The Meta-analytic Framework for the Evaluation of Surrogate Endpoints in Clinical Trials

Geert Molenberghs

Center for Statistics
Universiteit Hasselt, Belgium
geert.molenberghs@uhasselt.be
www.censtat.uhasselt.be

Biostatistical Centre
Katholieke Universiteit Leuven, Belgium
geert.molenberghs@med.kuleuven.be
www.kuleuven.ac.be/biostat/

universiteit hasselt
CENTRUM VOOR STATISTIEK

Non-clinical Statistics Conference, September 24, 2008

LEUVEN

Motivation

- **Primary motivation**
  - ▷ True endpoint is rare and/or distant
  - ▷ Surrogate endpoint is frequent and/or close in time

- **Secondary motivation**: True endpoint is
  - ▷ invasive
  - ▷ uncomfortable
  - ▷ costly
  - ▷ confounded by secondary treatments and/or competing risks

Definitions

Clinical Endpoint:

A characteristic or variable that reflects how a patient feels, functions, or survives.

Biomarker:

A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

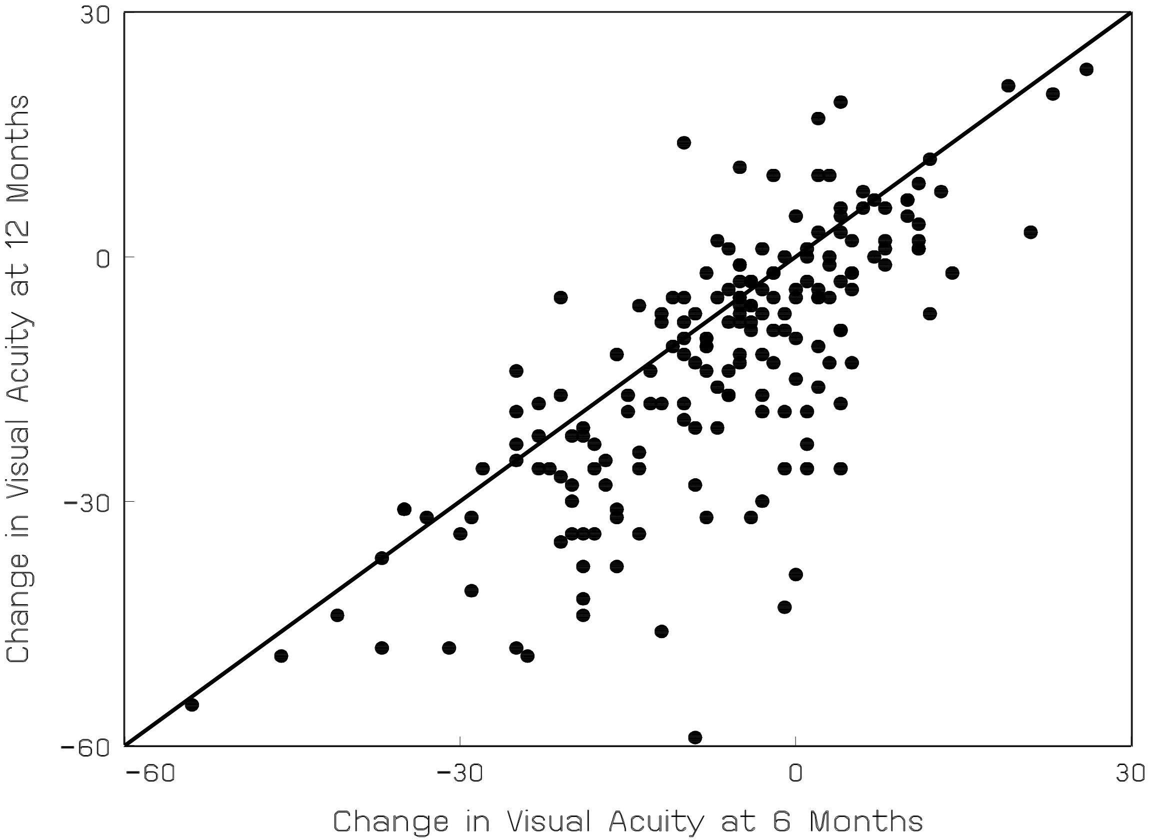
Surrogate Endpoint:

A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm).

