The Statistical Evaluation of Surrogate Endpoints in Clinical Trials – And Beyond

Geert Molenberghs

geert.molenberghs@uhasselt.be & geert.molenberghs@kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

Universiteit Hasselt & KU Leuven, Belgium

www.ibiostat.be

ESALQ, Piracicaba, Brazil

November 6, 2025

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

Motivation: Duration and Size

	True endpoint trial			Surrogate endpoint trial		
Event	**Endpoint**	**Size**	**Length**	**Endpoint**	**Size**	**Length**
MI	Death	4000	5 yrs	Cor. art. patency	200	90 min
MI	Death	4000	5 yrs	Eject. frac.	30	2-4 wks
Stroke	Stroke	25000	5 yrs	DBP	200	1-2 yrs

Wittes, Lakatos, and Probstfield (SiM 1989)

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)

Note: *A surrogate does not need to be a bio*marker.

Examples of Biomarkers in Oncology

Disease	Biomarker	Endpoint
Ovarian cancer	CA-125	Survival
Resected colorectal cancer	CEA	Time to recurrence
Germ-cell malgnancies	AFP	Survival
Gastrointestinal stromal tumors	PET scan	Survival
Hormone-dependent prostate cancer	PSA	Time to progression
Advanced prostate cancer	PSA	Survival

Candidate Surrogate Endpoints?

Disease	Surrogate	Type	True Endpoint	Type
Advanced cancer	Tumor response	Discr.	Survival	Surv.
Osteoporosis	BMD	Longit.	Fracture	Bin.
Cardiovascular	Ejection fraction	Cont.	MI	Bin.
Hypertension	Blood pressure	Cont.	Coronary HD	Bin.
Arrhythmias	Arrhythmias	Longit.	Survival	Surv.
HIV infection	CD4 counts	Longit.	AIDS	Surv.
AIDS	Viral load	Longit.	Survival	Surv.
Ophthalmology	Intraoccular press.	Cont.	Visual acuity	Cont.
Depression	Biomarkers	Cont.	Depression	Cont.

Candidate Surrogate Endpoints?

Disease	Surrogate	Type	True Endpoint	Type
Alzheimer	Amyloid	Cont.	CDR-SB	Cont.
Alzheimer	Tau	Cont.	CDR-SB	Cont.
Alzheimer	Amyloid & Tau	Cont.	CDR-SB	Cont.
Alzheimer	Other cogn. scale	Cont.	CDR-SB	Cont.
Alzheimer	Amyloid and/or Tau	Cont.	Various scales	Cont.
Alzheimer	⋮	Cont.	⋮	Cont.

Bad Precedents

Fleming and Demets (Ann Intern Med 1996)

N Engl J Med (1989, 306)

N Engl J Med (1991, 324)

False positive: Encainide and flecainide reduced the incidence of arrhythmias. These drugs were approved by FDA and an estimated 500,000 patients took them yearly in the US. The Cardiac Arrhythmia Suppression Trial (CAST) showed a 3-fold *increase* in death rate with anti-arrhythmic drugs!

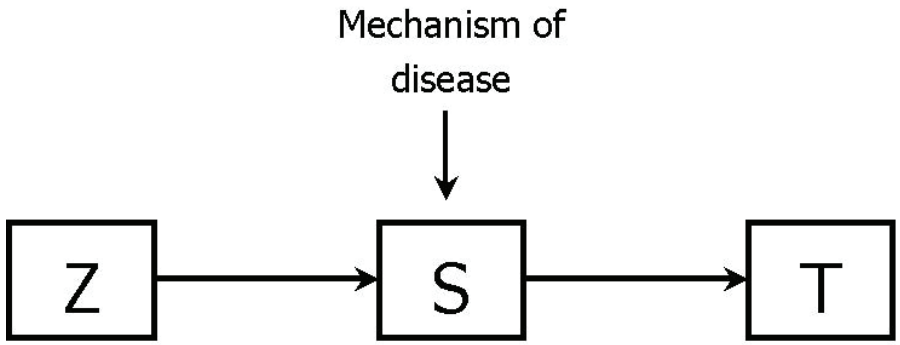
False negative: A trial in Chronic Granulomatous Disease showed no effect of γ-interferon on bacterial killing or superoxide production. Yet there was a 3-fold *decrease* in the rate of recurrent serious infections.

