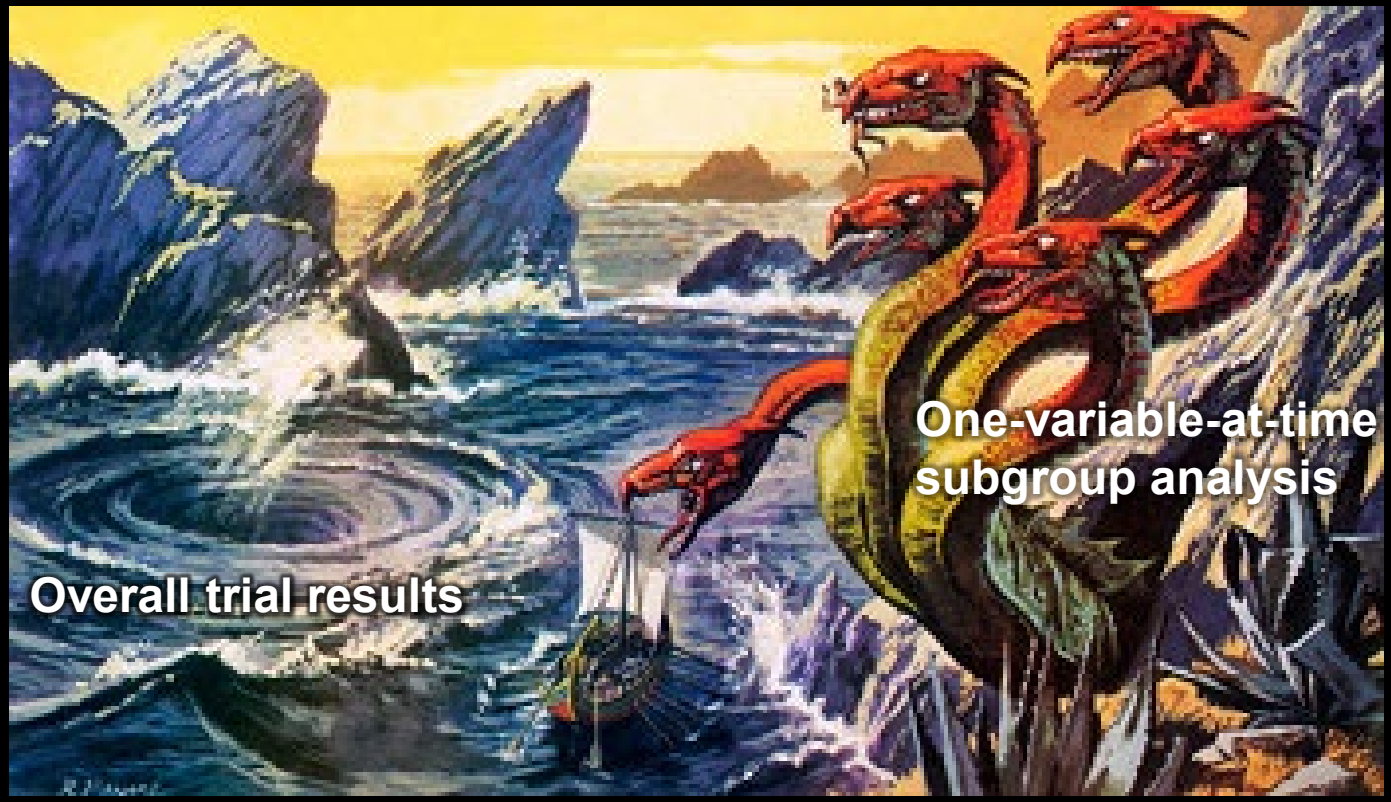
Overall average treatment effects and one-variable-at-a-time subgroup analysis: The Scylla and Charybdis of Evidence Based Medicine

David M. Kent, MD, MSc
Professor of Medicine, Neurology, Clinical and Translational Science,
Director, Predictive Analytics and Comparative Effectiveness (PACE) Center,
Institute for Clinical Research and Health Policy Studies, Tufts Medical Center

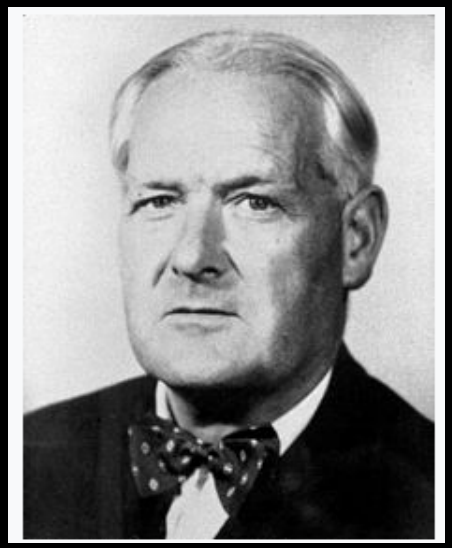
Scylla and Charybdis



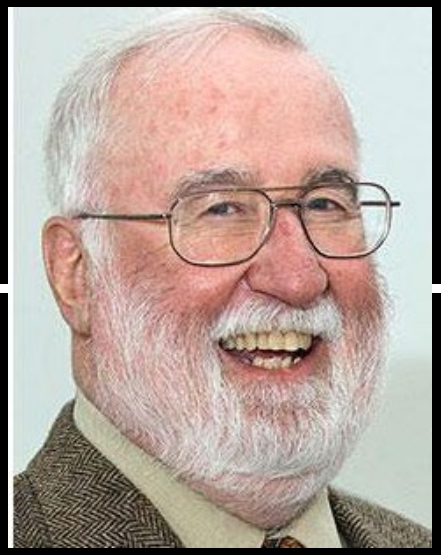
Limitations of RCTs for Clinical Decision Making

Limitations of RCTs for Clinical Decision Making

- While RCTs can determine the better treatment on average, they "do not answer the practicing doctor's question*: what is the most likely outcome when this particular drug is given to a particular patient?"*
  - Austin Bradford Hill



Evidence Based Medicine

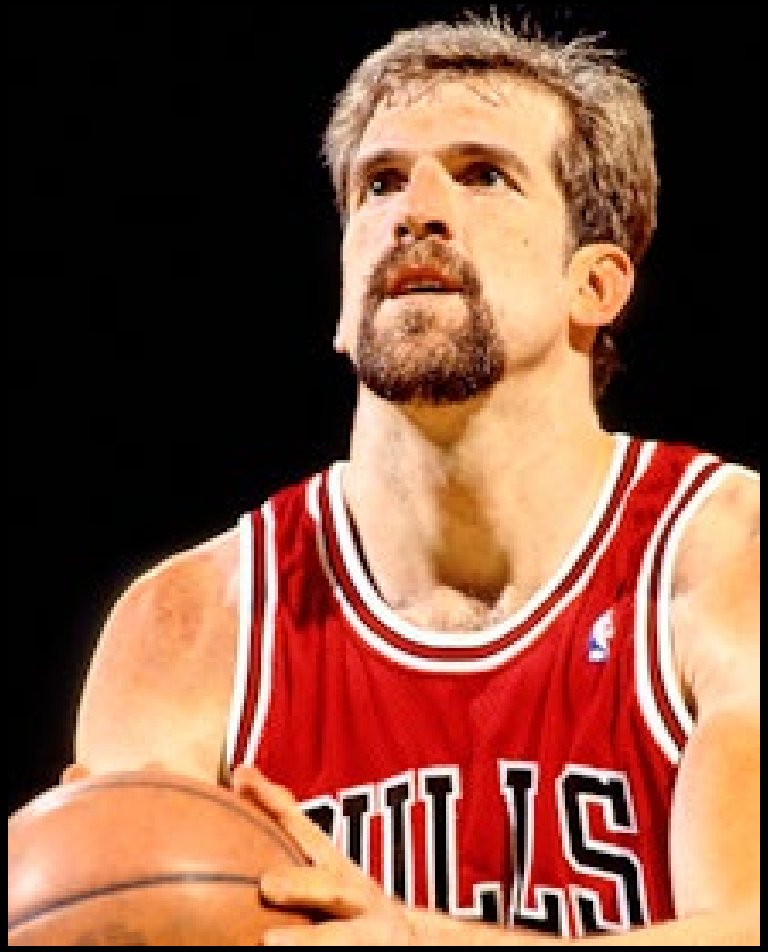


- Hill was wrong.
- RCTs should be used by doctors to determine what's best for individual patients.
- "Evidence based medicine is the conscientious, explicit, judicious and reasonable use of current best evidence in making decisions about the care of individual patients."