Biomarkers Definition Working Group (Clin Pharmacol Ther 2001)

Age-Related Macular Degeneration

Pharmacological Therapy for Macular Degeneration Study Group (1997)



Z: **Interferon-**α

S: **Visual acuity at 6 months**

T: **Visual acuity at 1 year**

N: 190 patients in 36 centers (# patients/center $\in$[2;18])

Definition and Single-Unit Model

Prentice (Bcs 1989)

"A test of H_0 of no effect of treatment on surrogate is equivalent to a test of H_0 of no effect of treatment on true endpoint."

$$\begin{aligned} S_j &= \mu_S + \alpha Z_j + \varepsilon_{Sj} \\ T_j &= \mu_T + \beta Z_j + \varepsilon_{Tj} \end{aligned} \qquad \Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{ST} \end{pmatrix}$$

$$T_j = \mu + \gamma S_j + \varepsilon_j$$

Prentice's Criteria and Measures

Prentice (1989), Freedman *et al* (1992)

Quantity	Estimate	Test
1 Effect of Z on T	β	$(T\|Z) \neq (T)$
2 Effect of Z on S	α	$(S\|Z) \neq (S)$
3 Effect of S on T	γ	$(T\|S) \neq (T)$
4 Effect of Z on T, given S	β_S	$(T\|Z,S) = (T\|S)$

$$\downarrow$$

Proportion Explained

$$PE = \frac{\beta - \beta_S}{\beta}$$

$$\swarrow \qquad\qquad \searrow$$

Relative Effect

$$RE = \frac{\beta}{\alpha}$$

Adjusted Association

$$\rho_Z = \text{Corr}(S, T|Z)$$

Prentice's Criteria and Measures

Prentice (1989), Freedman *et al* (1992)

	Quantity	Estimate	Test
1	Effect of Z on T	$\widehat{\beta} = 4.12(2.32)$	$p = 0.079$
2	Effect of Z on S	$\widehat{\alpha} = 2.83(1.86)$	$p = 0.13$
3	Effect of S on T	$\widehat{\gamma} = 0.95(0.06)$	$p < 0.0001$
4	Effect of Z on T, given S	$\widehat{\beta_S}$	

$$\downarrow$$

Proportion Explained

$$\widehat{PE} = 0.65 \quad [-0.22; 1.51]$$

$$\swarrow \qquad\qquad\qquad \searrow$$

Relative Effect

$$\widehat{RE} = 1.45 \quad [-0.48; 3.39]$$

Adjusted Association

$$\widehat{\rho}_Z = 0.75 \quad [0.69; 0.82]$$

Relationship and Problems

$$\begin{aligned} RE &= \frac{\beta}{\alpha} \\ \rho_Z &= \frac{\sigma_{ST}}{\sqrt{\sigma_{SS}\sigma_{TT}}} \\ PE &= \lambda \cdot \rho_Z \cdot \frac{\alpha}{\beta} = \lambda \cdot \rho_Z \cdot \frac{1}{RE} \end{aligned}$$

where

$$\lambda^2 = \frac{\sigma_{TT}}{\sigma_{SS}}$$

- Very wide confidence intervals for PE
- $PE \notin [0, 1]$

Use of Relative Effect and Adjusted Association

- The two new quantities have clear meaning

  ▷ **Relative Effect:** trial-level measure of surrogacy

  Can we translate the treatment effect on the surrogate to the treatment effect on the endpoint, in a sufficiently precise way*?*

  ▷ **Adjusted Association:** individual-level measure of surrogacy

  After accounting for the treatment effect, is the surrogate endpoint predictive for a patient's true endpoint?