Notation

Z: Treatment

S: Surrogate endpoint

T: True (or "final") endpoint

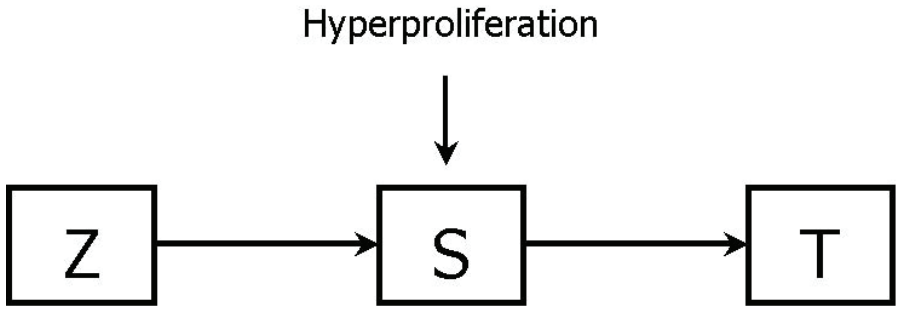


Example

Z: Dietary changes

S: Colorectal polyps

T: Colorectal adenocarcinomas



Schatzkin and Gail (Nature Reviews (Cancer) 2001)

Biological Concern

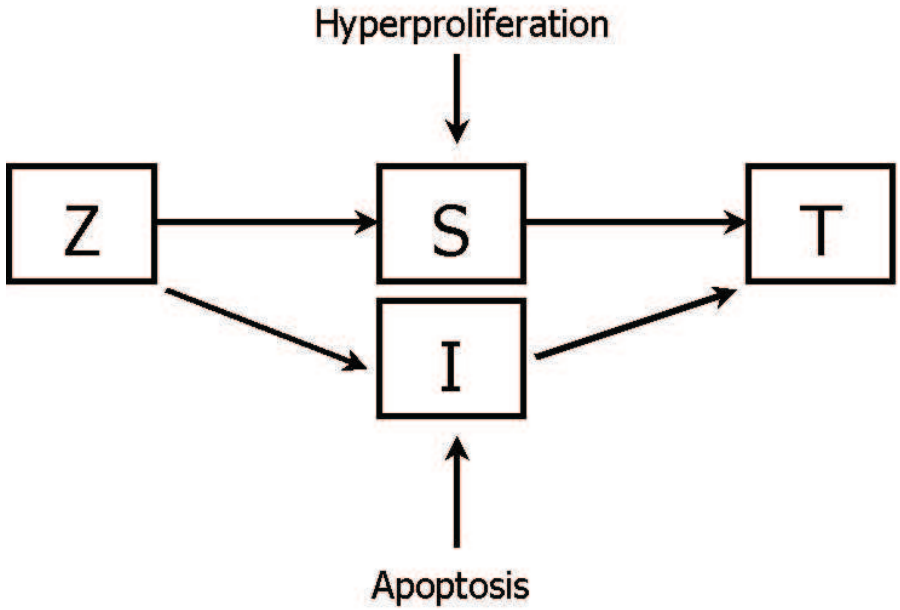
Z: Dietary changes

I: Intermediate step

S: Colorectal polyps

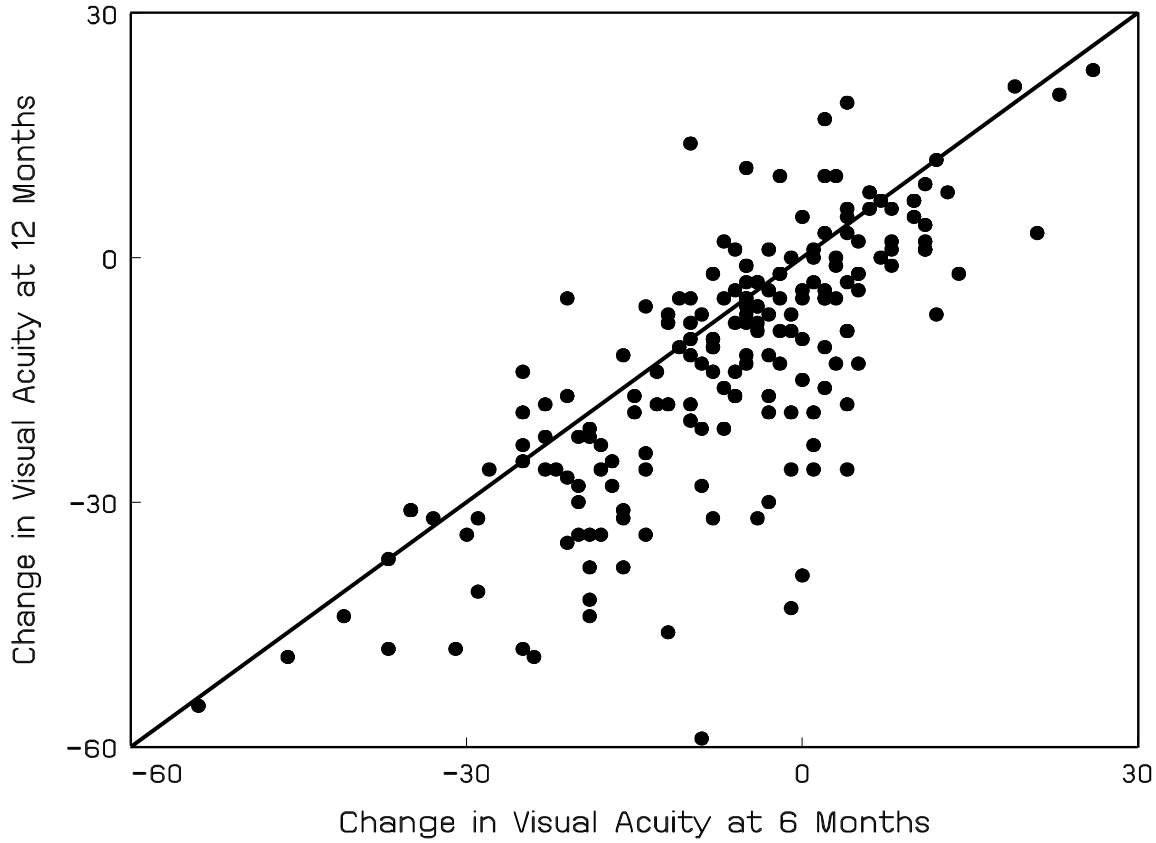
T: Colorectal adenocarcinomas

The final endpoint may be affected through several mechanisms, some of which do not involve the surrogate endpoint.



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])

Visual Acuity

V A L I D

A T I O N

O F S U R

R O G A T

E M A R K

E R S I N

R A N D O

M I Z E D

E X P E R

I M E N T

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 Definition and Criteria

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."

1. Treatment has significant effect on Surrogate
2. Treatment has significant effect on True Endpoint
3. Surrogate has significant effect on True Endpoint
4. Treatment is no longer significant on True Endpoint **given Surrogate**

Prentice's Definition and Criteria

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."

Crit. 1	$Z_j \longrightarrow T_j$	β
Crit. 2	$Z_j \longrightarrow S_j$	α
Crit. 3	$S_j \longrightarrow T_j$	γ
The Crit.	$Z_j \longrightarrow T_j \mid S_j$	β_S

From Prentice to Measures

Prentice (1989), Freedman *et al* (1992), Buyse and Molenberghs (1998)

Crit. 1	$Z_j \longrightarrow T_j$	β
Crit. 2	$Z_j \longrightarrow S_j$	α
Crit. 3	$S_j \longrightarrow T_j$	γ
The Crit.	$Z_j \longrightarrow T_j \mid S_j$	β_S

$\downarrow$

