association between the endpoints within a study
- Arises from each individual in a study contributing data towards each endpoint (or time point)
- Within-study correlations between endpoints are seldom reported in trials
CORRELATION TYPE 2: BETWEEN-STUDY CORRELATIONS

- Indicates how the true underlying endpoint summary values are related across studies
- Is induced by differences across studies such as age, changes in study characteristics, and threshold levels in diagnostic studies
- In diagnostic studies, for example, sensitivity and specificity are usually negatively correlated due to the use of different thresholds (and even for the same threshold)
- Can accurately describe the true bivariate relationship between the true log odds in a treatment group and the true log odds in a control group (baseline risk)
  - Effect modification with baseline risk as proxy for severity of illness

SURGICAL AND NON-SURGICAL TREATMENTS FOR PERIODONTAL DISEASE

- Between-study variance is much larger than within-study variance
- Small within-study variances result in similar UMVA and BVMA estimates

SCHOLASTIC APTITUDE TEST SCORES BETWEEN COACHED AND UNCOACHED STUDENTS

Large between-study variance AND within-study variance

Large variances result in differences between UMVA and BVMA

Math scores Verbal scores

Study 1

Study 7

Correlated Outcomes

EXAMPLE OF DIAGNOSTIC TESTS

- Sensitivity and specificity are measures of the accuracy of the diagnostic test
  - Sensitivity: Pr(test = + | disease = +)
  - Specificity: Pr(test = - | disease = -)

- For every test, there is one pair of sensitivity and specificity values
  - Within-study correlation not important because they are measured in different subgroups (specificity in disease “-” patients and sensitivity in disease “+” patients)

- Between-study correlation is important: sensitivity and specificity are negatively correlated

- Other sources of heterogeneity between studies may also be present
  - Thresholds to define “+” and “-” test results (explicit)
  - Observers, laboratories, or equipment (implicit)

- To pool data from multiple studies
  - One approach: combine data using standard methods for proportions
  - However, ignoring the correlation of the outcomes between studies is inappropriate
CORRELATED OUTCOMES
EXAMPLE OF DIAGNOSTIC TESTS

- BVMA of sensitivity and specificity estimates...
  - Between-study variations (heterogeneity) separately
  - 95% confidence intervals for each outcome
  - Confidence ellipse around the means of sensitivity and specificity
  - Also, the predictive ellipse for sensitivity and specificity for individual studies

- BVMA can also...
  - Obtain functions of sensitivity and specificity such as the diagnostic odds ratio
  - Depict summary receiver operating characteristic (ROC) curve
  - Examine covariates with potentially separate effects on sensitivity and specificity (e.g., for two or more diagnostic technologies)

EXAMPLE: DIAGNOSIS OF LYMPH NODE METASTASES

<table>
<thead>
<tr>
<th>Imaging test</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Diagnostic OR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphangiography (LAG)</td>
<td>0.67 (0.57, 0.76)</td>
<td>0.80 (0.73, 0.85)</td>
<td>8.13 (5.16, 12.82)</td>
</tr>
<tr>
<td>Computed tomography (CT)</td>
<td>0.49 (0.37, 0.61)</td>
<td>0.92 (0.88, 0.95)</td>
<td>11.34 (6.66, 19.30)</td>
</tr>
<tr>
<td>Magnetic resonance imaging (MRI)</td>
<td>0.56 (0.41, 0.70)</td>
<td>0.94 (0.90, 0.97)</td>
<td>21.42 (10.81, 42.45)</td>
</tr>
</tbody>
</table>

Specificity of LAG is significantly lower than that of CT (p=0.0002) and MRI (p=0.0001)

LAG has the highest sensitivity

DOR suggesting MRI is much better than others can be misleading

*Diagnostic OR = Ratio of odds of true "+" relative to odds of false "+"

Reitsma et al. 2005 J Clin Epi
BIVARIATE SUMMARY ESTIMATES AND 95% CONFIDENCE ELLIPSES

For each modality, the region containing the likely combination of the mean values

Confidence ellipses clearly show the differences in sensitivity and specificity of LAG compared with CT and MRI.

Reitsma et al. 2005 J Clin Epi

USING CORRELATIONS IN MVMA

- It allows a joint synthesis of the multiple endpoints
- It accounts for any correlation between endpoints that may exist both within studies and between studies
  - Useful for sample estimates of treatment effect
- Correlations allow a borrowing of strength across endpoints
  - Generally leads to corrected pooled estimates and standard errors
  - Standard errors tend to be lower relative to those from URMA
  - Especially when at least some endpoints are missing at random across studies
  - Narrower joint regions or distributions of effect
- MVMA reverts to separate UVMA of each endpoint when all correlations are zero and one endpoint does not borrow strength from another endpoint
PRACTICAL EXAMPLES: THE BIVARIATE CASE

SAS
- Proc mixed was used in MVMA over a decade ago

WINBUGS
- Markov chain Monte Carlo (MCMC) method

STATA
- `mvmeta`, with options:
  - `wscorr(expression)` – sensitivity analysis over a range of correlations
  - `wscorr(riley)` – fit the Alternative Method as suggested by Riley et al.

