The Statistical Evaluation of Surrogate Endpoints in Clinical Trials

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

EJ-PRD

November 18, 2022
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

<table>
<thead>
<tr>
<th>Event</th>
<th>True endpoint trial</th>
<th>Surrogate endpoint trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endpoint</td>
<td>Size</td>
</tr>
<tr>
<td>MI</td>
<td>Death</td>
<td>4000</td>
</tr>
<tr>
<td>MI</td>
<td>Death</td>
<td>4000</td>
</tr>
<tr>
<td>Stroke</td>
<td>Stroke</td>
<td>25000</td>
</tr>
</tbody>
</table>

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)
Examples of Biomarkers in Oncology

<table>
<thead>
<tr>
<th>Disease</th>
<th>Biomarker</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovarian cancer</td>
<td>CA-125</td>
<td>Survival</td>
</tr>
<tr>
<td>Resected colorectal cancer</td>
<td>CEA</td>
<td>Time to recurrence</td>
</tr>
<tr>
<td>Germ-cell malignancies</td>
<td>AFP</td>
<td>Survival</td>
</tr>
<tr>
<td>Gastrointestinal stromal tumors</td>
<td>PET scan</td>
<td>Survival</td>
</tr>
<tr>
<td>Hormone-dependent prostate cancer</td>
<td>PSA</td>
<td>Time to progression</td>
</tr>
<tr>
<td>Advanced prostate cancer</td>
<td>PSA</td>
<td>Survival</td>
</tr>
</tbody>
</table>
## Candidate Surrogate Endpoints?

<table>
<thead>
<tr>
<th>Disease</th>
<th>Surrogate</th>
<th>Type</th>
<th>True Endpoint</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced cancer</td>
<td>Tumor response</td>
<td>Discr.</td>
<td>Survival</td>
<td>Surv.</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>BMD</td>
<td>Longit.</td>
<td>Fracture</td>
<td>Bin.</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Ejection fraction</td>
<td>Cont.</td>
<td>MI</td>
<td>Bin.</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Blood pressure</td>
<td>Cont.</td>
<td>Coronary HD</td>
<td>Bin.</td>
</tr>
<tr>
<td>Arrhythmias</td>
<td>Arrhythmias</td>
<td>Longit.</td>
<td>Survival</td>
<td>Surv.</td>
</tr>
<tr>
<td>HIV infection</td>
<td>CD4 counts</td>
<td>Longit.</td>
<td>AIDS</td>
<td>Surv.</td>
</tr>
<tr>
<td>AIDS</td>
<td>Viral load</td>
<td>Longit.</td>
<td>Survival</td>
<td>Surv.</td>
</tr>
<tr>
<td>Depression</td>
<td>Biomarkers</td>
<td>Cont.</td>
<td>Depression</td>
<td>Cont.</td>
</tr>
</tbody>
</table>
Bad Precedents

Fleming and Demets (Ann Intern Med 1996)

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 $\gamma$-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

![Diagram showing the relationship between $Z$, $S$, and $T$.]

Mechanism of disease
Example

\(Z\): Dietary changes

\(S\): Colorectal polyps

\(T\): Colorectal adenocarcinomas

\[ Z \rightarrow S \rightarrow T \]

Hyperproliferation

_Schatzkin and Gail (Nature Reviews (Cancer) 2001)_
Biological Concern

$Z$: Dietary changes
$S$: Colorectal polyps
$I$: Intermediate step
$T$: 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)

\( Z: \text{Interferon-}\alpha \)

\( S: \text{Visual acuity at 6 months} \)

\( T: \text{Visual acuity at 1 year} \)

\( N: 190 \text{ patients in 36 centers (\# patients/center } \in [2;18]) \)
Visual Acuity