Proportion Explained

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

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

Relative Effect $\qquad$ **Adjusted Association**

$$\widehat{RE} = \frac{\beta}{\alpha} \qquad \widehat{\rho}_Z = \mathsf{Corr}(S_j, T_j | Z_j)$$

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

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

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

Relative Effect

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

Adjusted Association

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

Problems

- The **Proportion Explained** is not well defined:
  - ▷ It is not restricted to $[0, 1]$
  - ▷ It is sensitive to immaterial scale factors
  - ▷ Even when **Relative Effect** and **Adjusted Association** are perfect, **PE** is not necessarily equal to 1
- On top of this: the **PE** has very wide confidence limits
- The **Relative Effect** has some potential, but in a single trial it is based on a single point
- Extension needed!

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 Setting

Z_{ij}	$\longrightarrow$	T_{ij}	β_i
Z_{ij}	$\longrightarrow$	S_{ij}	α_i

- **Trial-level surrogacy:** Relation (α_i, β_i)
- **Individual-level surrogacy:** Relation $(S_{ij}, T_{ij}|Z_{ij})$
- **Let us not confound them**

Developed for a Variety of Settings

- Continuous (Gaussian) endpoints (linear mixed models)
- Binary endpoints (via latent variables)
- Time-to-event endpoints (via copulas)
- Surrogate and true endpoint of different type
- Longitudinally measured endpoints
- Multiple surrogates at the same time
- **Unified frameworks**