**SAS EXAMPLE**

**DIAGNOSTIC TEST**

```sas
proc mixed data=bi_meta method=remi cl;
  class study_id modality;
  model logit = dis*modality non_dis*modality /
    noint cl df=1000, 1000, 1000,
    1000, 1000, 1000;
  random dis non_dis / subject=study_id type=un;
  repeated / group=rec;
  parms / parmsdata=cov
    hold=4 to 91;
  contrast 'CT_sens vs LAG_sens'
    dis*modality 1 -1 0 /df=1000;
run;
```

Obtain estimates of mean sensitivity and specificity for each treatment

Random define between study covariance matrix

Repeat defines within-study covariance matrix

Dataset cov provides the within-study variance

Use contrast statement for testing hypotheses for differences in sensitivities and specificities

---

Reitsma et al. 2005 J Clin Epi

---

**SAS EXAMPLE**

**DIAGNOSTIC TEST**

From the `proc mixed` model, we want to get:
- Means of sensitivity and specificity
- Variance-covariance matrix which combines the between-study and within study variance-covariance

Solution for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>modality</th>
<th>Estimate</th>
<th>Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>dis*modality</td>
<td>1</td>
<td>-0.02090</td>
<td>0.2632</td>
<td>1000</td>
<td>-0.16</td>
<td>0.8721</td>
<td>0.05</td>
<td>-0.5142</td>
<td>0.4352</td>
</tr>
<tr>
<td>dis*modality</td>
<td>2</td>
<td>0.1775</td>
<td>0.2774</td>
<td>1000</td>
<td>0.16</td>
<td>0.8616</td>
<td>0.05</td>
<td>0.2717</td>
<td>1.1641</td>
</tr>
<tr>
<td>dis*modality</td>
<td>3</td>
<td>0.2459</td>
<td>0.3124</td>
<td>1000</td>
<td>0.80</td>
<td>0.4229</td>
<td>0.05</td>
<td>-0.3631</td>
<td>0.8359</td>
</tr>
<tr>
<td>non_dis*modality</td>
<td>1</td>
<td>2.4972</td>
<td>0.2252</td>
<td>1000</td>
<td>10.96</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td>2.0653</td>
<td>2.9389</td>
</tr>
<tr>
<td>non_dis*modality</td>
<td>2</td>
<td>1.3775</td>
<td>0.1678</td>
<td>1000</td>
<td>7.34</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td>1.0955</td>
<td>1.7494</td>
</tr>
<tr>
<td>non_dis*modality</td>
<td>3</td>
<td>-2.8143</td>
<td>0.2046</td>
<td>1000</td>
<td>-9.12</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td>2.2519</td>
<td>3.3767</td>
</tr>
</tbody>
</table>

Means of log sensitivity and specificity

Variance-covariance matrix

---

25

---

26
SAS EXAMPLE
PERIODONTAL DISEASE

```
proc mixed cl method=nl data=berkey;
  class trial type;
  model outcome=pd al
       /nounit s cl;
  random pd al /subject=trial type=un s;
  repeated type /subject=trial
group=trial type=un;
  parms /parmsdata=corrdata
  eqcons=4 to 10;
  • Starting values for between-trial
  variances and covariance
  • Variances within each trial
  • Covariances within each trial
run;
```

Estimate the means of the outcomes
Set up the between-trial variance-covariance matrix
Set up the within-trial variance-covariance matrix

\[
\begin{pmatrix}
\hat{\theta}_{ij} - N \left( \left( \hat{\theta}_d \right) \sum + C_i \right)
\end{pmatrix}
\]

Van Houwelingen et al. 2002 Stat Med

---

SAS EXAMPLE
PERIODONTAL DISEASE

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>pd</td>
<td>0.2448</td>
<td>0.04946</td>
<td>4</td>
<td>6.57</td>
<td>0.05</td>
<td>0.2425</td>
<td>0.4022</td>
<td></td>
</tr>
<tr>
<td>a1</td>
<td>-0.3379</td>
<td>0.07376</td>
<td>4</td>
<td>-4.54</td>
<td>0.05</td>
<td>-0.5256</td>
<td>-0.1585</td>
<td></td>
</tr>
</tbody>
</table>