- EBM proposed to repurpose RCTs from tools to establish causation into tools for *prediction* in single cases.

- What's best on average must be best for each individual.

The Fallacy of Division (Wennington's Fallacy)



Starters	PTS
B.J. Armstrong	16
Scottie Pippen	19
Michael Jordan	55
Toni Kukoc	3
Will Perdue	6
Reserves	**PTS**
Luc Longley	5
Corie Blount	2
Steve Kerr	5
Larry Krystkowiak	0
Bill Wennington	2
Pete Myers	0
Team Totals	**113**

"Michael and I combined for 57 points"
-Bill Wennington, 1995

- It is potentially misleading to draw inferences about individuals based on aggregated characteristics of the (heterogeneous) group to which they belong.
- How do we estimate "individual" treatment effects?

Clinical Trial: "Box Score"

ACTUAL OUTCOME

0 = alive
1 = dead

Subject Name	Without Treatment	With Treatment
SAM	0	
MARY		0
BOB	0	
BEN		0
CHRISTINE		0
NEIL	1	
MOHAMED		1
JENNIFER		1
PAUL	0	
NISHA	1	
MIGUEL	1	
LAYLA		0
PAUL	0	
EMANUEL		1
CHERYL		0
PATRICK	0	
OSCAR		1
JULIANNE	0	
THOMAS	0	
GEORGE		0

Individual Treatment Effects in a Deterministic Framework: Four possibilities

Without Treatment	With Treatment	
0	0	NO EFFECT
0	1	HARM
1	0	BENEFIT
1	1	NO EFFECT

Clinical Trial: "Box Score"

ACTUAL OUTCOME

COUNTER FACTUAL OUTCOME

0 = alive
1 = dead

Subject Name	Without Treatment	With Treatment
SAM	0	1
MARY	0	0
BOB	0	0
BEN	1	0
CHRISTINE	1	0
NEIL	1	1
MOHAMED	1	1
JENNIFER	1	1
PAUL	0	1
NISHA	1	1
MIGUEL	1	1
LAYLA	1	0
PAUL	0	0
EMANUEL	1	1
CHERYL	0	0
PATRICK	0	0
OSCAR	1	1
JULIANNE	0	0
THOMAS	0	0
GEORGE	1	0

Individual Treatment Effects in a Deterministic Framework: Four possibilities

Without Treatment	With Treatment	
0	0	NO EFFECT
0	1	HARM
1	0	BENEFIT
1	1	NO EFFECT

Clinical Trial: "Box Score"

ACTUAL OUTCOME
COUNTER FACTUAL OUTCOME

0 = alive
1 = dead

Subject Name	Without Treatment	With Treatment	
SAM	0	1	← HARM
MARY	0	0	
BOB	0	0	
BEN	1	0	← BENEFIT
CHRISTINE	1	0	← BENEFIT
NEIL	1	1	
MOHAMED	1	1	
JENNIFER	1	1	
PAUL	0	1	← HARM
NISHA	1	1	
MIGUEL	1	1	
LAYLA	1	0	← BENEFIT
PAUL	0	0	
EMANUEL	1	1	
CHERYL	0	0	
PATRICK	0	0	
OSCAR	1	1	
JULIANNE	0	0	
THOMAS	0	0	
GEORGE	1	0	← BENEFIT

Individual Treatment Effects in a Deterministic Framework: Four possibilities

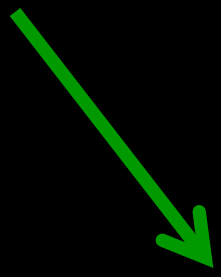
Without Treatment	With Treatment	
0	0	NO EFFECT
0	1	HARM
1	0	BENEFIT
1	1	NO EFFECT

Clinical Trial: "Box Score"

0 = alive
1 = dead

Without Treatment	With Treatment
0	1
0	0
0	0
1	0
1	0
1	1
1	1
1	1
0	1
1	1
1	1
1	0
0	0
1	1
0	0
0	0
1	1
0	0
0	0
1	0

BENEFIT



Proportion Dead	11/20 (55%)	9/20 (45%)

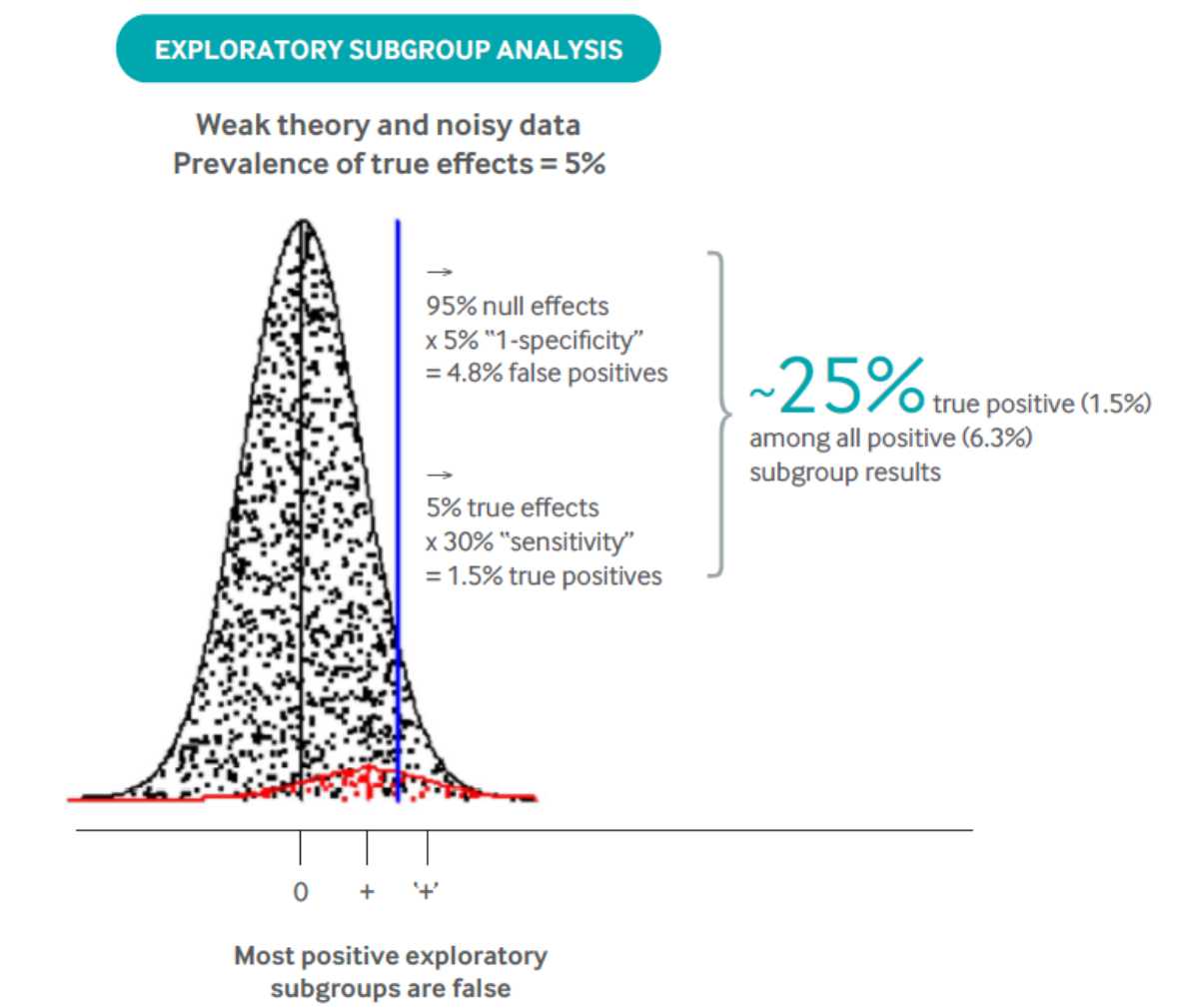
Conventional Approach: One variable at a time subgroup analysis

- This approach relies on the detection of "statistically significant" effect modifiers—contrasting *relative* treatment effects across levels of each subgrouping variable, one-variable-at-a-time.

