



Universiteit
Leiden
The Netherlands

The predictive approaches to treatment effect heterogeneity (PATH) statement

Kent, D.M.; Paulus, J.K.; Klaveren, D. van; D'Agostino, R.; Goodman, S.; Hayward, R.; ... ; Steyerberg, E.W.

Citation

Kent, D. M., Paulus, J. K., Klaveren, D. van, D'Agostino, R., Goodman, S., Hayward, R., ... Steyerberg, E. W. (2020). The predictive approaches to treatment effect heterogeneity (PATH) statement. *Annals Of Internal Medicine*, 172(1), 35-45. doi:10.7326/M18-3667

Version: Accepted Manuscript

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3185130>

Note: To cite this publication please use the final published version (if applicable).



Published in final edited form as:

Ann Intern Med. 2020 January 07; 172(1): 35–45. doi:10.7326/M18-3667.

The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement

David M. Kent, MD, MS¹, Jessica K. Paulus, ScD¹, David van Klaveren, PhD^{1,12}, Ralph D'Agostino, PhD², Steve Goodman, MD, MHS, PhD³, Rodney Hayward, MD⁴, John P.A. Ioannidis, MD, DSc³, Bray Patrick-Lake, MFS⁵, Sally Morton, PhD⁶, Michael Pencina, PhD⁵, Gowri Raman, MBBS, MS⁷, Joseph S. Ross, MD, MHS⁸, Harry P. Selker, MD, MSPH^{9,10}, Ravi Varadhan, PhD¹¹, Andrew Vickers, PhD¹³, John B. Wong, MD¹⁴, Ewout W. Steyerberg, PhD¹²

¹Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies (ICRHPS), Tufts Medical Center, Boston, MA, USA

²Department of Biostatistics and Epidemiology, Boston University, Boston, MA, USA ³Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

⁴Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

⁵Duke Clinical Research Institute (DCRI), Duke University, Durham, NC, USA ⁶Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA ⁷Center for Clinical Evidence Synthesis, ICRHPS, Tufts Medical Center, Boston, MA, USA

⁸Section of General Medicine, Department of Internal Medicine, School of Medicine and Department of Health Policy and Management, School of Public Health, Yale University, New Haven, CT, USA

⁹Center for Cardiovascular Health Services Research, ICRHPS, Tufts Medical Center, Boston, MA, USA ¹⁰Tufts Clinical and Translational Science Institute, Boston, MA, USA ¹¹Center on Aging and Health, Johns Hopkins University, Baltimore, MD, USA

¹²Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands ¹³Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

¹⁴Division of Clinical Decision Making, Tufts Medical Center, Boston, MA, USA

Abstract

Heterogeneity of treatment effects (HTE) refers to the non-random variation in the magnitude or direction of a treatment effect across levels of a covariate, as measured on a selected scale, against a clinical outcome. In randomized controlled trials (RCTs), it is typically examined through subgroup analysis contrasting effects in groups of patients defined one-variable-at-a-time (e.g. males versus female; old versus young). Herein, we present guidance on an alternative approach to HTE analysis, predictive HTE analysis. The goal of predictive HTE analysis is to provide patient-centered estimates of outcome risks with versus without the intervention, taking into account all relevant patient attributes simultaneously. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement was developed using: a multidisciplinary technical expert panel;

Corresponding Author: David M Kent, MD, MS, Director, Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington St, Box 63, Boston, MA 02111, dkent1@tuftsmedicalcenter.org.

targeted literature reviews; simulations to characterize potential problems with predictive approaches; and a deliberative process engaging the expert panel. We distinguished two categories of predictive HTE approaches: 1) a “risk modeling” approach wherein a multivariable model predicts the risk of an outcome and is applied to disaggregate patients within RCTs to define risk-based variation in benefit; and 2) an “effect modeling” approach wherein a model is developed on RCT data by incorporating a term for treatment assignment and interactions between treatment and baseline covariates. Both approaches can be used to predict differential absolute treatment effects, the most relevant scale for clinical decision-making. We developed four sets of guidance: 1) criteria to determine when risk modeling approaches are likely to identify clinically important HTE; 2) methodological aspects of risk modeling methods; 3) considerations for translation to clinical practice; and 4) considerations and caveats in the use of effect modeling approaches. The PATH Statement, together with the PATH Statement Explanation and Elaboration document, may guide future analyses and reporting of RCTs.