Means of outcomes 1 and 2

<table>
<thead>
<tr>
<th>Cov Param</th>
<th>Subject</th>
<th>Group</th>
<th>Estimate</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>trial</td>
<td></td>
<td>0.367092</td>
<td>0.05</td>
<td>0.001512</td>
<td>2.3978</td>
</tr>
<tr>
<td>UN(1,1)</td>
<td>trial</td>
<td></td>
<td>0.309365</td>
<td>0.05</td>
<td>-0.01408</td>
<td>0.03606</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>trial</td>
<td></td>
<td>0.036934</td>
<td>0.05</td>
<td>0.000770</td>
<td>0.1306</td>
</tr>
</tbody>
</table>

Variance-covariance matrix
model{  #Random-effects
  for (i in 1:noStudies){
    weight1[i]<-1/var1[i]
    rho[i]<-cov12[i]/sqrt(var1[i]*var2[i])  #Within-study correlation
    m2_prime[i]<-m[i,2]+rho[i]*sqrt(var2[i])/sqrt(var1[i])*y[i,1]-m[i,1]  #Mean of y2 | observed y1
    var2_prime[i]<-var2[i]*(1-pow(rho[i],2))  #Var of y2 | observed y1
    weight2[i]<-1/var2_prime[i]
  }
  for (i in 1:noStudies) {
    m[i,1]~dnorm(mean[1],prec[1])
    m[i,2]~dnorm(m.re[i],prec.vre)
    m.re[i]<-mean[2]+rho.tau*sqrt(tau.sq[2])/sqrt(tau.sq[1])*y[i,1]-m[i,1]  #Mean of effect 2 | effect 1
    v.re<tau.sq[2]*(1-pow(rho.tau,2))  #Var of effect 2 | effect 1
  }
  Prior specifications...  
}

Mavridis and Salanti 2013 Stat Methods Med Res
WINBUGS EXAMPLE
TREATMENTS FOR PERIODONTAL DISEASE
KNOWN COVARIANCES

```r
model{
  # Random-effects
  for (i in 1:noStudies){
    weight1[i]<-1/var1[i]
    rho[i]<-cov12[i]/sqrt(var1[i]*var2[i]) # Within-study correlation
    y[i,1]~dnorm(m[i,1],weight1[i])
    y[i,2]~dnorm(m2_prime[i],weight2[i])
    m2_prime[i]<-m[i,2]+rho[i]*#Mean of y2 | observed y1
      sqrt(var2[i])/sqrt(var1[i])*(y[i,1]-m[i,1])
    var2_prime[i]<-var2[i]*(1-pow(rho[i],2)) #Var of y2 | observed y1
    weight2[i]<-1/var2_prime[i]
  }
  for (i in 1:noStudies) {
    m[i,1]~dnorm(mean[1],prec[1])
    m[i,2]~dnorm(m.re[i],prec.vre)
    m.re[i]<-mean[2]+rho.tau*sqrt(tau.sq[2])/sqrt(tau.sq[1])*#Mean of effect 2 | effect 1
      (m[i,1]-mean[1]) #Mean of effect 2 | effect 1
    v.re<-tau.sq[2]*(1-pow(rho.tau,2)) #Var of effect 2 | effect 1
  }
  Prior specifications...
}
```

Mavridis and Salanti 2013 Stat Methods Med Res

---

WINBUGS EXAMPLE
TREATMENTS FOR PERIODONTAL DISEASE
UNKNOWN COVARIANCE

```r
model{
  # Random-effects
  for (i in 1:noStudies){
    weight1[i]<-1/var1[i]
    rho[i]~dbeta(1,1) # Uniform prior on [0,1]
    y[i,1]~dnorm(m[i,1],weight1[i])
    y[i,2]~dnorm(m2_prime[i],weight2[i])
    m2_prime[i]<-m[i,2]+rho[i]*#Mean of y2 | observed y1
      sqrt(var2[i])/sqrt(var1[i])*(y[i,1]-m[i,1])
    var2_prime[i]<-var2[i]*(1-pow(rho,2)) #Var of y2 | observed y1
    weight2[i]<-1/var2_prime[i]
  }
  for (i in 1:noStudies) {
    m[i,1]~dnorm(mean[1],prec[1])
    m[i,2]~dnorm(m.re[i],prec.vre)
    m.re[i]<-mean[2]+rho.tau*sqrt(tau.sq[2])/sqrt(tau.sq[1])*#Mean of effect 2 | effect 1
      (m[i,1]-mean[1]) #Mean of effect 2 | effect 1
    v.re<-tau.sq[2]*(1-pow(rho.tau,2)) #Var of effect 2 | effect 1
  }
  Prior specifications...
}
```