- **BUT:**

  The RE is based on a single trial $\Rightarrow$ regression through the origin, based on one point!

Analysis Based on Several Trials. . .

- **Context:**
  - ▷ multicenter trials
  - ▷ meta analysis
  - ▷ several meta-analyses

- **Extensions:**
  - ▷ **Relative Effect** $\longrightarrow$ **Trial-Level Surrogacy**

    How close is the relationship between the treatment effects on the surrogate and true endpoints, based on the various trials (units)?
  - ▷ **Adjusted Association** $\longrightarrow$ **Individual-Level Surrogacy**

    How close is the relationship between the surrogate and true outcome, after accounting for trial and treatment effects?

. . . Is Considered a Useful Idea

Albert *et al* (SiM 1998)

"There has been little work on alternative statistical approaches. A meta-analysis approach seems desirable to reduce variability. Nevertheless, we need to resolve basic problems in the interpretation of measures of surrogacy such as PE as well as questions about the biologic mechanisms of drug action."

Statistical Model

- **Model:**

$$
\begin{aligned}
S_{ij} &= \mu_{Si} + \alpha_i Z_{ij} + \varepsilon_{Sij} \\
T_{ij} &= \mu_{Ti} + \beta_i Z_{ij} + \varepsilon_{Tij}
\end{aligned}
$$

- **Error structure:**

$$
\Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix}
$$

Statistical Model

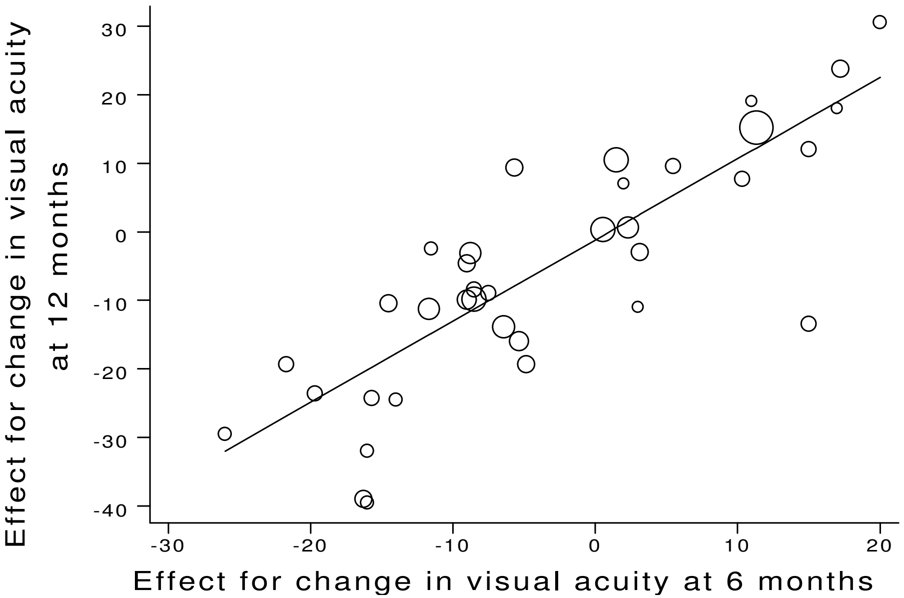
- **Model:**

$$\begin{aligned} S_{ij} &= \mu_{Si} + \alpha_i Z_{ij} + \varepsilon_{Sij} \\ T_{ij} &= \mu_{Ti} + \beta_i Z_{ij} + \varepsilon_{Tij} \end{aligned}$$

- **Trial-specific effects:**

$$\begin{pmatrix} \mu_{Si} \\ \mu_{Ti} \\ \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_T \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} m_{Si} \\ m_{Ti} \\ a_i \\ b_i \end{pmatrix} \qquad D = \begin{pmatrix} d_{SS} & d_{ST} & d_{Sa} & d_{Sb} \\ & d_{TT} & d_{Ta} & d_{Tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix}$$