Keywords

personalized medicine; predictive analytics; RCTs; heterogeneity of treatment effect; evidence based medicine; subgroups

Introduction

Medical treatment decisions by clinicians and patients are generally based—implicitly or explicitly—on predictions of outcomes under alternative treatment conditions. Under the paradigm of evidence-based medicine (EBM), the results of randomized controlled trials, singly or aggregated in meta-analysis, are the primary evidence used to support these predictions.

Popular approaches to EBM have encouraged the direct application of summary trial results to guide decision making for individuals, as though all patients meeting trial enrollment criteria are similarly likely to experience the benefits and harms of treatments. Yet there has been growing recognition that patients enrolled in trials typically differ from one another in many ways that might be relevant; in particular they can differ substantially in their risks of the outcome and in the balance of the benefits and harms of treatment. Thus, there is also growing recognition of the limitations of summary randomized controlled trial (RCT) results as tools for prediction for individualized clinical decision making, even among trial-eligible patients.^{1–3}

Understanding how a treatment’s effect can vary across patients,^{4–10} a concept described as heterogeneity of treatment effects (HTE), is central to the research agenda for both personalized (or precision) medicine and comparative effectiveness research. We define HTE as non-random variation in the magnitude or direction of a treatment effect across levels of a covariate (i.e. a patient attribute or set of attributes) against a clinical outcome. Since treatment effect can be measured in different scales (e.g. absolute risk difference, relative risk reduction), HTE is fundamentally a scale dependent concept (i.e. its presence or absence depends on which scale the effect is measured)¹¹.

There is extensive literature on conventional subgroup analyses, which serially divide the trial population into groups (e.g., male versus female, old versus young) and examine contrasts in relative treatment effects.^{12–22} Although potentially useful for exploring hypotheses about factors that modify a treatment effect, there are important limitations of these “one-variable-at-a-time” analyses. Briefly, low statistical power, multiplicity, and weak prior theory on relative effect modifiers make subgroup analyses prone to both false negative and false positive results.^{23–25} They also do not provide patient-centered treatment effect estimates, as patients have many attributes that simultaneously affect the outcome of interest and the benefits of treatment. The well-appreciated limitations of subgroup analysis have reinforced the reliance on summary trial results for clinical decision-making and the false impression that harm-benefit trade-offs are similar for all patients meeting trial enrollment criteria.^{3,6}

Predictive approaches to HTE analysis are designed to address some of the above limitations. The goal of predictive HTE analysis is to provide individualized predictions of treatment effect, specifically defined by the difference between expected potential outcome(s) of interest in a particular patient with one intervention versus an alternative—taking into account multiple relevant characteristics simultaneously.^{7,26} While there is expanding guidance regarding optimal approaches to prediction modeling (e.g., PROGRESS,^{27,28} TRIPOD²⁹) and some guidance for HTE analysis (e.g., Patient-Centered Outcomes Research Institute [PCORI] Standards for HTE³⁰), to our knowledge there are no consensus guidelines that address specific methodologic issues for predictive HTE analysis. Herein, we aim to provide guidance to predictive HTE analyses.

The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement outlines principles, criteria, and key considerations for applying predictive HTE approaches to clinical trials to provide patient-centered evidence in support of decision making. The primary focus of this effort is the identification of clinically important HTE—i.e. variation in the risk difference across patient subgroups potentially sufficient to span clinically-defined decision thresholds.^{6,9,26} In this article, we summarize criteria to determine when risk modeling approaches are likely to identify clinically important HTE, review critical methodological aspects of these methods, and identify considerations and caveats for translation to clinical practice. We focus on modeling strategies that use regression analysis, but acknowledge a wide set of other approaches (e.g., tree methods, machine learning). The PATH Statement is intended to be used in conjunction with the explanation and elaboration (E&E) document³¹, which expands the intent of each recommendation; describes the rationale for each; and discusses related analytic considerations, caveats, and reservations.