<table>
<thead>
<tr>
<th>V</th>
<th>A</th>
<th>L</th>
<th>I</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>T</td>
<td>I</td>
<td>O</td>
<td>N</td>
</tr>
<tr>
<td>O</td>
<td>F</td>
<td>S</td>
<td>U</td>
<td>R</td>
</tr>
<tr>
<td>R</td>
<td>O</td>
<td>G</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>E</td>
<td>M</td>
<td>A</td>
<td>R</td>
<td>K</td>
</tr>
<tr>
<td>E</td>
<td>R</td>
<td>S</td>
<td>I</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>A</td>
<td>N</td>
<td>D</td>
<td>O</td>
</tr>
<tr>
<td>M</td>
<td>I</td>
<td>Z</td>
<td>E</td>
<td>D</td>
</tr>
<tr>
<td>E</td>
<td>X</td>
<td>P</td>
<td>E</td>
<td>R</td>
</tr>
<tr>
<td>I</td>
<td>M</td>
<td>E</td>
<td>N</td>
<td>T</td>
</tr>
</tbody>
</table>
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.”

\[
S_j = \mu_S + \alpha Z_j + \varepsilon_{Sj}
\]
\[
T_j = \mu_T + \beta Z_j + \varepsilon_{Tj}
\]
\[
\Sigma = \begin{pmatrix}
\sigma_{SS} & \sigma_{ST} \\
\sigma_{ST} & \sigma_{TT}
\end{pmatrix}
\]
\[
T_j = \mu + \gamma S_j + \varepsilon_j
\]
## Prentice’s Criteria and Measures


<table>
<thead>
<tr>
<th>Quantity</th>
<th>Estimate</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Effect of $Z$ on $T$</td>
<td>$\beta$</td>
<td>$(T</td>
</tr>
<tr>
<td>2 Effect of $Z$ on $S$</td>
<td>$\alpha$</td>
<td>$(S</td>
</tr>
<tr>
<td>3 Effect of $S$ on $T$</td>
<td>$\gamma$</td>
<td>$(T</td>
</tr>
<tr>
<td>4 Effect of $Z$ on $T$, given $S$</td>
<td>$\beta_S$</td>
<td>$(T</td>
</tr>
</tbody>
</table>

\[ \downarrow \]

**Proportion Explained**

\[ PE = \frac{\beta - \beta_S}{\beta} \]

**Relative Effect**

\[ RE = \frac{\beta}{\alpha} \]

**Adjusted Association**

\[ \rho_Z = \text{Corr}(S, T|Z) \]
Prentice’s Criteria and Measures


<table>
<thead>
<tr>
<th>Quantity</th>
<th>Estimate</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Effect of $Z$ on $T$</td>
<td>$\hat{\beta} = 4.12(2.32)$</td>
<td>$p = 0.079$</td>
</tr>
<tr>
<td>2 Effect of $Z$ on $S$</td>
<td>$\hat{\alpha} = 2.83(1.86)$</td>
<td>$p = 0.13$</td>
</tr>
<tr>
<td>3 Effect of $S$ on $T$</td>
<td>$\hat{\gamma} = 0.95(0.06)$</td>
<td>$p &lt; 0.0001$</td>
</tr>
<tr>
<td>4 Effect of $Z$ on $T$, given $S$</td>
<td>$\hat{\beta}_S$</td>
<td></td>
</tr>
</tbody>
</table>

Proportion Explained

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

Relative Effect

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

Adjusted Association

$\hat{\rho}_Z = 0.75 \ [0.69; 0.82]$
Relationship and Problems

\[ RE = \frac{\beta}{\alpha} \]

\[ \rho_Z = \frac{\sigma_{ST}}{\sqrt{\sigma_{SS} \sigma_{TT}}} \]

\[ PE = \lambda \cdot \rho_Z \cdot \frac{\alpha}{\beta} = \lambda \cdot \rho_Z \cdot \frac{1}{RE} \]

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** → **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** → **Individual-Level Surrogacy**
    How close is the relationship between the surrogate and true outcome, after accounting for trial and treatment effects?
“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:**

\[
S_{ij} = \mu_{Si} + \alpha_i Z_{ij} + \varepsilon_{Si j}
\]

\[
T_{ij} = \mu_{Ti} + \beta_i Z_{ij} + \varepsilon_{Ti j}
\]

- **Error structure:**

\[
\Sigma = \begin{pmatrix}
\sigma_{SS} & \sigma_{ST} \\
\sigma_{ST} & \sigma_{TT}
\end{pmatrix}
\]
Statistical Model