Problems With Conventional Subgroup Analysis

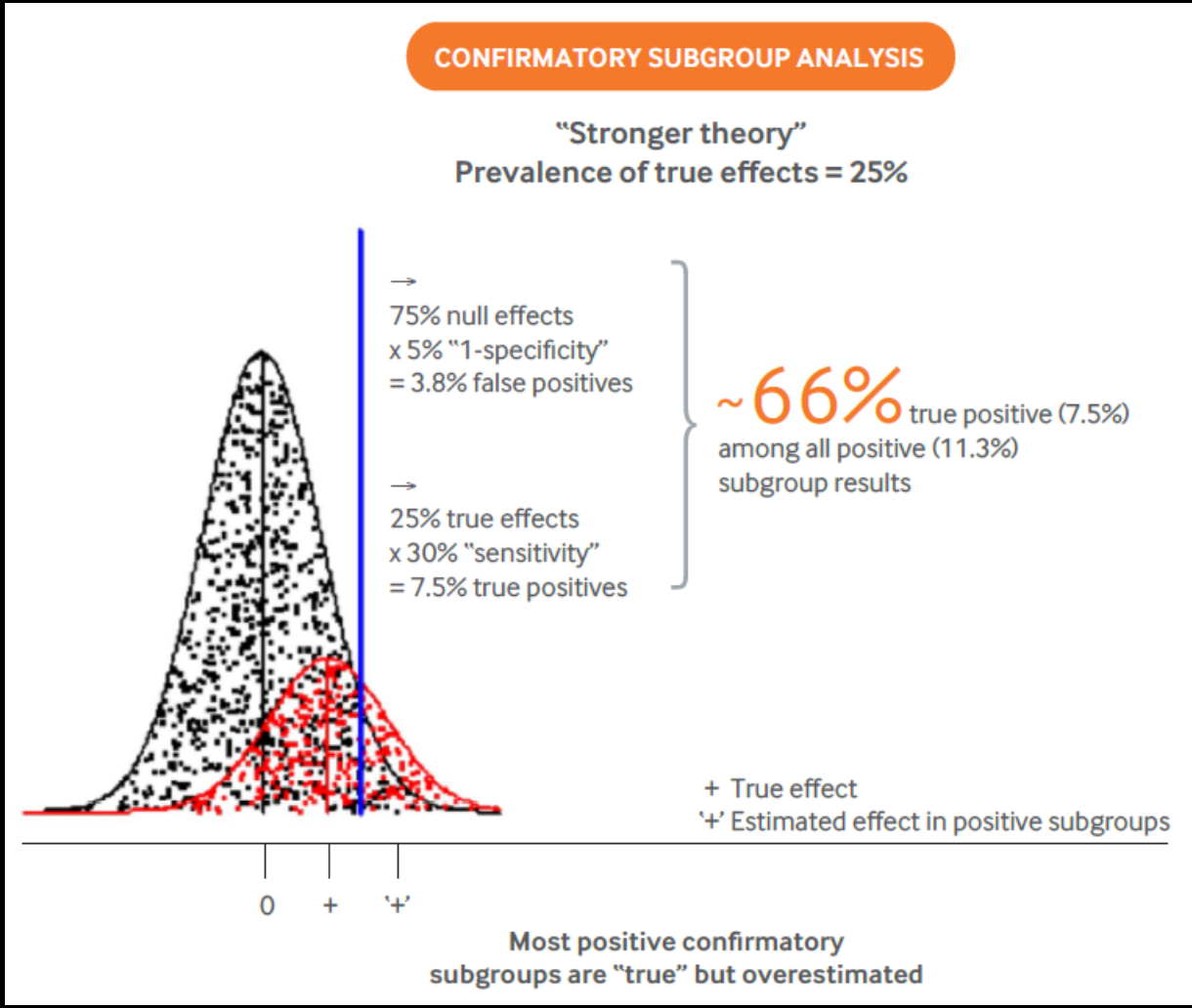
- Fail to detect HTE even when its there:
  - Low power
    - For covariates with 50% prevalence (e.g. gender): ~4 fold the sample size.
    - For covariates with 20% prevalence (e.g. common comorbidities): ~10 fold the sample size.
  - Compare Groups of patients that are more similar than dissimilar.
- Individuals patients belong to many different subgroups.
- Spurious False Positives

Why most subgroup effects are false or overestimated



Kent DM, Steyerberg E, van Klaveren D. BMJ 2018 (In Press).

Why most subgroup effects are false or overestimated



Kent DM, Steyerberg E, van Klaveren D. BMJ 2018 (In Press).

"Positive" subgroup analyses subsequently shown to be false

Observation	Refutation
Aspirin is ineffective in secondary prevention of stroke in women[29,30]	31

Rothwell PM. *Lancet* 2005;365(9454):176-86.

Interim Summary

- Determining the best treatment on average (the task of an RCT) is very different from determining the best treatment for an individual (the task of a good clinician) .
- Conventional subgroup analysis of clinical trials are typically inadequate and can also be misleading, and are not consistent with the clinically most important goal of *prediction*.

- The reference class problem is a model selection problem
- The goal of Personalized EBM can be conceived as the identification of an optimal subgrouping scheme, based on all relevant patient characteristics, that yields a more individualized reference class for each patient than the overall trial results.

Why Risk Based Subgroup Analysis Should be Routine

David M. Kent, MD, MSc
Professor of Medicine, Neurology, Clinical and Translational Science,
Director, Predictive Analytics and Comparative Effectiveness (PACE) Center,
Institute for Clinical Research and Health Policy Studies, Tufts Medical Center

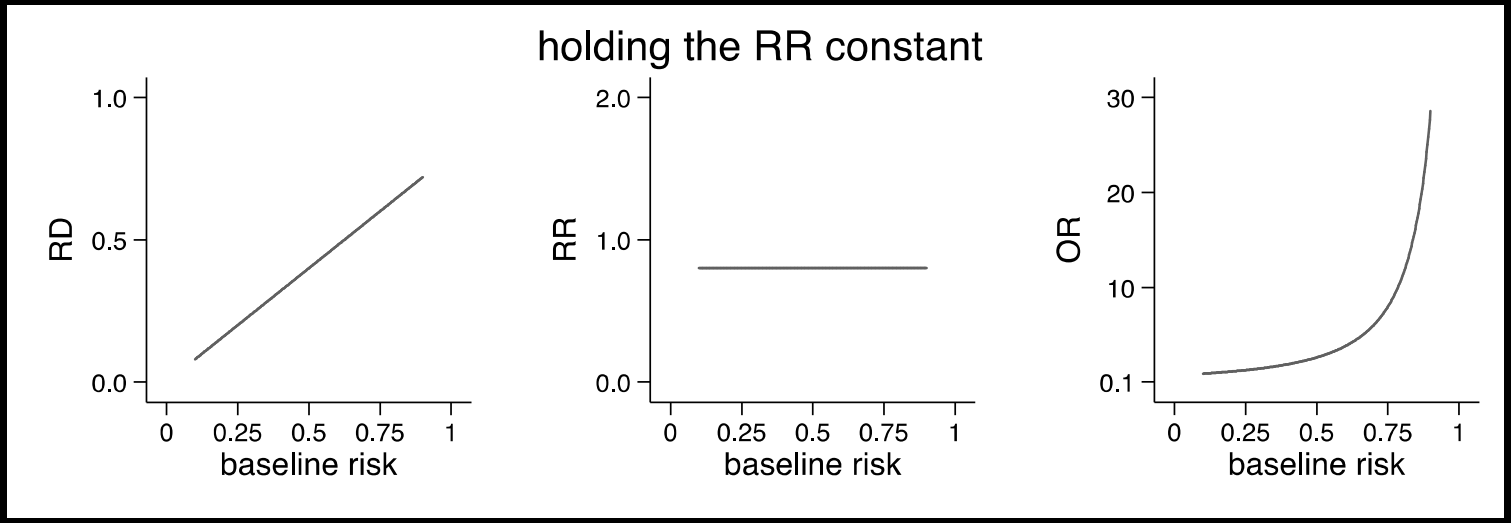
Why privilege risk-based HTE analysis?

- Risk is a known mathematical determinant of treatment effect.