Predictive HTE Approaches: Risk Modeling versus Effect Modeling

The main goal of predictive HTE analysis is to develop models that can predict which of two or more treatments will be better for a particular patient. We distinguish predictive HTE analyses from HTE analyses with other goals, such as those more focused on causal interaction—including both exploratory (hypothesis generating) and confirmatory (hypothesis testing) subgroup analysis. We direct the reader to the E&E for a fuller

discussion of differing concepts of “interaction” and HTE, and to the glossary for key definitions, as well as a recent narrative review.²⁶⁾

Predictive HTE analysis generally comprises two steps: 1) variable and model selection to define the reference class (or subgrouping) scheme; and 2) effect estimation across different strata of that scheme. Following a previous review²⁶, we distinguish two distinct approaches to predictive HTE analysis.²⁶ The first approach is a “risk modeling” approach in which first, a multivariable model that predicts the risk of an outcome is applied to stratify patients within trials to examine risk-based variation in treatment effects. In the second approach, “effect modeling,” a model is developed on RCT data with inclusion of a treatment assignment variable, and potential inclusion of treatment interaction terms. Both approaches can be used to predict differential absolute treatment effects, i.e., a difference in outcome risks under two alternative treatments. We describe each briefly below.

Risk Modeling—Risk modeling relies on the mathematical dependency of treatment effect on the control event rate (CER) (Table 1), an observable proxy for outcome risk. When risk is described through a combination of factors,³² the CER will typically vary considerably across the trial population. The absolute risk difference (RD)—the most clinically important effect measure—will generally vary across risk strata even if the relative risk is the same.³² When there is substantial variation in outcome risk across a trial population, there are often important differences in harm-benefit trade-offs^{3;26} (Figure 1). For this reason, risk models can be useful in identifying “clinically-important HTE,” which is evaluated on the absolute risk difference scale.

A “risk modeling” approach is typically performed in two steps. First, a multivariable regression model that predicts the risk of an outcome (usually the primary study outcome) is identified from external sources (an “external model”) or developed directly on the trial population without a term for treatment assignment (an “internal model”) (Table 2, Equation 1). This model is then applied to stratify patients within trials and examine risk-based variation in treatment effects.

Figure 1 shows a schematized (Figure 1A) and an actual (Figure 1B) example of the risk modeling approach to trial analysis, in which both relative effects and absolute effects vary by baseline risk across the trial population. In the examples shown, patients are divided using quantiles (e.g. quartiles) of risk and treatment effect is estimated in risk groups for purposes of reporting. Figure 2 of the E & E³¹ shows an alternative form of presentation in which outcomes are shown by predicted risk itself, rather than by quantiles. Variation in the treatment effect is tested statistically on the relative scale, here using an interaction between the linear predictor for risk and treatment (Table 2, Equation 2). Results are also presented on the absolute RD scale.

In the RITA-3 trial (Figure 1B), which compared an invasive approach to a non-invasive approach for acute coronary syndrome, it can be seen that the highest risk eighth of patients had an outcome rate approximately 8-fold higher than patients in the lowest risk quarter. In these high risk patients the benefits of therapy outweigh the harms, whereas low risk patients do not benefit.

For translation to clinical practice, modeling treatment effects across the full risk spectrum (i.e. as a continuous variable) can provide more individualized predictions of treatment effect. When relative effects across risk strata appear constant, a model with a constant treatment effect may suffice (Table 2, Equation 3). Including a treatment interaction with the linear predictor of risk permits the relative treatment effect to vary linearly across levels of risk (Table 2, Equation 2). Non-linear interactions between risk and treatment may also be considered. Importantly, the exact relation between risk and treatment effect will depend on the variables that are included in the risk model.