ARMD: Trial-Level Surrogacy



- **Prediction:**
  - ▷ *What do we expect ?*

    $$E(\beta + b_0 | m_{S0}, a_0)$$

  - ▷ *How precisely can we estimate it ?*

    $$\text{Var}(\beta + b_0 | m_{S0}, a_0)$$

- **Estimate:**
  - ▷ $R^2_{\text{trial}} = 0.692$ (95% C.I. $[0.52; 0.86]$)

ARMD: Individual-Level Surrogacy

- **Individual-level association:**

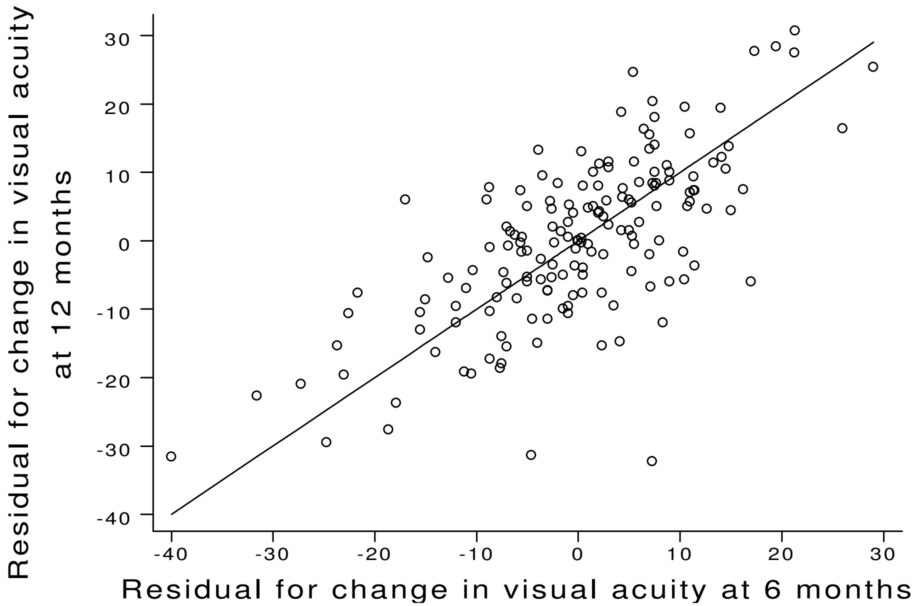
  $\rho_Z = R_{\text{indiv}} = \text{Corr}(\varepsilon_{Ti}, \varepsilon_{Si})$

- **Estimate:**

  ▷ $R^2_{\text{indiv}} = 0.483$ (95% C.I. $[0.38; 0.59]$)

  ▷ $R_{\text{indiv}} = 0.69$ (95% C.I. $[0.62; 0.77]$)

  ▷ Recall $\rho_Z = 0.75$ (95% C.I. $[0.69; 0.82]$)



A Number of Case Studies

	Age-related macular degeneration	Advanced ovarian cancer	Advanced colorectal cancer
Surrogate **True**	Vis. Ac. (6 months) Vis. Ac. (1 year)	Progr.-free surv. Overall surv.	Progr.-free surv. Overall surv.
Prentice Criteria 1–3 (p value)			
Association (Z, S)	0.31	0.013	0.90
Association (Z, T)	0.22	0.08	0.86
Association (S, T)	< 0.001	< 0.001	< 0.001
Single-Unit Validation Measures (estimate and 95% C.I.)			
Proportion Explained	$0.61[-0.19; 1.41]$	$1.34[0.73; 1.95]$	$0.51[-4.97; 5.99]$
Relative Effect	$1.51[-0.46; 3.49]$	$0.65[0.36; 0.95]$	$1.59[-15.49, 18.67]$
Adjusted Association	$0.74[0.68; 0.81]$	$0.94[0.94; 0.95]$	$0.73[0.70, 0.76]$
Multiple-Unit Validation Measures (estimate and 95% C.I.)			
R^2_{trial}	$0.69[0.52; 0.86]$	$0.94[0.91; 0.97]$	$0.57[0.41, 0.72]$
R^2_{indiv}	$0.48[0.38; 0.59]$	$0.89[0.87; 0.90]$	$0.57[0.52, 0.62]$