ARMD: Trial-Level Surrogacy

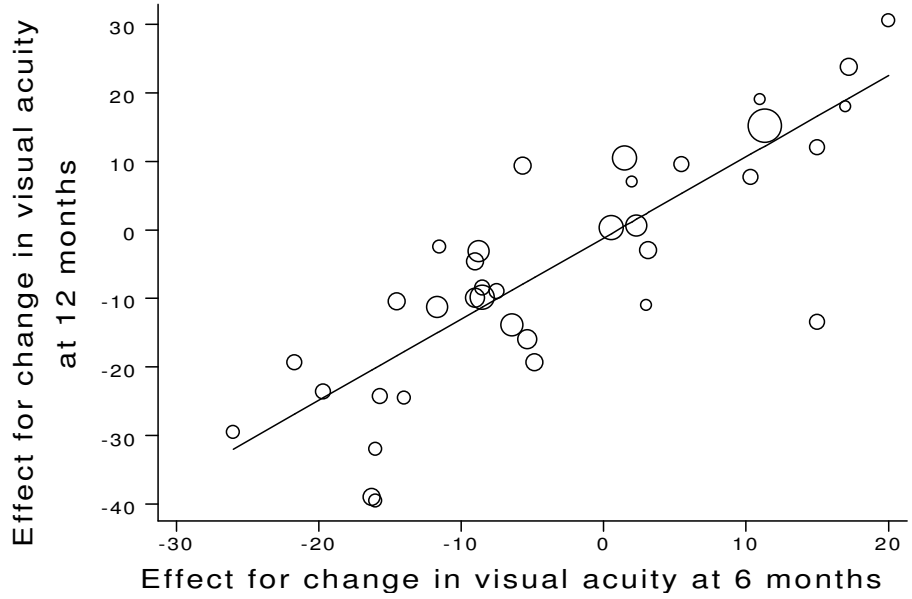
- **Prediction:**
  - ▷ *What do we expect ?*

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

  - ▷ *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. $[0.52; 0.86]$)



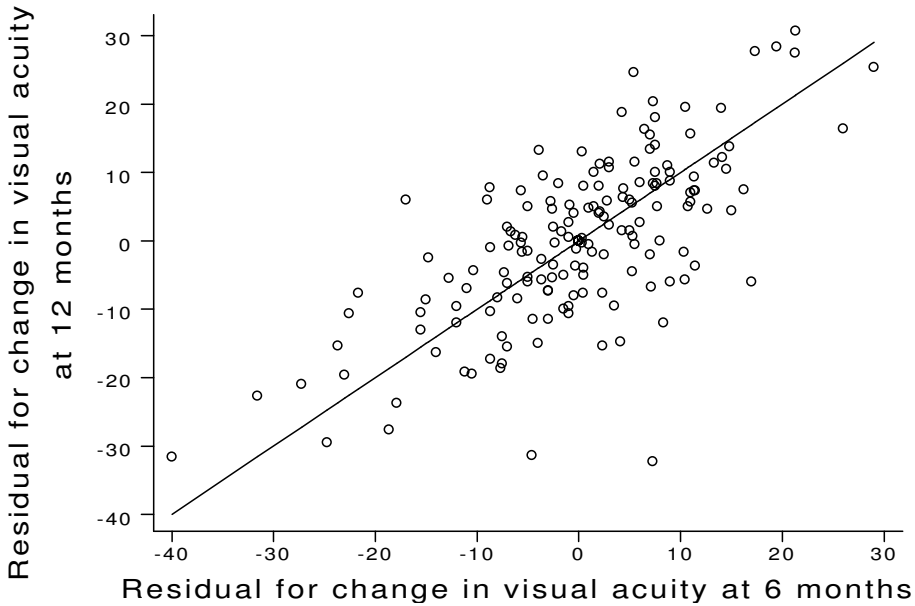
ARMD: Individual-Level Surrogacy

- **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	Vis. Ac. (6 months)	Progr.-free surv.	Progr.-free surv.
True	Vis. Ac. (1 year)	Overall 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.)			
$R^2_{\textbf{trial}}$	$0.69[0.52; 0.86]$	$0.94[0.91; 0.97]$	$0.57[0.41, 0.72]$
$R^2_{\textbf{indiv}}$	$0.48[0.38; 0.59]$	$0.89[0.87; 0.90]$	$0.57[0.52, 0.62]$

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.)			
$R^2_{\textbf{trial}}$	$0.56[0.43; 0.68]$	$0.58[0.45; 0.71]$	$0.70[0.44; 0.96]$
$R^2_{\textbf{indiv}}$	$0.51[0.47; 0.55]$	$0.52[0.48; 0.56]$	$0.55[0.47; 0.62]$

Advanced Ovarian Cancer

	Marginal hazards	
Copula	Ovarian	Colorectal
	Trial-level R^2	
Clayton	0.867 $[0.788, 0.946]$	0.542 $[0.349, 0.735]$
Hougaard	0.900 $[0.839, 0.960]$	0.556 $[0.367, 0.746]$
	Individual-Level Kendall's τ	
Clayton	0.871 $[0.860, 0.883]$	0.603 $[0.560, 0.646]$
Hougaard	0.853 $[0.842, 0.863]$	0.632 $[0.597, 0.667]$