Effect Modeling—In a second, “effect modeling,” approach, a regression model is developed directly on RCT data with inclusion of risk predictors, a treatment assignment variable, and potential inclusion of treatment interaction terms (Table 2, Equation 4). Because these models include interaction terms between treatment and baseline covariates, they are vulnerable to some of the same problems that undermine conventional subgroup analysis (low power, multiplicity, and limited prior knowledge about important effect modifiers).²⁶ While on rare occasion there are highly credible treatment effect interactions that should be taken into account when predicting effects in individuals, data-driven effect models are prone to bias due to false or exaggerated interaction effects,^{31;33} particularly when tests of statistical significance are used for the selection of interaction terms.^{33;34} Methods to address these concerns include using “penalization” to shrink model coefficients to avoid overfitting^{35;36} or using a two-step process in which one first develops the model on randomized data to define the subgrouping (or reference class) scheme and uses a second data set for the treatment effect estimation.^{37;38} Because: 1) these more flexible approaches are vulnerable to overfitting and to false discovery of promising subgroup effects (or require very large databases well powered for the detection of interaction effects); 2) practical experience with these methods is limited; and 3) robust literature comparing the rapidly emerging different approaches is lacking thus far,^{39–44} the technical expert panel (TEP) elected to limit methodological recommendations only to those circumstances where highly credible treatment effect interactions have been identified; while we recognize data-driven effect modeling as a promising area of research, we limit our comments to underscoring caveats and various concerns. For a discussion of emerging data driven effect modeling approaches, we direct readers to a recent review.⁴⁵

Developing the PATH Statement

To develop the PATH Statement criteria and considerations, we adopted an approach that combined expert opinion, review of the literature, and simulation studies (detailed elsewhere).³³

The PATH Technical Expert Panel—We assembled a panel of 16 experts that would represent various perspectives on these analyses. The PATH Technical Expert Panel was comprised of experts in HTE, prediction modeling, clinical trials, and guideline development and a patient advocate (full roster available in Appendix Table 1).

Literature Review—The Technical Expert Panel co-chairs (DMK, ES) led a literature review of important papers related to the conduct of predictive HTE analyses.²⁶ We

developed a library of relevant methodological and applied articles on the topic of predictive HTE analysis in randomized trials, and articles related to interaction testing and subgroup analysis. Additional articles were solicited from panelists and organized in an online library accessible to panelists. In addition, an Evidence Review Committee (chaired by JBW) conducted a systematic scoping review (detailed elsewhere⁴⁶) to identify methodological studies of predictive HTE analysis in RCTs that use regression methods. The goal was to generate an annotated bibliography to guide and support the development of the PATH Statement and to inform future work.

Consensus-Building Process—A modified Delphi process⁴⁷ was used to build consensus among panelists to address the main PATH Statement objectives. Consensus building was facilitated through the use of ThinkTank™, a cloud-based collaborative platform.

From February to June 2018, the PATH Statement co-chairs convened five 2-hour webinar meetings and one 4-hour in-person meeting to develop consensus. The first meeting was designed to define the scope, timeline, and expectations for the PATH Statement. Two additional webinars were convened in April and July 2019 to consider alterations to recommendations in response to reviewer comments.

During each webinar, criteria were scored on a 5-point scale for Agreement, Importance, and Feasibility of assessment. A trained facilitator was present during webinar meetings to moderate discussion, structure verbal and electronic communication, and review agreement on criteria. After recording each vote, an open discussion followed that was centered on areas of disagreement. Items with a standard deviation (SD) of greater than 1 and/or mean ratings of less than 4 (on 5-point scales) were prioritized for discussion as low-consensus criteria. Following discussion, some criteria were deleted, consolidated, or revised, and panelists were given the opportunity to revise their judgments with a re-vote before the next meeting. The criteria for consensus agreement were reaching mean ratings of over 4 and SD ≤ 1 on the 5-point agreement scale after rounds of discussion and re-rating. In the week following a PATH Statement meeting, items reaching consensus and those with limited agreement were distributed to members via email for refinement. The results of the final votes on criteria and considerations and caveats are provided in Appendix Tables 2-4.