Common Measures of Treatment Effect

Risk Reduction (RR)	Definition
Absolute RR	EER-CER
Relative RR	$1 - \frac{EER}{CER}$
Odds Ratio	$\frac{EER/(1-EER)}{CER/(1-CER)}$
CER=control event rate EER=experimental event rate	

An Illustration of Scale Dependence of HTE over Baseline Outcome Risk



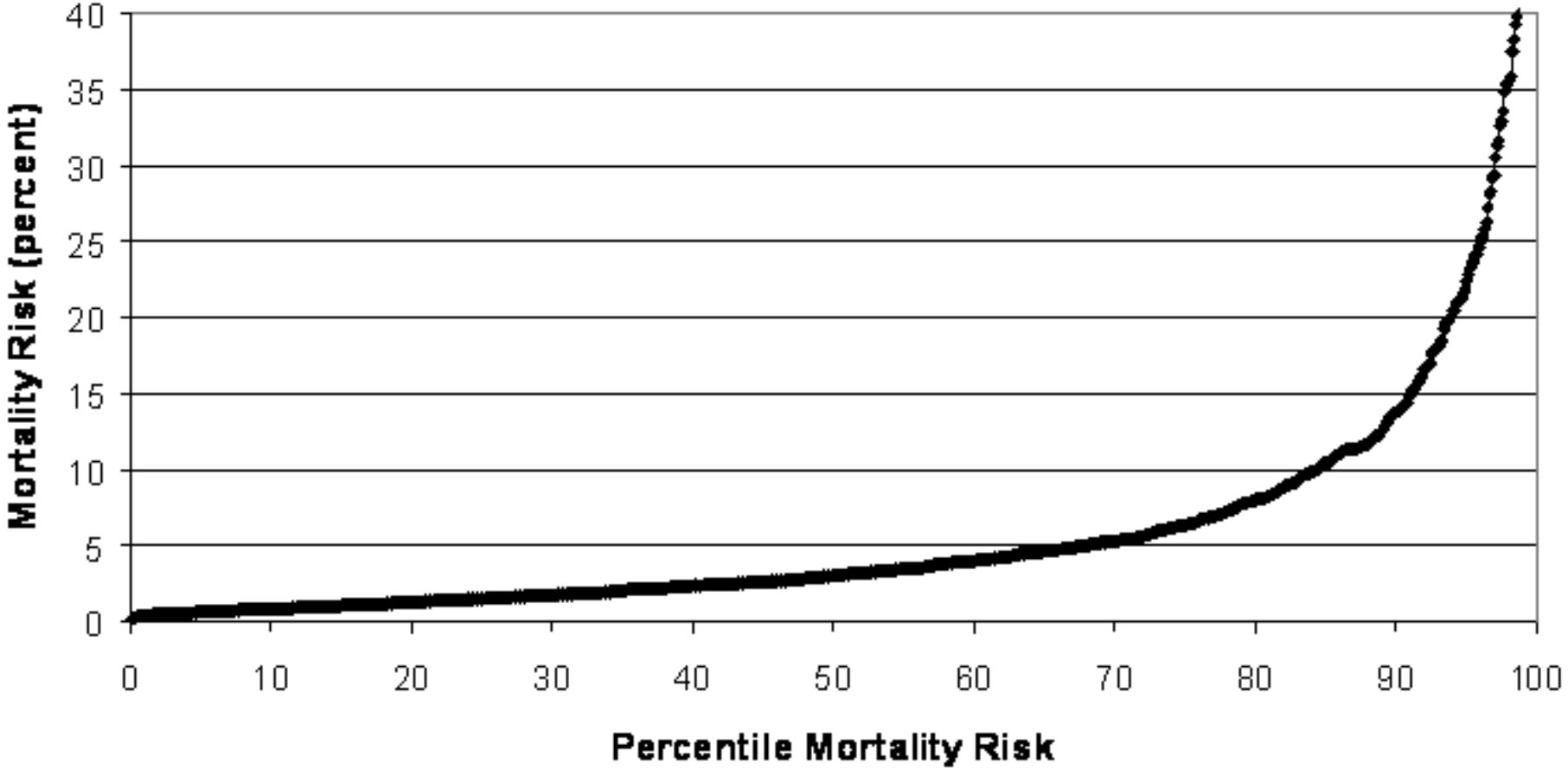
Dahabreh, Hayward, Kent, IJE, 2016

PACE

Why privilege risk-based HTE analysis?

- Risk is a known mathematical determinant of treatment effect.
- When baseline risk heterogeneity is present (and the treatment effect is non-zero), there is always HTE.
- Risk provides a summary measure that takes into account multiple variables that are relevant; provides "patient-centered" evidence.

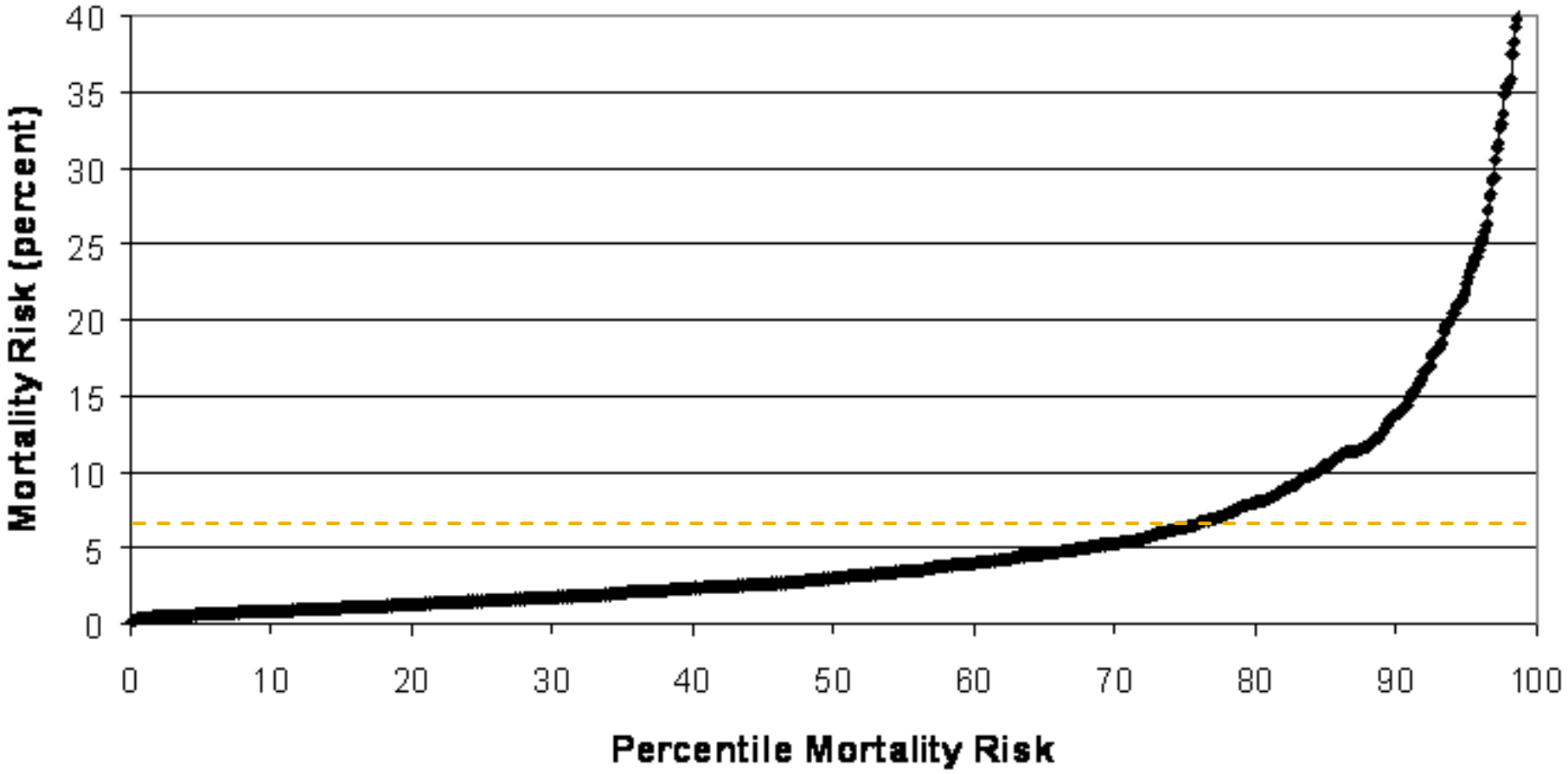
Figure 1: Distribution of Mortality Risk with Thrombolytic Thearpy in Patients with Acute Myocardial Infarction