Two Longitudinal Endpoints

- Surrogate measured longitudinally over time
- True endpoint measured longitudinally over time
- There is variability in the true endpoint from patient to patient over time, across the entire time axis
- How much of that variability can be explained by the surrogate sequence?
- The problem is comparable to "explained variability" in multivariate methodology
- Several (related) measures can be constructed:

- Several (related) measures can be constructed:
  - ▷ **VRF:** variance reduction factor
  - ▷ θ_p**:** based on canonical correlation
  - ▷ R^2_Λ**:** based on likelihood ratio test statistic
  - ▷ **LRF:** likelihood reduction factor

- LRF can be applied to all endpoint data types, not only continuous endpoints

One More: An Information-theoretic Approach

- Several (related) measures can be constructed:
  - ▷ **VRF:** variance reduction factor
  - ▷ θ_p**:** based on canonical correlation
  - ▷ R^2_Λ**:** based on likelihood ratio test statistic
  - ▷ **LRF:** likelihood reduction factor
  - ▷ R^2_h**: information-theoretic measure of surrogacy**
- It can be applied to **trial-level surrogacy & individual-level surrogacy**

Some Implications

- **PE & R_h^2:**

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

- In human language:
  - ▷ The good old Proportion Explained and the information-theory measure are similar in structure
  - ▷ But R_h^2 is principled – whereas PE is not

- **Prediction:** It is easier to predict a true endpoint from a surrogate endpoint, when:
  - ▷ The measure of **surrogacy** (e.g., R_h^2) is closer to 1
  - ▷ The **true endpoint** is subject to less measurement error (cf. CD4)

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 (latent correlation)	0.67	[0.55,0.76]
Plackett-Dale odds ratio ψ	25.12	[14.66;43.02]

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:**

Cox model **without** surrogate:

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

Cox model **with** surrogate:

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

Advanced Colorectal Cancer

Parameter	Estimate (95% C.I.) 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 $i = 0$, where only the surrogate is recorded.
- How well can we predict the treatment effect on the true endpoint from the treatment effect on the surrogate endpoint?
- This depends on three factors...

▷ How well is the treatment effect estimated in the new trial?

This depends on the sample size of the new trial

▷ How precise has the meta-analytic evaluation been done?

This depends on the size of the meta-analysis

▷ How good is the surrogacy?

This depends on the variability in the true endpoint & on the R^2_{trial}

- This suggests the question:

> **How large does the treatment effect on the surrogate have to be, to predict a significant treatment effect on the true endpoint**

- Often very high!

- Reflection on the role of surrogates. . .

Potential Outcomes

Alonso, Van der Elst, Molenberghs (Statistical Modeling 2016)

- Setting:

Potential outcomes	(T_{0j}, T_{1j})
Individual causal effect	$\Delta_{Tj} = T_{1j} - T_{0j}$
Expected causal effect	$\beta = E(T_{1j} - T_{0j})$
Surrogate	S_j