Overview: Case Studies

	Schizoph. Study I (138 units)	Schizoph. Study I (29 units)	Schizoph. Study II
Surrogate	— PANSS —		
True	— CGI —		
Prentice Criteria 1–3 (p value)			
Association (Z, S)	0.016		0.835
Association (Z, T)	0.007		0.792
Association (S, T)	< 0.001		< 0.001
Single-Unit Validation Measures (estimate and 95% C.I.)			
Proportion Explained	$0.81[0.46; 1.67]$		$-0.94[\infty]$
Relative Effect	$0.055[0.01; 0.16]$		$-0.03[\infty]$
Adjusted Association	$0.72[0.69; 0.75]$		$0.74[0.69; 0.79]$
Multiple-Unit Validation Measures (estimate and 95% C.I.)			
R^2_{trial}	$0.56[0.43; 0.68]$	$0.58[0.45; 0.71]$	$0.70[0.44; 0.96]$
R^2_{indiv}	$0.51[0.47; 0.55]$	$0.52[0.48; 0.56]$	$0.55[0.47; 0.62]$

Two Longitudinal Endpoints

First Stage

$$\begin{aligned} T_{ijt} &= \mu_{T_i} + \beta_i Z_{ij} + \theta_{T_i} t_{ijt} + \varepsilon_{T_{ijt}} \\ S_{ijt} &= \mu_{S_i} + \alpha_i Z_{ij} + \theta_{S_i} t_{ijt} + \varepsilon_{S_{ijt}} \end{aligned}$$

$$\Sigma_i = \begin{pmatrix} \sigma_{TTi} & \sigma_{STi} \\ \sigma_{STi} & \sigma_{SSi} \end{pmatrix} \otimes R_i$$

Second Stage

$$\begin{pmatrix} \mu_{S_i} \\ \mu_{T_i} \\ \alpha_i \\ \beta_i \\ \theta_{S_i} \\ \theta_{T_i} \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_T \\ \alpha \\ \beta \\ \theta_S \\ \theta_T \end{pmatrix} + \begin{pmatrix} m_{S_i} \\ m_{T_i} \\ a_i \\ b_i \\ \tau_{S_i} \\ \tau_{T_i} \end{pmatrix}$$

Evaluation Measures?

A Sequence of Measures

- **Variance Reduction Factor VRF:**

$$VRF = \frac{\Sigma_i \{\text{tr}(\Sigma_{TTi}) - \text{tr}(\Sigma_{(T|S)i})\}}{\Sigma_i \, \text{tr}(\Sigma_{TTi})}$$

- **Canonical-correlation Root-statistic Based Measure θ_p:**

$$\theta_p = \sum_i \frac{1}{Np_i} \text{tr}\left\{\left(\Sigma_{TTi} - \Sigma_{(T|S)i}\right) \Sigma_{TTi}^{-1}\right\}$$

- **Canonical-correlation Root-statistic Based Measure R_Λ^2:**

$$R_\Lambda^2 = \frac{1}{N} \sum_i (1 - \Lambda_i),$$

where

$$\Lambda_i = \frac{|\Sigma_i|}{|\Sigma_{TTi}| \, |\Sigma_{SSi}|}$$

A Sequence of Measures

- **The Likelihood Reduction Factor LRF:**

  ▷ Consider a pair of models:

$$g_T(T_{ij}) = \mu_{T_i} + \beta_i Z_{ij}$$

$$g_T(T_{ij}) = \theta_{0_i} + \theta_{1i} Z_{ij} + \theta_{2i} S_{ij}$$

  ▷ G_i^2 log-likelihood ratio for comparison of both models

  ▷ The proposed measure:

$$\text{LRF} = 1 - \frac{1}{N} \sum_i \exp\left(-\frac{G_i^2}{n_i}\right)$$

An Information-theoretic Approach

- Can we unify all previous proposals?