PACE

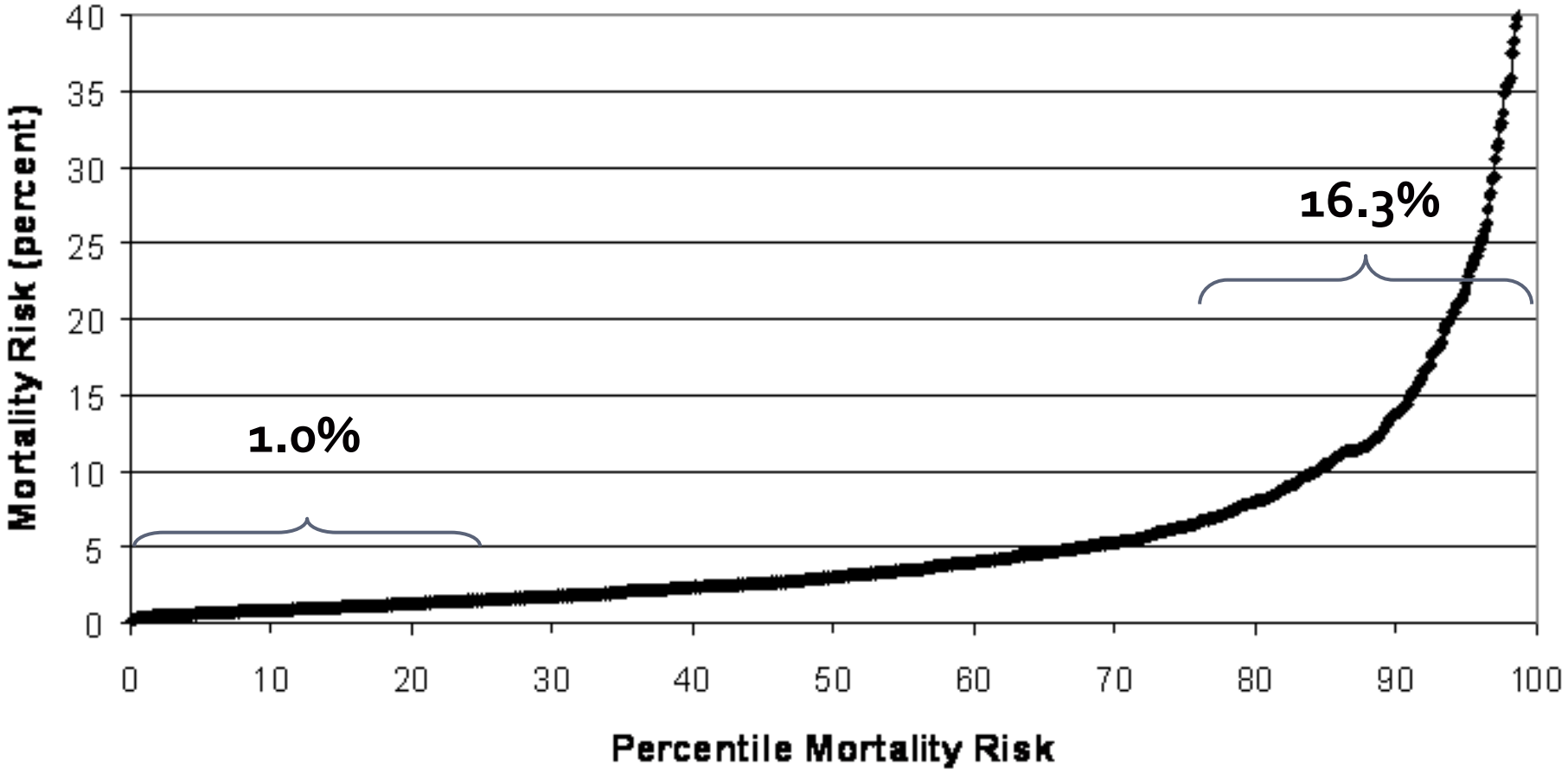
Kent DM, et al. *J Gen Intern Med* 2002; 17:887-94.

Figure 1: Distribution of Mortality Risk with Thrombolytic Thearpy in Patients with Acute Myocardial Infarction



Kent DM, et al. *J Gen Intern Med* 2002; 17:887-94.

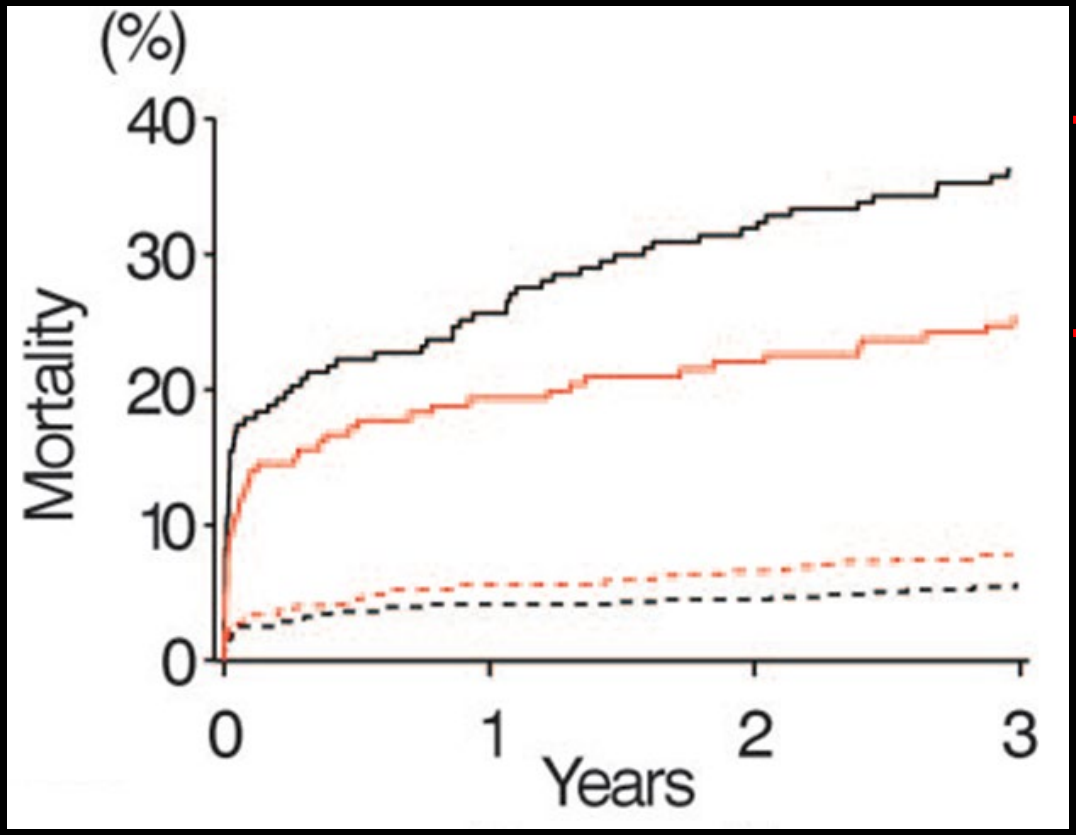
Figure 1: Distribution of Mortality Risk with Thrombolytic Thearpy in Patients with Acute Myocardial Infarction



PACE

Kent DM, et al. *J Gen Intern Med* 2002; 17:887-94.

DANAMI-2



High Risk

Low Risk

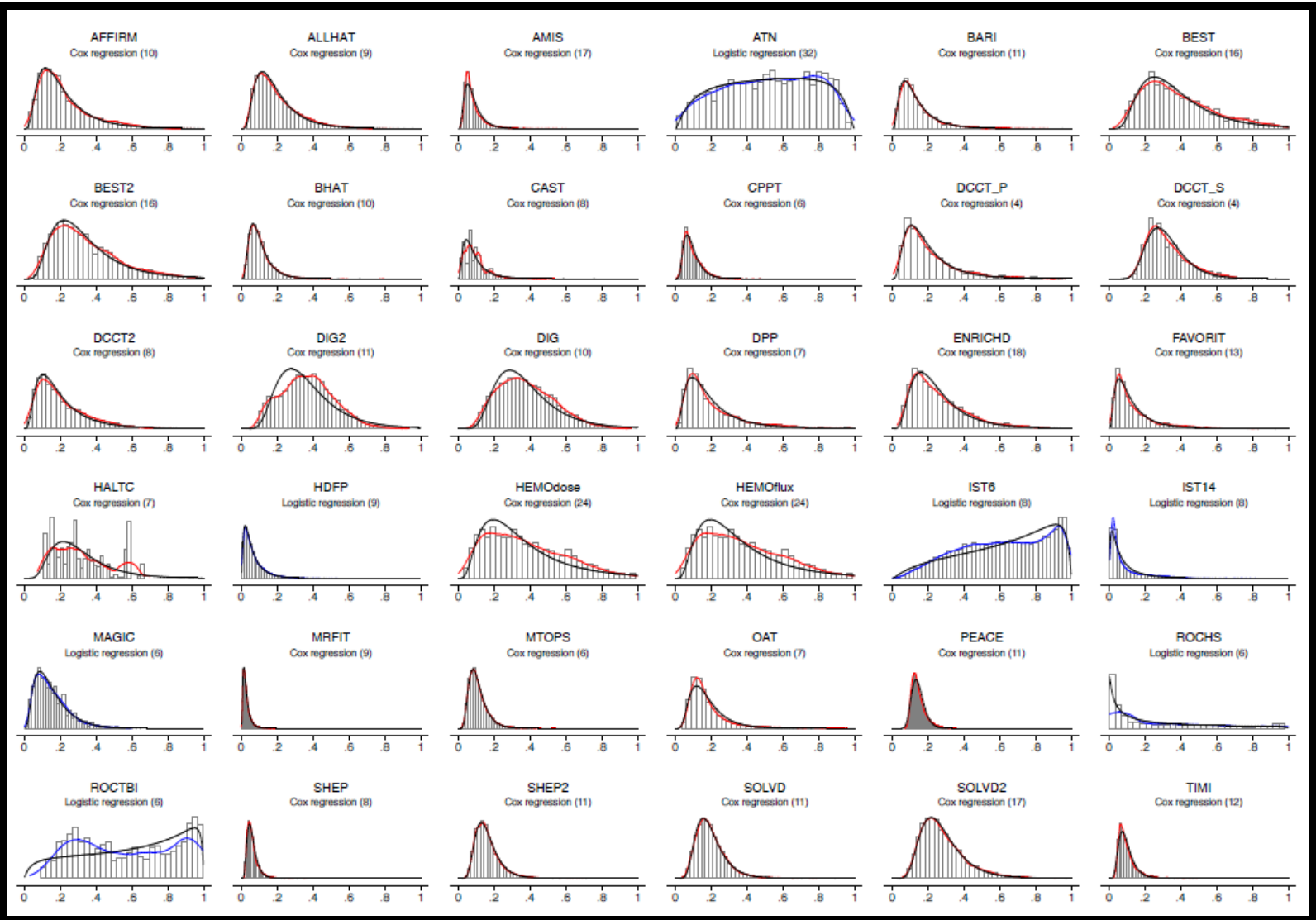
Number at risk				
TIMI 0-4	Fx	556	533	531
	PA	578	546	540
TIMI ≥ 5	Fx	207	154	141
	PA	186	150	145