- **Predictive causal association:**

$$\rho_\psi = \mathsf{corr}(\Delta_{Tj}, S_j)$$

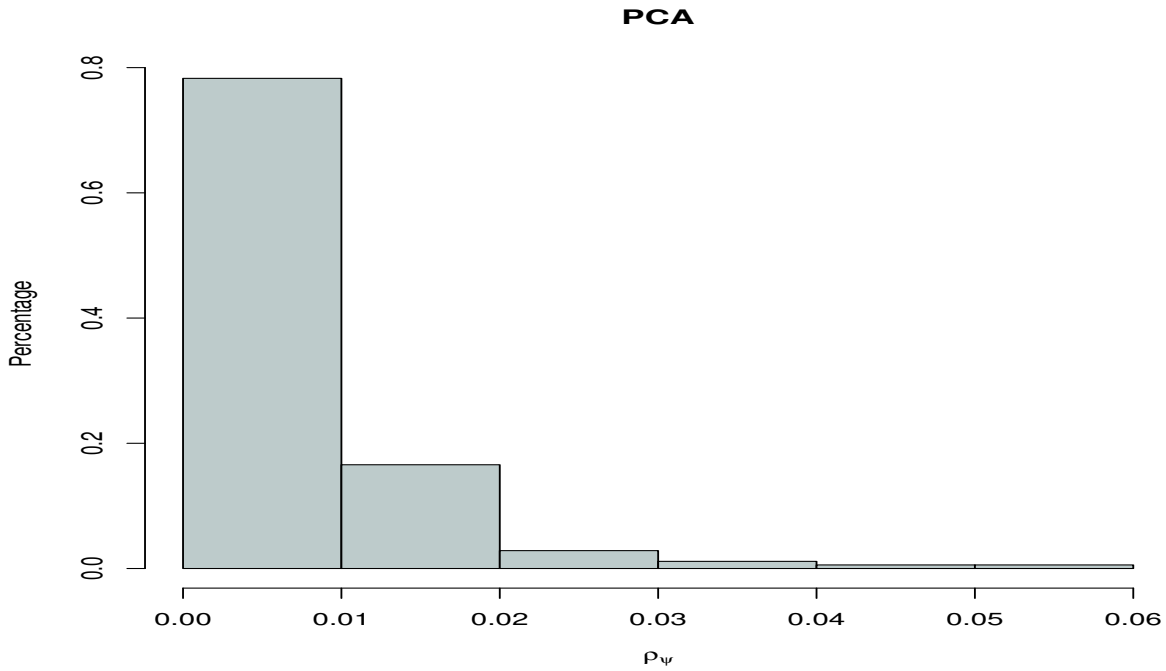
- **Predictive causal association:**

$$\rho_\psi = \text{corr}(\Delta_{Tj}, S_j)$$

- How well can Δ_{Tj} be predicted from the surrogate S_j?
  - ▷ When the variability of Δ_{Tj} is small
  - ▷ When the correlation ρ_ψ is large

- **(Un)identifiability**:
  - ▷ We need the correlation between T_0 and T_1
  - ▷ But, of course, it cannot be identified
  - ▷ So, what do we do?

$\Rightarrow$ **Sensitivity analysis:**


PCA
Percentage
0.0
0.2
0.4
0.6
0.8
0.00
0.01
0.02
0.03
0.04
0.05
0.06
ρ_ψ

Rubin's Model for Causal Inference: One Step Further

- For each patient there exists: $\boldsymbol{Y} = (T_0, T_1, S_0, S_1)'$
  - ▷ Of course, we observe either (T_0, S_0) or (T_1, S_1), but never both
- Nevertheless, define the so-called **Individual Causal Effects:**

$$\boldsymbol{\Delta} = (\Delta T, \Delta S)'$$

  - ▷ $\Delta T = T_1 - T_0$
  - ▷ $\Delta S = S_1 - S_0$
- The **non-identifiability** is, once again, built in.
- We carry on, because we know that sensitivity analysis can help us out.

- Over the entire population, we are interested in the **Expected Causal Effects:**

$$\beta = \mathsf{E}(\Delta T) = E(T_1 - T_0)$$

$$\alpha = \mathsf{E}(\Delta S) = E(S_1 - S_0)$$

- They are each identified in randomized clinical trials.

- But their relationship is **not**.

We Start From Rubin's Model for Causal Inference

$$\boldsymbol{Y} = \begin{pmatrix} T_0 \\ T_1 \\ S_0 \\ S_1 \end{pmatrix} \sim N\left[\boldsymbol{\mu} = \begin{pmatrix} \mu_{T0} \\ \mu_{T1} \\ \mu_{S0} \\ \mu_{S1} \end{pmatrix}, \boldsymbol{\Sigma} = \left(\begin{array}{cc|cc} \sigma_{T0T0} & \sigma_{T0T1} & \sigma_{T0S0} & \sigma_{T0S1} \\ \sigma_{T0T1} & \sigma_{T1T1} & \sigma_{T1S0} & \sigma_{T1S1} \\ \hline \sigma_{T0S0} & \sigma_{T1S0} & \sigma_{S0S0} & \sigma_{S0S1} \\ \sigma_{T0S1} & \sigma_{T1S1} & \sigma_{S0S1} & \sigma_{S1S1} \end{array}\right)\right]$$

$$\boldsymbol{\Delta} = \begin{pmatrix} \Delta T \\ \Delta S \end{pmatrix} = \begin{pmatrix} T_1 - T_0 \\ S_1 - S_0 \end{pmatrix} \sim N\left(\boldsymbol{\mu}_\Delta, \boldsymbol{\Sigma}_\Delta\right),$$

We shall say that S is a **good surrogate** for T if and only if ΔS conveys a substantial amount of **information** on ΔT.

Individual Causal Association (ICA)

- We shall say that S is a **good surrogate** for T if and only if ΔS conveys a substantial amount of **information** on ΔT.

$$\rho_\Delta = \text{corr}(\Delta T, \Delta S).$$

- ρ_Δ is built from the correlations between the components of (T_0, T_1, S_0, S_1).