- Shannon (1916–2001) defined entropy of a distribution:

$$h(Y) = E[-\log(f(Y))]$$

- Conditional version:

$$h(Y|X=x) = E_{Y|X}[\log f_{Y|X}(Y|X=x)] \qquad \text{and} \qquad I(Y|X) = E_X[h(Y|X=x)]$$

- The amount of uncertainty (entropy) that is expected to be removed if the value of X is known:

$$I(X,y) = h(Y) - h(Y|X)$$

An Information-theoretic Approach

- Informational measure of association R_h^2:

$$R_h^2 = R_h^2 = \frac{EP(Y) - EP(Y|X)}{EP(Y)}$$

with

$$EP(X) = \frac{1}{(2\pi e)^n} e^{2h(X)}$$

- Version for N trials:

$$R_h^2 = \sum_{i=1}^{N_q} \alpha_i R_{hi}^2 = 1 - \sum_{i=1}^{N_q} \alpha_i e^{-2I_i(S_i,T_i)},$$

where the α_i form a convex combination.

Relationships With Previous Definitions

- All have desirable behavior within $[0, 1]$ for continuous endpoints

- All can be embedded within a family

- θ_p is symmetric in S and T whereas the VRF is not

- θ_p is invariant w.r.t. linear bijective transformations; VRF only when they are orthogonal

- R^2_Λ and later ones also apply to non-Gaussian settings

Relationships With Previous Definitions

- Later ones specialize to earlier ones

- They all reduce to the R^2_{indiv} for cross-sectional Gaussian outcomes

- Longitudinal normal setting:

$$R^2_h = R^2_\Lambda \qquad \text{if} \qquad \alpha_i = N_q^{-1}$$

- General setting:

$$\text{LRF} \xrightarrow{P} R^2_h$$

  when the number of subjects per trial approaches ∞

Other Implications

- Relationship with Prentice's main criterion and the Data Processing Inequality:

$$
\begin{aligned}
f(T|Z,S) = F(T|S) \quad &\Rightarrow \quad Z \rightarrow S \rightarrow T \\
&\Rightarrow \quad I(T,Z|S) = 0 \\
&\Rightarrow \quad I(Z,S) \geq I(Z,T)
\end{aligned}
$$

- PE and R_h^2:

$$
\mathsf{PE} = 1 - \frac{\beta_S}{\beta} \qquad \longleftrightarrow \qquad R_h^2 = 1 - \frac{\mathsf{EP}(\beta_i|\alpha_i)}{\mathsf{EP}(\beta_i)}
$$

Fano's Inequality

- Fano's Inequality:

$$E\left[(T-g(S))^2\right] \quad \geq \quad EP(T)(1-R_h^2)$$

  ▷ Left hand side is prediction error

  ▷ Applies regardless of distributional form and predictor function $g(\cdot)$

  ▷ **"How large does R_h^2 have to be?"** $\longleftarrow$ The answer depend crucially on the power entropy of T

Schizophrenia Trial

- **Continuous Outcomes:**
  - ▷ $VRF_{\text{ind}} = 0.39$ with 95% C.I. $[0.36; 0.41]$
  - ▷ $R^2_{\text{trial}} = 0.85$ with 95% C.I. $[0.68; 0.95]$
- **Binary Outcomes:**

Parameter	Estimate	95% C.I.
Trial-level R^2_{trial} measures		
Information-theoretic	0.49	[0.21,0.81]
Probit	0.51	[0.18,0.78]
Plackett-Dale	0.51	[0.21,0.81]
Individual-level measures		
R^2_h	0.27	[0.24,0.33]
R^2_hmax	0.39	[0.35,0.48]
Probit	0.67	[0.55,0.76]
Plackett-Dale ψ	25.12	[14.66;43.02]
Fano's lower-bound	0.08	