PCI

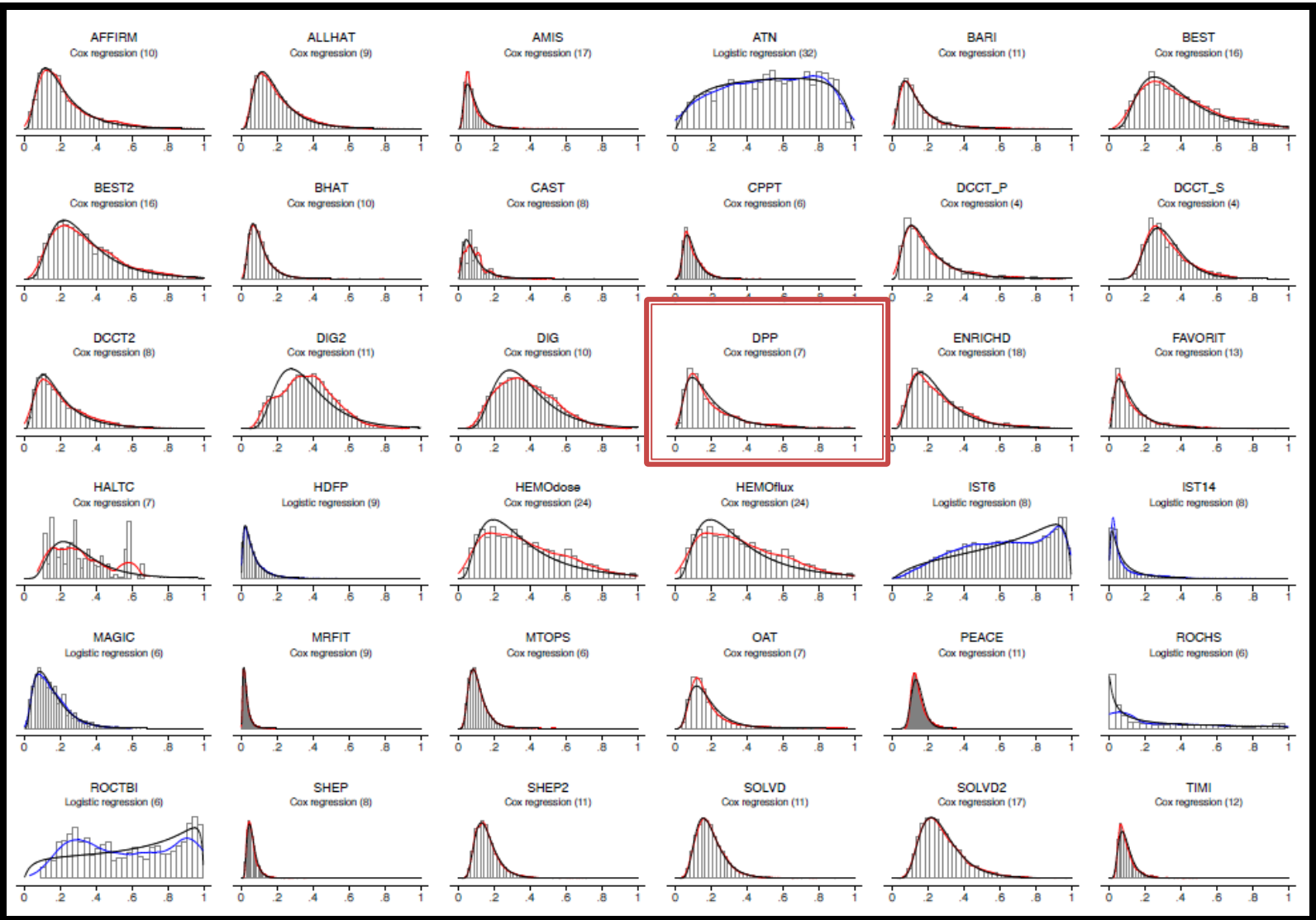
Medical Therapy

Thune JJ, et al. *Circulation* 2005,112:2017-2021.