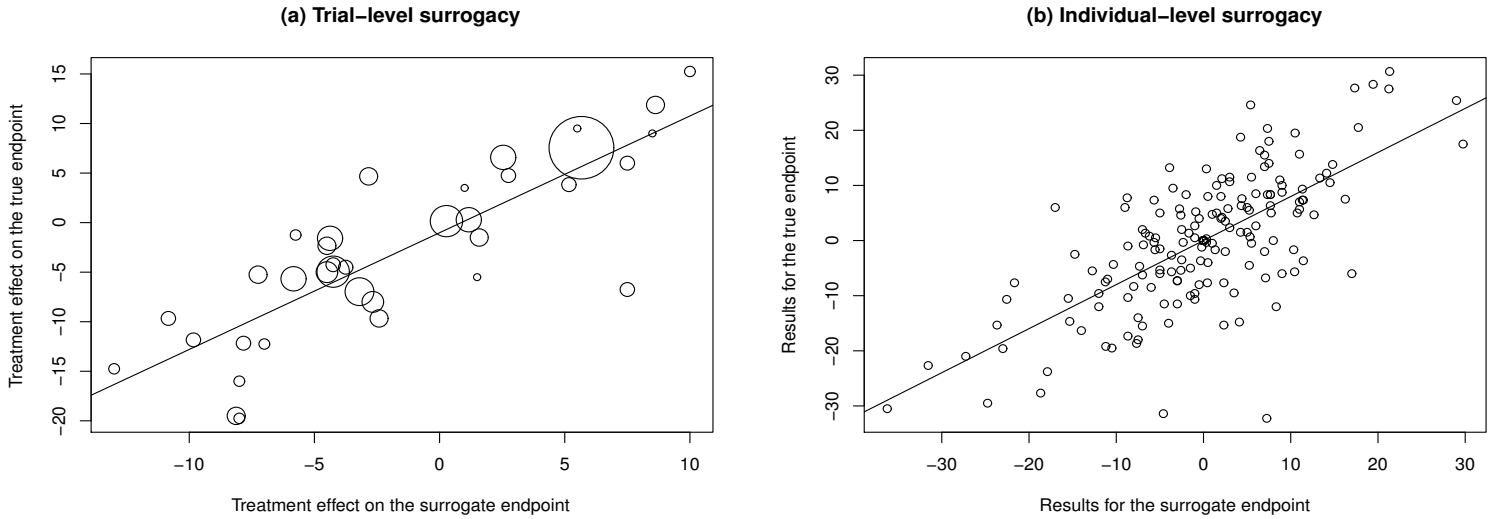
- Some, but not all correlations between variables identifiable from data.

- Assumptions to identify them are often implausible and impossible to test – – so we do not do that!

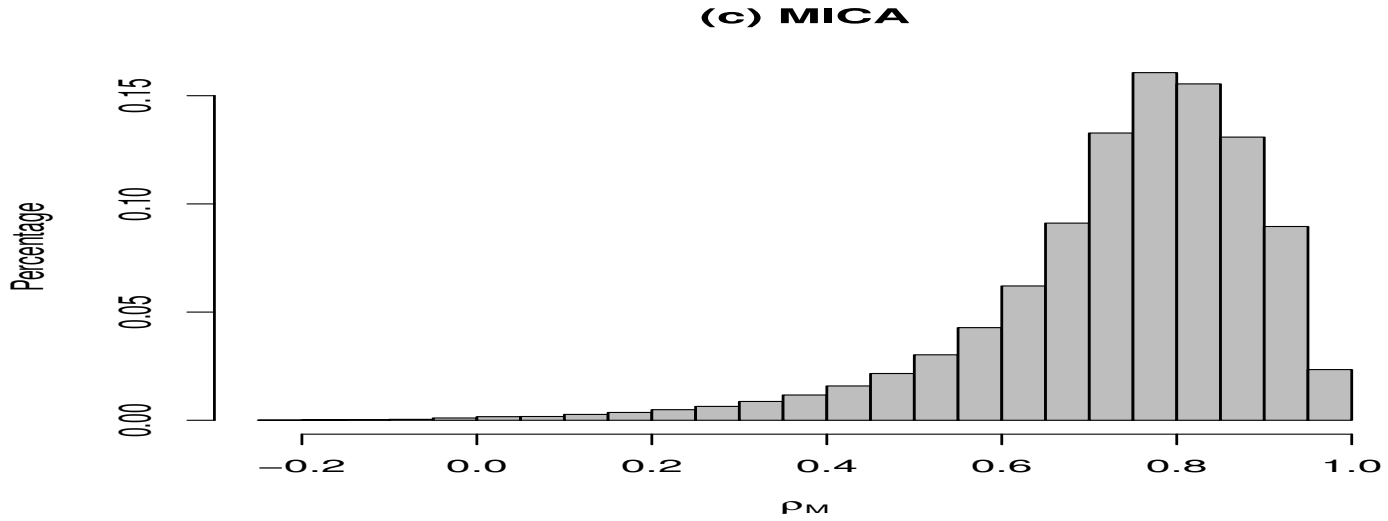
- Information coming from:
  - ▷ design
  - ▷ data
  - ▷ assumptions $\longrightarrow$ **sensitivity analysis**

- This approach can be followed in a meta-analysis as well.