• Model:

\[ S_{ij} = \mu_S i + \alpha_i Z_{ij} + \varepsilon_{Sij} \]
\[ T_{ij} = \mu_T i + \beta_i Z_{ij} + \varepsilon_{Tij} \]

• Trial-specific effects:

\[
\begin{pmatrix}
\mu_{S_i} \\
\mu_{T_i} \\
\alpha_i \\
\beta_i
\end{pmatrix}
\begin{pmatrix}
\mu_S \\
\mu_T \\
\alpha \\
\beta
\end{pmatrix}
\begin{pmatrix}
\mu_{S_i} \\
\mu_{T_i} \\
\alpha_i \\
\beta_i
\end{pmatrix}
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}
**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:
  \[ \rho_Z = R_{\text{indiv}} = \text{Corr}(\varepsilon_T, \varepsilon_S) \]

- Estimate:
  \[ R_{\text{indiv}}^2 = 0.483 \text{ (95\% C.I. [0.38; 0.59])} \]
  \[ R_{\text{indiv}} = 0.69 \text{ (95\% C.I. [0.62; 0.77])} \]
  \[ \text{Recall } \rho_Z = 0.75 \text{ (95\% C.I. [0.69; 0.82])} \]
## A Number of Case Studies

<table>
<thead>
<tr>
<th>Surrogate True</th>
<th>Age-related macular degeneration</th>
<th>Advanced ovarian cancer</th>
<th>Advanced colorectal cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vis. Ac. (6 months)</td>
<td>Progr.-free surv.</td>
<td>Progr.-free surv.</td>
<td></td>
</tr>
<tr>
<td>Vis. Ac. (1 year)</td>
<td>Overall surv.</td>
<td>Overall surv.</td>
<td></td>
</tr>
</tbody>
</table>

### Prentice Criteria 1–3 ($p$ value)

<table>
<thead>
<tr>
<th>Association ($Z, S$)</th>
<th>0.31</th>
<th>0.013</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association ($Z, T$)</td>
<td>0.22</td>
<td>0.08</td>
<td>0.86</td>
</tr>
<tr>
<td>Association ($S, T$)</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

### Single-Unit Validation Measures (estimate and 95% C.I.)

<table>
<thead>
<tr>
<th>Proportion Explained</th>
<th>0.61[−0.19; 1.41]</th>
<th>1.34[0.73; 1.95]</th>
<th>0.51[−4.97; 5.99]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Effect</td>
<td>1.51[−0.46; 3.49]</td>
<td>0.65[0.36; 0.95]</td>
<td>1.59[−15.49; 18.67]</td>
</tr>
<tr>
<td>Adjusted Association</td>
<td>0.74[0.68; 0.81]</td>
<td>0.94[0.94; 0.95]</td>
<td>0.73[0.70, 0.76]</td>
</tr>
</tbody>
</table>

### Multiple-Unit Validation Measures (estimate and 95% C.I.)

<table>
<thead>
<tr>
<th>$R^2_{\text{trial}}$</th>
<th>0.69[0.52; 0.86]</th>
<th>0.94[0.91; 0.97]</th>
<th>0.57[0.41, 0.72]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2_{\text{indiv}}$</td>
<td>0.48[0.38; 0.59]</td>
<td>0.89[0.87; 0.90]</td>
<td>0.57[0.52, 0.62]</td>
</tr>
</tbody>
</table>
## Overview: Case Studies

<table>
<thead>
<tr>
<th></th>
<th>Schizophrenia Study I (138 units)</th>
<th>Schizophrenia Study I (29 units)</th>
<th>Schizophrenia Study II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surrogate</strong></td>
<td>— PANSS —</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>True</strong></td>
<td>— CGI —</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Prentice Criteria 1–3 (p value)

<table>
<thead>
<tr>
<th>Association</th>
<th>Schizophrenia Study I (138 units)</th>
<th>Schizophrenia Study I (29 units)</th>
<th>Schizophrenia Study II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(Z, S)$</td>
<td>0.016</td>
<td>0.835</td>
<td></td>
</tr>
<tr>
<td>$(Z, T)$</td>
<td>0.007</td>
<td>0.792</td>
<td></td>
</tr>
<tr>
<td>$(S, T)$</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>