Predicted Risk Distributions in RCTs



Predicted Risk Distributions in RCTs

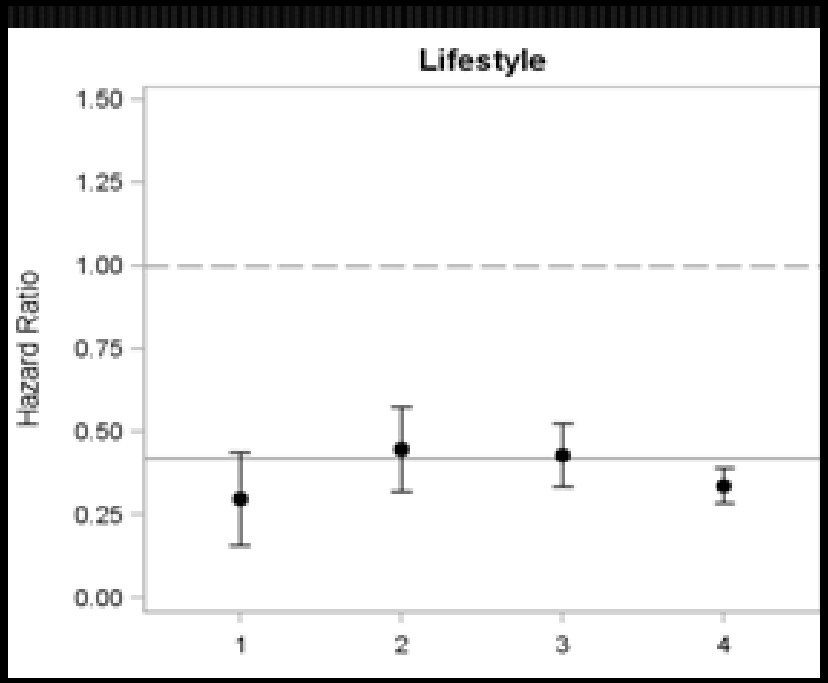


Diabetes Prevention Program (DPP) Randomized Controlled Trial

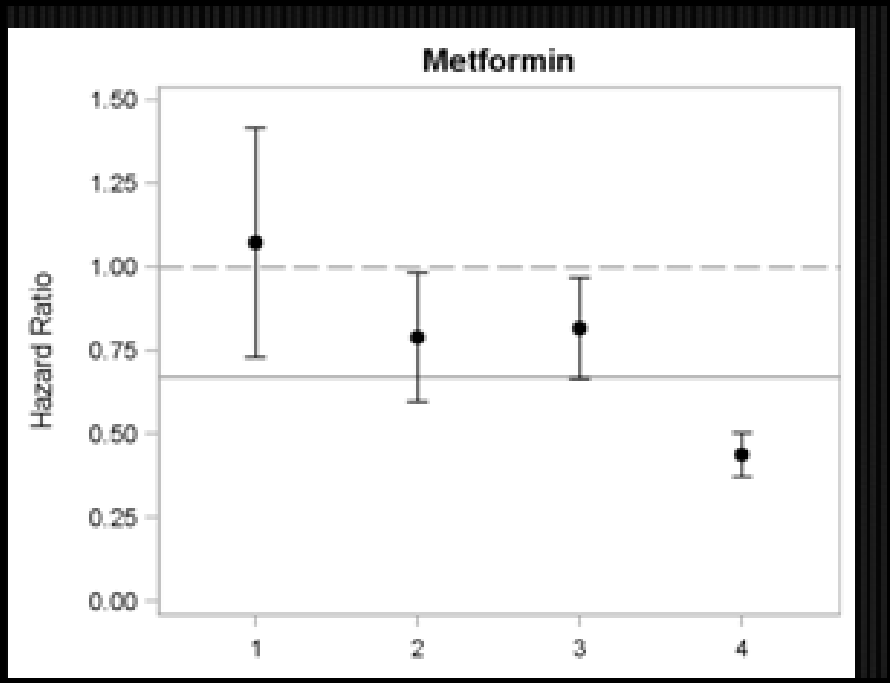
- Participants: 3060 nondiabetic persons with evidence of impaired glucose metabolism.
- Intervention: Intervention groups received metformin or a lifestyle-modification program.
- Main Outcome Measure: Development of diabetes

The DPP study was conducted by the DPP Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

DPP Risk Stratified Results

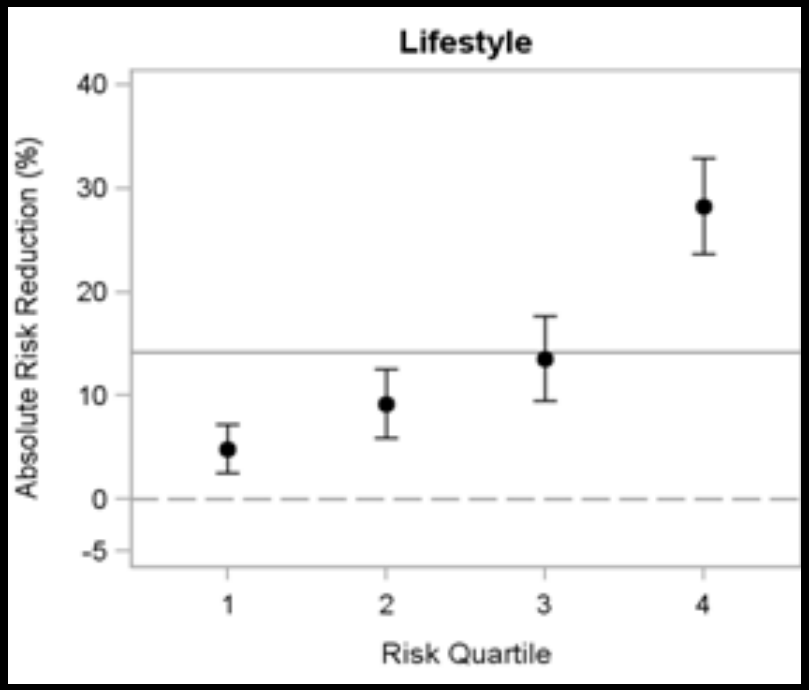


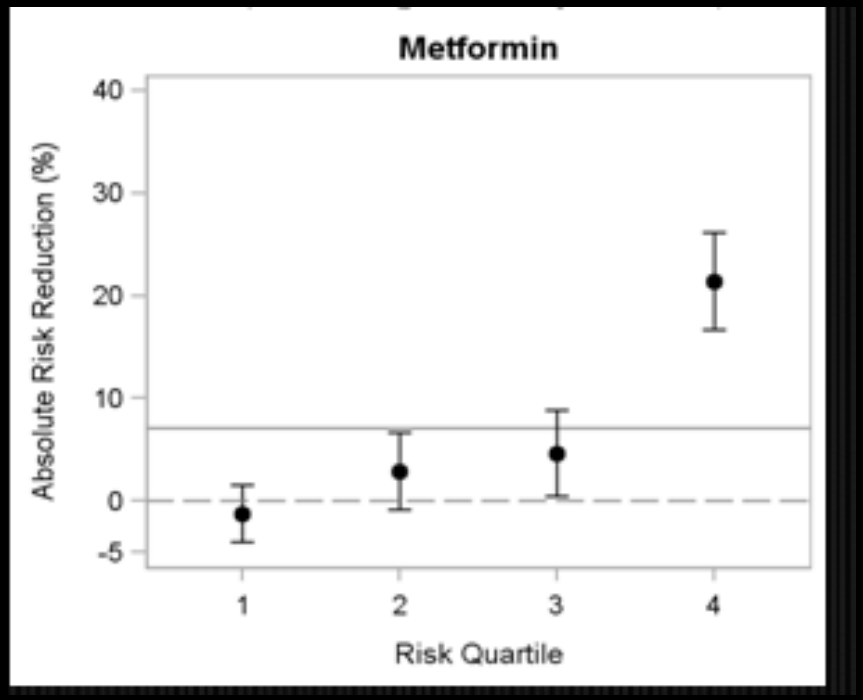
p value = NS



p value = 0.0008

DPP Risk Stratified Results





Risk based analyses can reveal counter-intuitive findings

- Overall effectiveness results may be driven by a relatively small group of influential (typically high risk) patients;
- The typical (median) risk patient is frequently at considerably lower risk than the overall average;
- The average benefit seen in the summary result often over estimates the benefit (on the RD scale) in most patients (and may obscure harm in many).

METHODOLOGY Open Access

Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal

TRIALS

David M Kent[1*], Peter M Rothwell[2], John PA Ioannidis[1,3], Doug G Altman[4], Rodney A Hayward[5]

1. Evaluate and **report on the distribution of risk** in the overall study population and in the separate treatment arms of the study by using a risk prediction model or index.
2. Primary subgroup analyses should include reporting how relative and absolute risk reduction varies in a **risk-stratified analysis**.
3. Any additional **primary subgroup analysis should be pre-specified** and limited to patient attributes with strong a prior pathophysiological or empirical justification.
4. Conduct and **report on secondary (exploratory) subgroup analyses separate** from primary subgroup comparisons.
5. All analyses conducted must be reported and statistical testing of HTE should be done using **appropriate methods** (such as interaction terms) and avoiding over-interpretation.


PACE
thebmj
STATE OF THE ART REVIEW
Annals of Internal Medicine
The Predictive Appro
Statement: Explan
David M. Kent, MD, MS; David v
Steve Goodman, MD, MHS, PhD
Michael Pencina, PhD; Gowri
Andrew Vickers, PhD; John
The PATH (Predictive Appr
neity) Statement was deve
provide guidance for, p
treatment effects (HTE) in clini
HTE analysis is to provide patient-cent
risk with versus without the intervention, tak
relevant patient attributes simultaneously, to suppo
sonalized clinical decision making than can be made
basis of only an overall average treatment effect. The author
distinguished 2 categories of predictive HTE approaches (a "risk-
modeling" and an "effect-modeling" approach) and developed
4 sets of guidance statements: criteria to determine when risk-
modeling approaches are likely to identify clinically meaningful
THE LEARNING HEALTH SYSTEM SERIES
CARING FOR THE INDIVIDUAL PATIENT
Understanding Heterogeneous Treatment Effects
David Kent, Jessica Paulus, Mahnoor Ahmed, and Danielle Whicher, Editors
NATIONAL ACADEMY OF MEDICINE
Annals of
The Pre
(PATH
David M.
Steve Go
RTING METHODS
rogeneity
ck-Lake, MFS; Sally Morton, PhD;
MSPH; Ravi Varadhan, PhD;
pplied to disaggregate patients
ed variation in benefit, and an
herein a model is developed on
m for treatment assignment and
and baseline covariates. Both
dict differential absolute treat-
ale for clinical decision making.
methods, consid- guidance: criteria to determine
and considerations are likely to identify clinically
proaches. They dis- pects of risk-modeling meth-
enumerate research to clinical practice, and con-
signed to generate se of effect-modeling ap-
planation and elabora- gether with its explanation
and rationale of each ide future analyses and re-
lated analytic consider-
Annals.org
Annals.org
November 2019.
g on 12 November 2019.
For author
This article was
37

Risk modeling vs. Effect modeling

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement: Explanation and Elaboration

Table 2. Equations Corresponding to Risk-Modeling and Effect-Modeling Approaches

Risk modeling

A multivariable regression model f that predicts the risk for an outcome based on risk predictors x_i is identified or developed:

Equation 1: $risk = f(\alpha + \beta_1 \times x_1 + \ldots + \beta_p \times x_p)$

Variation in the treatment effect across risk can be tested statistically on the relative scale through the interaction between a linear predictor of risk ($lp = \beta_1 \times x_1 + \ldots + \beta_p \times x_p$) and treatment assignment tx:

Equation 2: $risk = f(\alpha + \beta_{tx} \times tx + \beta_{lp} \times lp + \delta_{lp} \times lp \times tx)$

Including a treatment interaction with the linear predictor of risk permits the relative treatment effect to vary linearly across levels of risk (and permits testing of the statistical significance of this interaction effect, δ_{lp}).