- We denote the meta-analytic version by ρ_M.


(a) Trial−level surrogacy
Treatment effect on the true endpoint
Treatment effect on the surrogate endpoint
−20
−15
−10
−5
0
5
10
15
−10
−5
0
5
10

(b) Individual−level surrogacy
Results for the true endpoint
Results for the surrogate endpoint
−30
−20
−10
0
10
20
30
−30
−20
−10
0
10
20
30


(c) MICA
Percentage
0.00
0.05
0.10
0.15
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
pM

Multivariate Surrogate Endpoints

Van der Elst, Ong, Stijven, Alonso, Van Keilegom, Geys, Eisele, Molenberghs (2024)

- It has often implicitly been assumed that the treatment effect on the true endpoint can be accurately predicted based on the treatment effect on a single surrogate endpoint

- Given the complex nature of many diseases and the various therapeutic pathways in which a treatment can impact the true endpoint, this assumption may be overly optimistic

- The methodology can be extended to the setting of multiple surrogates

- Setting:
  - ▷ True endpoint T
  - ▷ A collection of candidate surrogates $(S_1,\ S_2, \ldots,\ S_p)^T$
  - ▷ Treatment Z (0/1)

- The **Individual Causal Association** takes the form of a multiple correlation R^2_H

- Computational challenges can be efficiently handled

  Florez *et al.* (2022)

- At the individual level:

$$\begin{aligned} T &= \mu_T + \beta Z + \varepsilon_T \\ S_1 &= \mu_{S_1} + \alpha_1 Z + \varepsilon_{S_1} \\ S_2 &= \mu_{S_2} + \alpha_2 Z + \varepsilon_{S_2} \\ &\vdots \\ S_p &= \mu_{S_p} + \alpha_p Z + \varepsilon_{S_p} \end{aligned}$$

At the individual level: **multivariate adjusted association**: the multiple correlation coefficient of T given the surrogates: γ^2_Δ.

Meta-analysis Without Sharing Data: Federated Data Analysis

- Standardized data format across sponsors:

i	Z_i	S_i	T_i
⋮	⋮	⋮	⋮

- Software code provided to sponsors
- **Trial-specific analysis in-house**
- Estimation results provided to central core
- **Combination analysis at central core**

Work Flow for Federated Data Analysis

- **Step 1.** Apply federated work flow on simulated data
- **Step 2.** Analyze natural history data $\Longrightarrow$ Suggestions for potential surrogates

  Maybe. . . individual-level surrogacy only

- **Step 3.** Apply federated work flow to placebo arms
- **Step 4.** Apply federated work flow to placebo & active arms

Choice of Endpoints

- **True endpoint:**
  - ▷ CDR-SB or ...
  - ▷ Several endpoints together
  - ▷ Measured once, change from baseline, longitudinal sequence, ...
- **Surrogate endpoint:**
  - ▷ Amyloid plaques
  - ▷ TAU PET
  - ▷ Measured once, change from baseline, longitudinal sequence, ...
  - ▷ Region
  - ▷ Standardization
  - ▷ Several endpoints together

Natural History Data: ADNI

- **True endpoint:** CDR-SB
- **Surrogate endpoints:** Tau PET SUVR in inferior temporal, caudal, middle frontal, precuneus, meta temporal
- **Sample size:** 357 patients
- **Accommodating baseline covariates:** age, sex, APOE $\epsilon 4$ allele, ABETA centiloid at initial tau scan, education
- **Hierarchies in the data:**
  - ▷ **Longitudinal data:** time flexibly accommodated: important!
  - ▷ **Meta-analysis:** site: not needed!
- **Evaluation:** VRF: How much variability is explained by the surrogate:
  - ▷ **longitudinal:** throughout a patient's profile
  - ▷ **cross-sectional:** at a given point in time

Natural History Data: ADNI: Results

Subgroup	VRF (patient)	VRF (cross-sect.)
CN & MCI & Dementia	0.28	0.029
CN	0.12	0.007
MCI	0.23	0.004
Dementia	0.32	0.149

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.
- Keep individual-level and trial-level surrogacy separate!
- Statisticians **evaluate** surrogates – not more than that...
- Whether it will be used in practice is a different matter:
  - ▷ A regulatory question
  - ▷ Conversation with sponsor and patient (representatives)
  - ▷ For a certain amount of time or for good...