Role of the Funding Source—Development of the PATH Statement was supported through a PCORI contract, the Predictive Analytics Resource Center [SA.Tufts.PARC.OSCO.2018.01.25]. This work was also informed by a 2018 conference (“Evidence and the Individual Patient: Understanding Heterogeneous Treatment Effects for Patient-Centered Care”) convened by the National Academy of Medicine and funded through a PCORI Eugene Washington Engagement Award (1900-TMC). The funding source had no role in study design, data collection, analysis, preparation of the manuscript, or decision to submit the manuscript for publication. The views, statements, and opinions presented in this work are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors, or Methodology Committee.

The PATH Statement

The PATH Statement is comprised of four sets of guidance on the conduct of predictive HTE analyses with further explanation in the E&E document³¹ that is briefly summarized below.

First, we developed consensus criteria to define the decision-making, data, design and analytic context in which the application of risk modeling approaches is likely to yield clinically informative results (Box A). The Technical Expert Panel (TEP) agreed that risk modeling approaches should generally be applied when there is an overall treatment effect, and considered cautiously when there is not, given the potential for false positive findings. The criteria emphasized conditions where there is likely to be substantial outcome-risk heterogeneity, and where this risk heterogeneity is likely to induce important heterogeneity in benefit-harm trade-offs. This will occur when substantial and identifiable heterogeneity of outcome risk in the trial population is anticipated (i.e. a broad case mix) and when treatments are associated with a non-trivial amount of serious harm or burden. Figure 1B shows a clinical example where outcome rates vary dramatically from the high to low risk group and the presence of a small amount of treatment-related harm may be sufficient to erode much (or all) of the benefit in low risk patients. Pragmatic and data-related criteria emphasized the availability of several large, randomized, well-conducted clinical trials of contemporary interventions for individual patient meta-analysis, and prioritized cases when the variables needed for prediction are routinely available in clinical care. Two additional criteria did not reach consensus (Appendix). Prior work suggests that risk modeling might tend to be more informative when the incidence of the outcome is lower (which can lead to more heterogeneity and skewness of risk³²), but the expert panel was divided on the importance of this criterion.

In Box B, we present guidance on best methodological practices to conduct risk modeling approaches to identify HTE, including overarching guidance, identifying or developing a risk model, and applying the model and reporting results. This guidance emphasizes the application of a high-quality externally-developed risk model to stratify trial results. Alternatively, when a compatible external model is not available, an internal (or endogenous) risk model, developed on the entire trial population without a term for treatment assignment,^{48;49} can be applied. In either case, the analysis plan should be fully specified prior to examining the data. Internal models developed only on the control arm (to predict baseline risk, if untreated) are prone to inducing or exaggerating interactions between treatment and risk due to differential fit between the two treatment arms. Even slight overfitting on the control arm could bias treatment effect estimates across levels of risk.^{33;49;50} Our guidance on reporting the results of risk-stratified analyses underscores the importance of reporting serious treatment-related harms within each strata to support evaluation of strata-specific benefit-harm trade-offs and the appropriate conduct of statistical hypothesis testing.

Translation of findings from predictive approaches to HTE analyses into clinical practice is a complex topic that includes many issues, such as human factors aspects of model implementation, coping with missing patient characteristics in real world clinical practice, risk communication, and model transportability. While a detailed discussion of these challenges is beyond the scope of this project, a set of caveats and considerations for the

translation of predictive HTE analyses to clinical practice were developed and are presented in Box C. While results should be reported on both absolute and relative scales, the importance of interpreting treatment effects on the clinically-relevant scale (absolute RD rather than relative effects) was emphasized,⁵¹ as was the importance of external validation and calibration of the risk model used for trial stratification to the target population of interest.