### Single-Unit Validation Measures (estimate and 95% C.I.)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Schizophrenia Study I (138 units)</th>
<th>Schizophrenia Study I (29 units)</th>
<th>Schizophrenia Study II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Explained</td>
<td>0.81[0.46; 1.67]</td>
<td></td>
<td>−0.94[∞]</td>
</tr>
<tr>
<td>Relative Effect</td>
<td>0.055[0.01; 0.16]</td>
<td></td>
<td>−0.03[∞]</td>
</tr>
<tr>
<td>Adjusted Association</td>
<td>0.72[0.69; 0.75]</td>
<td></td>
<td>0.74[0.69; 0.79]</td>
</tr>
</tbody>
</table>

### Multiple-Unit Validation Measures (estimate and 95% C.I.)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Schizophrenia Study I (138 units)</th>
<th>Schizophrenia Study I (29 units)</th>
<th>Schizophrenia Study II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2_{\text{trial}}$</td>
<td>0.56[0.43; 0.68]</td>
<td>0.58[0.45; 0.71]</td>
<td>0.70[0.44; 0.96]</td>
</tr>
<tr>
<td>$R^2_{\text{indiv}}$</td>
<td>0.51[0.47; 0.55]</td>
<td>0.52[0.48; 0.56]</td>
<td>0.55[0.47; 0.62]</td>
</tr>
</tbody>
</table>
Binary Endpoints

\[
\begin{align*}
\tilde{S}_{ij} &= \mu_s + m_{S_i} + (\alpha + a_i)Z_{ij} + \varepsilon_{S_{ij}}, \\
\tilde{T}_{ij} &= \mu_T + m_{T_i} + (\beta + b_i)Z_{ij} + \varepsilon_{T_{ij}},
\end{align*}
\]

where \(\tilde{S}_{ij}\) and \(\tilde{T}_{ij}\) are normally distributed, latent variables:

\[
S_{ij} = \begin{cases} 
1 & \text{if } \tilde{S}_{ij} > 0 \\
0 & \text{if } \tilde{S}_{ij} \leq 0
\end{cases} \quad T_{ij} = \begin{cases} 
1 & \text{if } \tilde{T}_{ij} > 0 \\
0 & \text{if } \tilde{T}_{ij} \leq 0
\end{cases}
\]

- multilevel probit model
- Plackett-Dale model
- pseudo-likelihood
Two-Stage Model for Survival

Stage I

- Survival model for the surrogate endpoint
- Survival model for the true endpoint
- Association function to couple both: copula
  Association can be represented as Kendall’s $\tau$
- Two copula functions:
  \begin{itemize}
  \item \textit{Clayton (1978)}:
    \[ C_\delta(u, v) = (u^{1-\delta} + v^{1-\delta})^{\frac{1}{1-\delta}}, \quad \delta > 1 \]
  \item \textit{Hougaard (1986)}:
    \[ C_\delta(u, v) = \exp\left[-\left\{(-\ln u)^{\frac{1}{\delta}} + (-\ln v)^{\frac{1}{\delta}}\right\}^\delta\right], \quad 0 < \delta < 1 \]
  \end{itemize}
The Clayton Copula

\[ C_\theta(u, v) = (u^{1-\theta} + v^{1-\theta} - 1)^{\frac{1}{1-\theta}}, \quad \theta > 1 \]

- \( \theta > 1 \Rightarrow S \) and \( T \) positively associated
- \( \theta \to 1 \Rightarrow S \) and \( T \) independent
- Kendall’s \( \tau = (\theta - 1)/(\theta + 1) \)
- “late” dependence
The Hougaard Copula

\[ C_\theta(u, v) = \exp\left[ -\left\{ \left( -\ln u \right)^{\frac{1}{\theta}} + \left( -\ln v \right)^{\frac{1}{\theta}} \right\}^\theta \right], \quad 0 < \theta < 1 \]

- \[ \theta < 1 \Rightarrow S \text{ and } T \text{ positively associated} \]
- \[ \theta \to 1 \Rightarrow S \text{ and } T \text{ independent} \]
- \[ \tau = 1 - \theta \]
- “early” dependence
Stage II