When relative effects across risk strata seem constant, a model with a constant treatment effect may suffice:

Equation 3: $risk = f(\alpha + \beta_{tx} \times tx + \beta_1 \times x_1 + \ldots + \beta_p \times x_p)$,

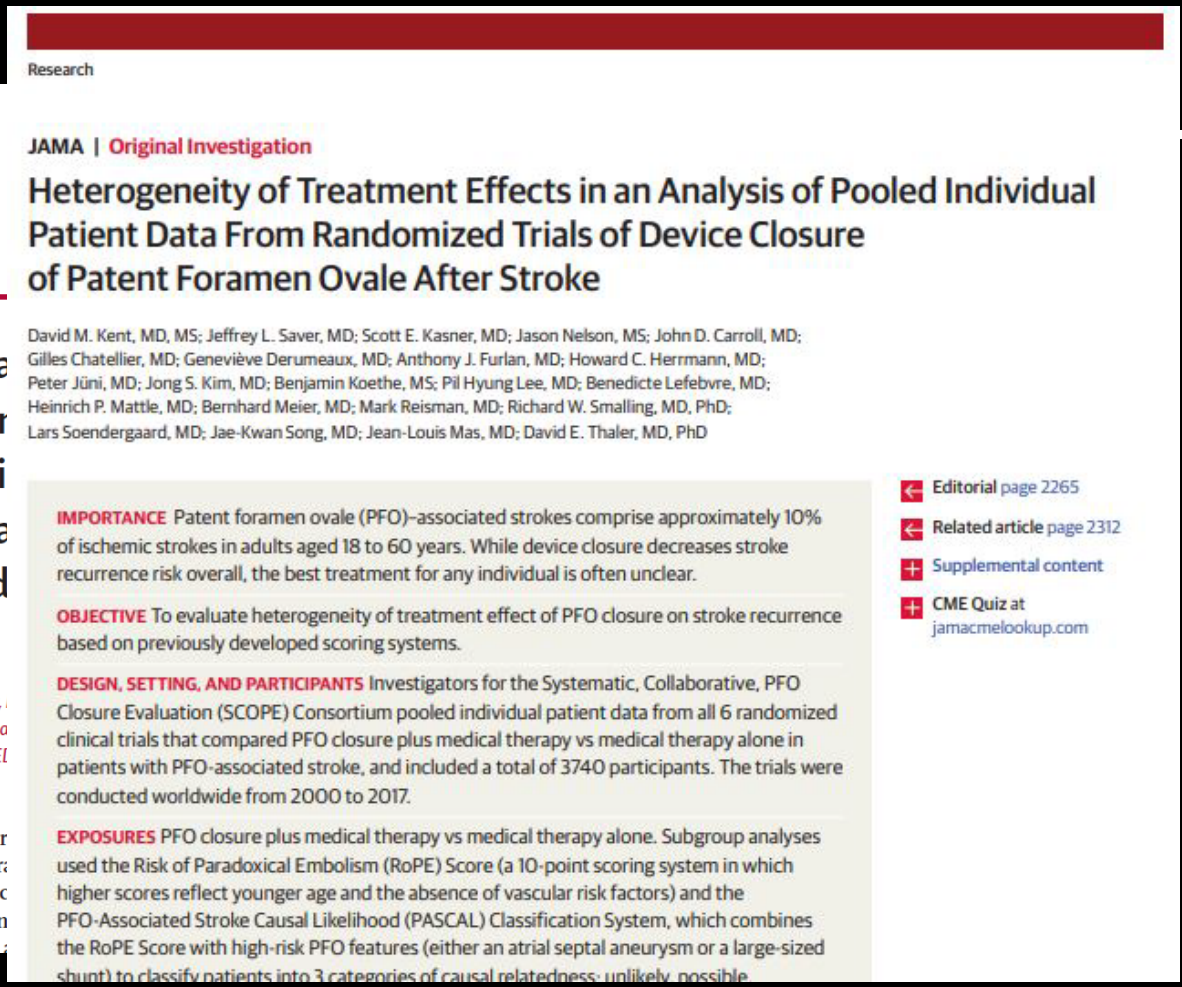
where the parameter β_{tx} represents a constant risk reduction on the log hazard or log odds scale for treated ($tx = 1$) versus control ($tx = 0$) patients.

Effect modeling

A regression model f is developed on RCT data with inclusion of risk predictors x_i, a treatment assignment variable tx, and potential treatment interaction terms ($x_i \times tx$):

Equation 4: $risk = f(\alpha + \beta_{tx} \times tx + \beta_1 \times x_1 + \ldots + \beta_p \times x_p + \delta_t \times x_1 \times tx + \ldots + \delta_p \times x_p \times tx)$

RCT = randomized controlled trial.



Research

JAMA | Original Investigation

Heterogeneity of Treatment Effects in an Analysis of Pooled Individual Patient Data From Randomized Trials of Device Closure of Patent Foramen Ovale After Stroke

David M. Kent, MD, MS; Jeffrey L. Saver, MD; Scott E. Kasner, MD; Jason Nelson, MS; John D. Carroll, MD; Gilles Chatellier, MD; Geneviève Derumeaux, MD; Anthony J. Furlan, MD; Howard C. Herrmann, MD; Peter Jüni, MD; Jong S. Kim, MD; Benjamin Koethe, MS; Pil Hyung Lee, MD; Benedicte Lefebvre, MD; Heinrich P. Mattle, MD; Bernhard Meier, MD; Mark Reisman, MD; Richard W. Smalling, MD, PhD; Lars Soendergaard, MD; Jae-Kwan Song, MD; Jean-Louis Mas, MD; David E. Thaler, MD, PhD

IMPORTANCE Patent foramen ovale (PFO)–associated strokes comprise approximately 10% of ischemic strokes in adults aged 18 to 60 years. While device closure decreases stroke recurrence risk overall, the best treatment for any individual is often unclear.

OBJECTIVE To evaluate heterogeneity of treatment effect of PFO closure on stroke recurrence based on previously developed scoring systems.

DESIGN, SETTING, AND PARTICIPANTS Investigators for the Systematic, Collaborative, PFO Closure Evaluation (SCOPE) Consortium pooled individual patient data from all 6 randomized clinical trials that compared PFO closure plus medical therapy vs medical therapy alone in patients with PFO-associated stroke, and included a total of 3740 participants. The trials were conducted worldwide from 2000 to 2017.

EXPOSURES PFO closure plus medical therapy vs medical therapy alone. Subgroup analyses used the Risk of Paradoxical Embolism (RoPE) Score (a 10-point scoring system in which higher scores reflect younger age and the absence of vascular risk factors) and the PFO-Associated Stroke Causal Likelihood (PASCAL) Classification System, which combines the RoPE Score with high-risk PFO features (either an atrial septal aneurysm or a large-sized shunt) to classify patients into 3 categories of causal relatedness: unlikely, possible

Editorial page 2265
Related article page 2312
Supplemental content
CME Quiz at jamacmelookup.com

Redevelopment and va individualise decision r surgical revascularisati artery disease: seconda randomised controlled validation

Kuniaki Takahashi, Patrick W Serruys, Valentin Fuster, Jung-Min Ahn, Arie Pieter Kappetein, Stuart J Head, Da David van Klaveren, on behalf of the SYNTAXES, FREEL

Summary

Background Randomised controlled trials ar interventions, and typically report the avera vary between patients, basing treatment dec be suboptimal. We aimed to develop an in strategy in patients with complex coronary

Summary

- Heterogeneity of outcome risk is ubiquitous.
- Heterogeneity of outcome risk inevitably gives rise to heterogeneity of treatment effect.
- One variable at a time subgroup analyses are inadequate (and prone to spurious false positive results).
- Risk based subgroup analyses can do better.