When *highly credible* relative effect modifiers have been identified, they should be incorporated into prediction models using multiplicative treatment-by-covariate interaction terms (Box D). Credibility should be evaluated using rigorous multidimensional criteria (such as described by the Instrument for assessing the Credibility of Effect Modification ANalyses (ICEMAN tool)⁵²) and should not rely solely on statistical criteria (such as p-value thresholds). We anticipate that highly credible interactions will generally be rare, when using conventionally sized randomized clinical trials. Given the limited practical experience with data driven effect modeling (i.e. where statistical approaches are used to explore and select out effect modifiers on the relative scale), the vulnerability of these flexible approaches to falsely discovering promising subgroup effects, the growing variety of competing approaches, and the absence of a robust literature comparing the different approaches, the technical expert panel restricted their recommendations to a set of considerations and caveats for these more “aggressive” modeling techniques (i.e., those using more degrees of freedom for estimation of the differential treatment effect). These statements emphasized the need to address the potential for overfitting that can lead to signals of treatment benefit and harm even for interventions that are completely ineffective and innocuous.³³ This is a particular problem for modeling of RCT data, where power for statistical interaction of relative effects is generally limited, as are opportunities for external validation.

The PATH Explanation and Elaboration (E&E) Document

A supporting explanation and elaboration (E&E)³¹ document is provided to provide clarifications and to expand on the motivation and reservations regarding items in the PATH Statement. Each recommendation is explained in more detail and clinical applications of methods, supporting methodological evidence and caveats or limitations, are provided where relevant. The E&E document also describes special considerations for evaluating models that predict benefit. The literature reviews conducted as part of this project were used to justify the rationale for guidance statements or criteria. Development of the E&E document was completed with input from panelists after the PATH Statement criteria reached final consensus, with several face-to-face meetings, teleconferences, and iterations among the authors. Additional revisions were made after sharing the document with the whole PATH Statement group before final approval.

Discussion

The field of evidence-based medicine (EBM) has historically emphasized the results of randomized trials (and their meta-analysis) as the best evidence for clinical decision making. However, patients have multiple attributes that potentially influence the probability of the

outcome of interest and the benefits of treatment. The goal of Personalized EBM can be conceived as the identification of an optimal subgrouping (or reference class) scheme, based on all relevant patient characteristics, that yields more individualized treatment effect estimates for each patient than the average trial result, thus improving overall outcomes.²⁶ In fact, the most common definition of EBM from over 20 years ago anticipates the need to make decisions for individual patients.⁵³ Thus, the goal of personalization has been at the core of EBM since its inception, although the limitations of summary trial results in supporting this goal have been inconsistently recognized.

The PATH Statement was developed to address this gap. While the PATH Statement and E&E are supported by a substantial and growing body of evidence, and while they build on prior efforts to offer methodological guidance on predictive approaches to HTE,^{5,6} this represents the first such guidance developed with a diverse set of experts and stakeholders with differing views and perspectives that involved an iterative process of discussion, feedback, and revisions. The guidance thus aims to assist a diverse set of relevant stakeholders, including researchers, regulators, industry professionals, and guideline writing bodies.

We focused on risk modeling, and acknowledge that the more comprehensive and flexible effect modeling approach (incorporating treatment interaction terms with individual effect modifiers) holds promise—particularly in data sets that are substantially larger than conventional RCTs. Capturing the benefits of effect modeling while avoiding the potential harms of overfitting is an area of intense research interest^{35;37;38;48;54–60} and a central challenge for future study. Nevertheless, there is considerable evidence that risk modeling approaches can frequently provide clinically important insights—beyond that provided by overall trial results—that can directly improve decision making.^{39–44;61–63}

The PATH Statement focuses on regression-based prediction in randomized trials. There is a broader and evolving tool kit for data-driven approaches to predicting patient benefit, including machine learning techniques.^{45;64} As experience grows, we anticipate stronger methodological and evidentiary guidance to more fully understand the appropriate contexts and the advantages and limitations of these more flexible modeling methods. Additionally, while observational studies appear to offer many important advantages to conventional RCTs for more refined analyses (i.e., enhanced statistical power and patient heterogeneity), it is not well understood when these data may be sufficiently de-biased for reliable treatment effect or HTE determination.^{9;65} These are issues our PATH Statement identified as high research priorities.