- Mixed effects:
  \[
  \begin{pmatrix}
  \alpha_i \\
  \beta_i 
  \end{pmatrix} = 
  \begin{pmatrix}
  \alpha \\
  \beta 
  \end{pmatrix} +
  \begin{pmatrix}
  a_i \\
  b_i 
  \end{pmatrix}
  \]

Error structure of random effects:

\[
D = \begin{pmatrix}
  d_{aa} & d_{ab} \\
  d_{ab} & d_{bb} 
\end{pmatrix}
\]
### Advanced Ovarian Cancer

<table>
<thead>
<tr>
<th>Copula</th>
<th>Ovarian</th>
<th>Colorectal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marginal hazards</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial-level $R^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>0.867 [0.788, 0.946]</td>
<td>0.542 [0.349, 0.735]</td>
</tr>
<tr>
<td>Hougaard</td>
<td>0.900 [0.839, 0.960]</td>
<td>0.556 [0.367, 0.746]</td>
</tr>
<tr>
<td><strong>Individual-Level $\tau$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clayton</td>
<td>0.871 [0.860, 0.883]</td>
<td>0.603 [0.560, 0.646]</td>
</tr>
<tr>
<td>Hougaard</td>
<td>0.853 [0.842, 0.863]</td>
<td>0.632 [0.597, 0.667]</td>
</tr>
</tbody>
</table>
Two Longitudinal Endpoints

First Stage

\[ T_{ijt} = \mu_T + \beta_i Z_{ij} + \theta_T t_{ijt} + \varepsilon_{Tijt} \]
\[ S_{ijt} = \mu_S + \alpha_i Z_{ij} + \theta_S t_{ijt} + \varepsilon_{Sijt} \]

Second Stage

\[ \begin{pmatrix} \mu_{S_i} \\ \mu_{Ti} \\ \alpha_i \\ \beta_i \\ \theta_{Si} \\ \theta_{Ti} \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_T \\ \alpha \\ \beta \\ \theta_S \\ \theta_T \end{pmatrix} + \begin{pmatrix} m_{S_i} \\ m_{Ti} \\ a_i \\ b_i \\ \tau_{Si} \\ \tau_{Ti} \end{pmatrix} \]

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

Evaluation Measures?
A Sequence of Measures

- Variance Reduction Factor VRF:
  \[ VRF = \frac{\sum_i \{ \text{tr}(\Sigma TT_i) - \text{tr}(\Sigma (T|S)_i) \}}{\sum_i \text{tr}(\Sigma TT_i)} \]

- Canonical-correlation Root-statistic Based Measure \( \theta_p \):
  \[ \theta_p = \sum_i \frac{1}{Np_i} \text{tr} \left\{ \left( \Sigma TT_i - \Sigma (T|S)_i \right) \Sigma TT_i^{-1} \right\} \]

- Canonical-correlation Root-statistic Based Measure \( R^2_\Lambda \):
  \[ R^2_\Lambda = \frac{1}{N} \sum_i (1 - \Lambda_i), \]
  where
  \[ \Lambda_i = \frac{|\Sigma_i|}{|\Sigma TT_i| |\Sigma SS_i|} \]
A Sequence of Measures

- **The Likelihood Reduction Factor LRF:**

  ▶ Consider a pair of models:

  \[
g_T(T_{ij}) = \mu T_i + \beta_i Z_{ij}
  \]

  \[
g_T(T_{ij}) = \theta_0 i + \theta_1 Z_{ij} + \theta_2 S_{ij}
  \]

  ▶ \(G_i^2\) 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:

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

• 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** $R^2_h$:

  \[
  R^2_h = R^2_h = \frac{EP(Y) - EP(Y|X)}{EP(Y)}
  \]

  with

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

- **Version for $N$ trials**:

  \[
  R^2_h = \sum_{i=1}^{N_q} \alpha_i R^2_{h,i} = 1 - \sum_{i=1}^{N_q} \alpha_i e^{-2I_i(S_i,T_i)},
  \]

  where the $\alpha_i$ 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

- $\theta_p$ is symmetric in $S$ and $T$ whereas the VRF is not

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

- $R_A^2$ and later ones also apply to non-Gaussian settings
Relationships With Previous Definitions

• Later ones specialize to earlier ones

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

• Longitudinal normal setting:

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

• General setting:

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

when the number of subjects per trial approaches $\infty$
Other Implications

• Relationship with Prentice’s main criterion & Data Processing Inequality:

\[ f(T|Z,S) = F(T|S) \implies Z \rightarrow S \rightarrow T \]

\[ \implies I(T, Z|S') = 0 \]

\[ \implies I(Z, S') \geq I(Z, T) \]

• PE & \( R^2_h \):

\[ PE = 1 - \frac{\beta_S}{\beta} \quad \longleftrightarrow \quad R^2_h = 1 - \frac{\text{EP}(\beta_i|\alpha_i)}{\text{EP}(\beta_i)} \]
Fano’s Inequality

- Fano’s Inequality:

\[
E \left[ (T - g(S))^2 \right] \geq EP(T)(1 - R_{zh}^2)
\]

▷ Left hand side is prediction error

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

▷ “How large does \( R_{zh}^2 \) have to be?” \( \text{←} \) The answer depend crucially on the power entropy of \( T \)
Schizophrenia Trial

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

- **Binary Outcomes:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial-level $R^2_{\text{trial}}$ measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information-theoretic</td>
<td>0.49</td>
<td>[0.21, 0.81]</td>
</tr>
<tr>
<td>Probit</td>
<td>0.51</td>
<td>[0.18, 0.78]</td>
</tr>
<tr>
<td>Plackett-Dale</td>
<td>0.51</td>
<td>[0.21, 0.81]</td>
</tr>
<tr>
<td><strong>Individual-level measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2_{h}$</td>
<td>0.27</td>
<td>[0.24, 0.33]</td>
</tr>
<tr>
<td>$R^2_{h,\text{max}}$</td>
<td>0.39</td>
<td>[0.35, 0.48]</td>
</tr>
<tr>
<td>Probit</td>
<td>0.67</td>
<td>[0.55, 0.76]</td>
</tr>
<tr>
<td>Plackett-Dale $\psi$</td>
<td>25.12</td>
<td>[14.66; 43.02]</td>
</tr>
<tr>
<td>Fano’s lower-bound</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>
Age-related Macular Degeneration Trial

- Both outcomes binary:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>[95% C.I.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2_{\text{trial}}$</td>
<td>0.3845</td>
<td>[0.1494;0.6144]</td>
</tr>
<tr>
<td>$R^2_h$</td>
<td>0.2648</td>
<td>[0.2213;0.3705]</td>
</tr>
<tr>
<td>$R^2_{h\text{max}}$</td>
<td>0.4955</td>
<td>[0.3252;0.6044]</td>
</tr>
</tbody>
</table>
$S$: Time to progression/death

$T$: Time to death

• **Models:**

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

\[ h_{ij}(t) = h_{i0}(t)\exp\{\beta_{Si} Z_{ij} + \gamma_i S_{ij}(t)\} \]
## Advanced Colorectal Cancer

### Estimate (95% C.I.)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dataset I</th>
<th>Dataset II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial-level measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{R}^2_{\text{trial}}$ (separate models)</td>
<td>0.82 [0.40;0.95]</td>
<td>0.85 [0.53;0.96]</td>
</tr>
<tr>
<td>$\hat{R}^2_{\text{trial}}$ (Clayton copula)</td>
<td>0.88 [0.59;0.98]</td>
<td>0.82 [0.43;0.95]</td>
</tr>
<tr>
<td>$\hat{R}^2_{\text{trial}}$ (Hougaard copula)</td>
<td></td>
<td>0.75 [0.00;1.00]</td>
</tr>
<tr>
<td><strong>Individual-level measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{R}^2_{ih}$</td>
<td>0.84 [0.82;0.85]</td>
<td>0.83 [0.82;0.85]</td>
</tr>
<tr>
<td>Percentage of censoring</td>
<td>19%</td>
<td>55%</td>
</tr>
</tbody>
</table>

The Statistical Evaluation of Surrogate Endpoints in Clinical Trials
Prediction in a New Trial

- Consider a new trial $i = 0$:

\[ S_{0j} = \mu_{s0} + \alpha_0 Z_{0j} + \varepsilon_{s0j} \]