Age-related Macular Degeneration Trial

- **Both outcomes binary:**

Parameter	Estimate	[95% C.I.]
R^2_{trial}	0.3845	[0.1494;0.6144]
R^2_h	0.2648	[0.2213;0.3705]
$R^2_h\text{max}$	0.4955	[0.3252;0.6044]

Advanced Colorectal Cancer

S: Time to progression/death

T: Time to death

- Models:

$$h_{ij}(t) = h_{i0}(t)\text{exp}\{\beta_i Z_{ij}\}$$

$$h_{ij}(t) = h_{i0}(t)\text{exp}\{\beta_{Si} Z_{ij} + \gamma_i S_{ij}(t)\}$$

Advanced Colorectal Cancer

	Estimate (95% C.I.)	
Parameter	**Dataset I**	**Dataset II**
Trial-level measures		
$\hat{R}^2_{\text{trial}}$ (separate models)	0.82 [0.40;0.95]	0.85 [0.53;0.96]
$\hat{R}^2_{\text{trial}}$ (Clayton copula)	0.88 [0.59;0.98]	0.82 [0.43;0.95]
$\hat{R}^2_{\text{trial}}$ (Hougaard copula)		0.75 [0.00;1.00]
Individual-level measures		
$\hat{R}^2_h$	0.84 [0.82;0.85]	0.83 [0.82;0.85]
Percentage of censoring	19%	55%

Prediction in a New Trial

- Consider a new trial $i = 0$:

$$S_{0j} = \mu_{S0} + \alpha_0 Z_{0j} + \varepsilon_{S0j}$$

- **Prediction variance:**

$$\text{Var}(\beta + b_0 | \mu_{S0}, \alpha_0, \vartheta) \approx f\{\text{Var}(\widehat{\mu}_{S0}, \widehat{\alpha}_0)\} + f\{\text{Var}(\widehat{\vartheta})\} + (1 - R^2_{\text{trial}})\text{Var}(b_0)$$

- where
  - ▷ $f(\cdot)$ are appropriate functions of the parameters involved
  - ▷ ϑ contains all fixed effects

Prediction in a New Trial

- Meaning of the three terms:
  - ▷ **Estimation error in both the meta-analysis and the new trial:**

    all three terms apply

  - ▷ **Estimation error in the meta-analysis only:**

$$\mathsf{Var}(\beta + b_0 | \mu_{S0}, \alpha_0, \vartheta) \approx f\{\mathsf{Var}(\widehat{\vartheta})\} + (1 - R^2_{\text{trial}})\mathsf{Var}(b_0)$$

  - ▷ **No estimation error:**

$$\mathsf{Var}(\beta + b_0 | m_{S0}, a_0) = (1 - R^2_{\text{trial}})\mathsf{Var}(b_0)$$

The Surrogate Threshold Effect

- **STE:** The smallest treatment effect upon the surrogate that predicts a significant treatment effect on the true endpoint

- *Various versions:*

  ▷ STE$_{N,n}$: STE for a finite meta-analysis and a finite new trial

  ▷ STE$_{N,\infty}$: STE for a finite meta-analysis and an infinite new trial

  ▷ STE$_{\infty,\infty}$: STE when both the meta-analysis and the new trial are infinitely large

Practical Conclusions

- *Are surrogate endpoints useful in practice?*

- An investigator wants to be able to predict the effect of treatment on T, based on the observed effect of treatment on S.

- R^2_{trial}, R^2_{indiv}, (ψ, τ), VRF, θ_p, R^2_Λ LRF, R^2_h, $\ldots$: quantification of surrogacy in a meta-analytic setting

- Prediction: useful in a *new* trial

Methodological Conclusions

- **Basis for new assessment strategy**
  - ▷ trial-level surrogacy
  - ▷ individual-level surrogacy
- **Requirements**
  - ▷ Was required: joint model for surrogate and true endpoint
  - ▷ Was required: acknowledgment of the hierarchical structure
  - ▷ Matters simplify with information-theoretic approach