Mavridis and Salanti 2013 Stat Methods Med Res
### PRACTICAL SUMMARY

<table>
<thead>
<tr>
<th>SAS Proc Mixed</th>
<th>WinBUGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenient</td>
<td>Requires knowledge of distribution functions (e.g. prior, likelihood, etc.)</td>
</tr>
<tr>
<td>Flexibility in covariance structure</td>
<td>NA</td>
</tr>
<tr>
<td>Requires within-study correlations to be known</td>
<td>Does not require known within-study correlation estimates</td>
</tr>
<tr>
<td>Provides estimates of population means and between-trial covariance matrix</td>
<td>Provides posterior distributions for all parameters</td>
</tr>
<tr>
<td>Larger standard deviations with few studies</td>
<td>Different prior specifications can lead to different results, especially when there are only few studies</td>
</tr>
</tbody>
</table>

---

### MVMA LIMITATIONS AND OTHER CONSIDERATIONS

Or the harbinger of gloom

Source: [http://britcomgreats.blogspot.co.uk/2012/11/one-foot-in-grave.html](http://britcomgreats.blogspot.co.uk/2012/11/one-foot-in-grave.html)

---

**John W. Stevens**  
ScHARR, University of Sheffield
TWO PERSPECTIVES

One view:

“In the realm of research synthesis . . . the consequences of accounting for (modeling) dependence or ignoring it are not well understood” (Becker et al., 2004).

Another view:

“Psychologist: one who thinks no univariate data are Normally distributed but that all multivariate data are.” Senn, S

“Multivariate analysis will bring you many dimensions but you will not enjoy them.” Senn, S

“Multivariate analysis: such pretty pictures.” Senn, S

BACK TO BASICS (1)

- Univariate data can be analysed exactly according to an appropriate likelihood function for each treatment arm
  - E.g. a Binomial distribution for the number of observed events out of the number of patient randomised
- Multivariate meta-analyses generally resort to analysing study-specific treatment effects
  - E.g. the sample log odds ratio depending on the estimate of the variance of the log odds ratio
- Study-specific treatment effects are then assumed to be (multivariate) normally distributed
The performance associated with study-specific treatment effects will depend on:
- Sample size (i.e. Central Limit Theorem)
- How close the probability of an event is to 0 or 1
- The presence of zero values

Comparative effectiveness
- In this case the main question may be:
  “Is there evidence of a treatment effect and can the treatment work in patients not included in the study?”
- Parameter estimates may be only slightly changed unless there is substantial missing data
- The effort involved in a MVMA may not be worthwhile
Health technology assessment

In this case the objective may be to characterise the joint distribution for absolute treatment effects such as:

- The mean systolic and diastolic blood pressure on each treatment arm
- The proportion of patients in each ACR and EULAR category on each treatment arm
- The mean time to disease progression and death on each treatment arm

Allowing for correlation between parameters can substantially reduce uncertainty in output parameters (Briggs et al., 2006)

Characterising the joint uncertainty about input parameters and propagating the uncertainty through to the output may affect ICERs and expected net benefit, particularly in non-linear models

PROBLEMS WITH WITHIN-STUDY CORRELATION

Estimating within-study correlations would not be an issue if individual patient-level data were available for each study

In practice, this is unrealistic and we generally deal with published data for some or all studies

When within-study correlations are unavailable solutions include:

- Impute within-study correlations from available studies (ignores uncertainty)
- Perform sensitivity analyses over a range of assume values for the within-study correlations (computer intensive)
- Use a Bayesian approach (i.e. multiple imputation) based on an informative prior distribution for the within-study correlations using external data or expert opinion (coherent and transparent)
OTHER LIMITATIONS

- **Publication bias**
  - If the objective is to correlate primary and secondary outcomes and secondary outcomes are not missing at random then this may bias both primary and secondary effects.

- **Ordered categorical data**
  - We can correlate treatment effects from multiple ordered categorical data outcomes but it is not clear (to me) how to use the estimates from a MVMA to generate absolute treatment effects.

- **Time-to-event**
  - We can correlate (log) hazard ratios for multiple outcomes such as OS and PFS assuming that hazards are proportional, else absolute effects such as mean PFS and OS will be biased (as in the univariate case).

SUMMARY

- “Except when within-study variation is very small relative to between-study variation, the within-study correlation plays a crucial role in the amount or borrowing of strength.” Riley RD, 2009
- The question being addressed affects the appropriateness of the scale of measurement (as with univariate analyses)
- An important feature of multivariate meta-analyses is dealing with within-study correlation when they are unknown
- Diagnostic test accuracy is a special case of a multivariate meta-analysis in which sensitivity and specificity are correlated between studies but there is no within-study correlation.
REFERENCES

- Gleser and Olkin. 1994 Stochastically dependent effect sizes. In *The Handbook of Research Syntheses* (eds Cooper and Hedges)