- **Prediction variance:**

\[
\text{Var}(\beta + b_0 | \mu_{s0}, \alpha_0, \vartheta) \approx f \{ \text{Var}(\bar{\mu}_{s0}, \bar{\alpha}_0) \} + f \{ \text{Var}(\vartheta) \} + (1 - R^2_{\text{trial}}) \text{Var}(b_0)
\]

- where

  - $f(\cdot)$ are appropriate functions of the parameters involved
  - $\vartheta$ 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:**
    
    $$\text{Var}(\beta + b_0 | \mu_{S0}, \alpha_0, \vartheta) \approx f\{\text{Var}(\vartheta)\} + (1 - R_{\text{trial}}^2)\text{Var}(b_0)$$
  
  ▶ **No estimation error:**
    
    $$\text{Var}(\beta + b_0 | m_{S0}, a_0) = (1 - R_{\text{trial}}^2)\text{Var}(b_0)$$
The Surrogate Threshold Effect

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

- **Various versions:**
  - $\Delta \text{STE}_{N,n}$: STE for a finite meta-analysis and a finite new trial
  - $\Delta \text{STE}_{N,\infty}$: STE for a finite meta-analysis and an infinite new trial
  - $\Delta \text{STE}_{\infty,\infty}$: STE when both the meta-analysis and the new trial are infinitely large
## Potential Outcomes

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

- **Setting:**

<table>
<thead>
<tr>
<th>Potential outcomes</th>
<th>$(T_{0j}, T_{1j})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual causal effect</td>
<td>$\Delta_{Tj} = T_{1j} - T_{0j}$</td>
</tr>
<tr>
<td>Expected causal effect</td>
<td>$\beta = E(T_{1j} - T_{0j})$</td>
</tr>
<tr>
<td>Surrogate</td>
<td>$S_j$</td>
</tr>
</tbody>
</table>
• Normality:

\[ Y_j = \begin{pmatrix} T_{0j} \\ T_{1j} \\ S_j \end{pmatrix} \text{ normal } \Rightarrow \begin{pmatrix} \Delta T_j \\ S_j \end{pmatrix} \text{ normal} \]

• Predictive causal association:

\[ \rho_\psi = \text{corr}(\Delta T_j, S_j) \]

• Relation with measure of prediction accuracy (cf. Fano):

\[ E \left[ \{ \Delta T_j - g(S_j) \}^2 \right] = (1 - \rho_\psi^2) \sigma_{\Delta T} \]
• **(Un)identifiability:**

\[ \rho_{T_0 T_1} \text{ not identifiable} \]

⇒ **Sensitivity analysis:**

![Sensitivity analysis graph](image)
Rubin’s Model for Causal Inference

• For each patient there exists: \( Y = (T_0, T_1, S_0, S_1)' \)
  \( \triangleright (T_i, S_i) \) would be observed under the condition \( Z = i, i = 0, 1. \)

• Individual causal effects: \( \Delta = (\Delta T, \Delta S)' \) where
  \( \triangleright \Delta T = T_1 - T_0 \)
  \( \triangleright \Delta S = S_1 - S_0 \)

• **Fundamental problem of causal inference:** \( Y \) and, hence, \( \Delta \) often not observable.

• Expected causal effect: \( \beta = E(\Delta T) \) and \( \alpha = E(\Delta S) \).
Identifiability of the expected causal treatment effects

- Expected causal treatment effects identifiable under three conditions

  **Consistency**: If $Z = z$ for a given subject then $Y_z = Y$ for that subject

  **Conditional exchangeability**: There is no unmeasured confounding: Given baseline covariates $L$, $Y_z \perp Z|L = l$ for each possible value $z$ of $Z$ and $l$ of $L$

  **Positivity**: If $f_L(l) \neq 0$ then $f_{Z|L}(z|l) > 0$

- In randomized clinical trials all conditions hold and

  $$\beta = E(T|Z = 1) - E(T|Z = 0) \quad \text{and} \quad \alpha = E(S|Z = 1) - E(S|Z = 0)$$

- The methodology proposed in the following sections is based only on the individual causal treatment effects and it is valid if consistency holds, i.e., it could also be applied to observational data.
Rubin’s Model for Causal Inference

\[ Y = \begin{pmatrix} T_0 \\ T_1 \\ S_0 \\ S_1 \end{pmatrix} \sim N \begin{pmatrix} \mu_{T0} \\ \mu_{T1} \\ \mu_{S0} \\ \mu_{S1} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{T0T0} & \sigma_{T0T1} & \sigma_{T0S0} & \sigma_{T0S1} \\ \sigma_{T0T1} & \sigma_{T1T1} & \sigma_{T1S0} & \sigma_{T1S1} \\ \sigma_{T0S0} & \sigma_{T1S0} & \sigma_{S0S0} & \sigma_{S0S1} \\ \sigma_{T0S1} & \sigma_{T1S1} & \sigma_{S0S1} & \sigma_{S1S1} \end{pmatrix} \]
Individual Causal Association (ICA)

Given the aforementioned distributional assumptions one has that

\[ \Delta = \begin{pmatrix} \Delta T \\ \Delta S \end{pmatrix} = AY = \begin{pmatrix} T_1 - T_0 \\ S_1 - S_0 \end{pmatrix} \sim N(\mu_\Delta, \Sigma_\Delta), \]

where \( \Sigma_\Delta = A\Sigma A' \), \( \mu_\Delta = A\mu = (\beta, \alpha)' \) and \( A \) contrast matrix.

**Fundamental question**: Given a treatment \( Z \), when should one say that \( S \) is a “good surrogate endpoint” for \( T \)?

**Definition**: We shall say that \( S \) is a good surrogate for \( T \) if and only if \( \Delta S \) conveys a substantial amount of information on \( \Delta T \).
Individual Causal Association (ICA)

- Informational coefficient of correlation

\[ R_H^2 = 1 - e^{I(\Delta T, \Delta S)} = \rho_\Delta^2 \]

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

- Individual causal association: Under homoscedasticity \( \sigma_{T0T0} = \sigma_{T1T1} = \sigma_T \) and \( \sigma_{S0S0} = \sigma_{S1S1} = \sigma_S \)

\[ \rho_\Delta = \text{corr}(\Delta T, \Delta S) = \frac{\rho_{T0S0} + \rho_{T1S1} - \rho_{T1S0} - \rho_{T0S1}}{2\sqrt{(1 - \rho_{T0T1})(1 - \rho_{S0S1})}} \]

- \( \rho_{T0S0} \) and \( \rho_{T1S1} \) identifiable under consistency.

- All the other correlations are not identifiable from the data.
Relationship Causal Inference
Meta-analytic Paradigm

Alonso et al. *(Biometrics 2015)*

- Setting:

\[
Y_j = \begin{pmatrix}
T_{0j} \\
T_{1j} \\
S_{0j} \\
S_{0j}
\end{pmatrix} \quad \implies \quad \Delta_j = \begin{pmatrix}
\Delta T_j \\
\Delta S_j
\end{pmatrix} = \begin{pmatrix}
T_{1j} - T_{1j} \\
S_{1j} - S_{0j}
\end{pmatrix}
\]
• **Individual causal association (ICA):**

\[ \rho_\Delta = \text{corr}(\Delta_{Tj}, \Delta_{Sj}) \]

• **Joint distribution unidentifiable**

• Capture assumptions in **causal diagrams** → reduced forms of \( \rho_\Delta \)

• Information coming from:
  ▶ design
  ▶ data
  ▶ assumptions \( \rightarrow \) sensitivity
Meta-analytic formulation:

\[ \Delta T_{ij} = \beta_i + \varepsilon \Delta T_{ij} \]
\[ \Delta S_{ij} = \alpha_i + \varepsilon \Delta S_{ij} \]

Meta-analytic Individual Causal Association:

\[ \rho_M = \text{corr}(\Delta T_{ij}, \Delta S_{ij}) \]
(a) Trial-level surrogacy

(b) Individual-level surrogacy

(c) MICA

The Statistical Evaluation of Surrogate Endpoints in Clinical Trials
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 $S$.

- $R^2_{\text{trial}}, R^2_{\text{indiv}} (\psi, \tau), \text{VRF}, \theta_p, R^2_{\Lambda} \text{LRF}, R^2_h, \ldots$: 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
  - Promising causal-inference/meta-analytic framework