While predictive HTE methods can often be usefully applied to individual large clinical trials,^{39–44;61–63} the PATH Statement authors recognize that fully realizing the goals of improved evidence personalization also depends on: (1) increased collaborative efforts to create pooled data substrates that are more conducive to these analyses than individual trials, and (2) the implementation of innovative trial designs, including those sampling larger and broader populations, that may enrich the heterogeneity of clinical trial populations. The PATH Statement is intended to encourage and motivate these innovations.

We also need research to better integrate clinical prediction into practice,^{66;67} to understand how to individualize clinical practice guidelines, to establish or extend reporting guidelines,²⁹ to establish new models of data ownership to facilitate data pooling,⁶⁸ and to re-engineer the clinical research infrastructure to support substantially larger clinically-integrated trials sufficiently powered to determine HTE.⁶⁹ Many recent and ongoing organizational and technical advances should enable this evolution.^{68;70–73} The collaborative work in the field of genetic epidemiology⁷⁴ may serve as a useful model for HTE prediction if we are to optimally address many of the challenges to individualizing evidence.

The PATH TEP recognizes the inherent difficulties and fundamental limitations of using group data to estimate treatment effects in individuals.⁹ In particular, individual treatment effects are inherently unobservable (in parallel arms studies) and individual patients do not have uniquely identifiable risks.^{9;26;75–77} Nevertheless, clinicians are required to make decisions one patient at a time. The PATH Statement provides guidance for analytic approaches that seek to advance our ability to provide more patient-centered treatment effect estimates. We present it as an important formative step in a long-term research effort to better personalize evidence from comparative effectiveness data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge the excellent technical support from Mark Adkins, Teddy Balan, Jeff Chang, and Dan Sjoberg, in analyses included in figures and supporting appendices. We also thank the Annals of Internal Medicine editors and reviewers, whose thoughtful feedback greatly improved this work. We thank Jennifer Lutz and Christine Lundquist for assistance with copyediting and creating exhibits.

Reference List

- (1). Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995; 345(8965):1616–1619. [PubMed: 7783541]
- (2). Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. *Lancet* 2005; 365(9455):256–265. [PubMed: 15652609]
- (3). Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007; 298(10):1209–1212. [PubMed: 17848656]
- (4). Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004; 82(4):661–687. [PubMed: 15595946]
- (5). Hayward RA, Kent DM, Vijan S, Hofer TP. Reporting clinical trial results to inform providers, payers, and consumers. *Health Aff (Millwood)* 2005; 24(6):1571–1581. [PubMed: 16284031]
- (6). Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010; 11:85. [PubMed: 20704705]
- (7). Varadhan R, Segal JB, Boyd CM, Wu AW, Weiss CO. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *J Clin Epidemiol* 2013; 66(8):818–825. [PubMed: 23651763]
- (8). Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA* 2014; 311(4):405–411. [PubMed: 24449319]

- (9). Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol* 2016; 45(6):2184–2193. [PubMed: 27864403]
- (10). Davidoff F. Can Knowledge About Heterogeneity in Treatment Effects Help Us Choose Wisely? *Ann Intern Med* 2017; 166(2):141–142. [PubMed: 27820948]
- (11). Greenland S, Lash TL, Rothman KJ. Concepts of Interaction In: Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*. 3rd ed Philadelphia: Lippincott Williams and Wilkins; 2008.
- (12). Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; 365(9454):176–186. [PubMed: 15639301]
- (13). Lagakos SW. The challenge of subgroup analyses--reporting without distorting. *N Engl J Med* 2006; 354(16):1667–1669. [PubMed: 16625007]
- (14). Hernandez AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J* 2006; 151(2):257–264. [PubMed: 16442886]
- (15). Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007; 357(21):2189–2194. [PubMed: 18032770]
- (16). Furberg CD, Byington RP. What do subgroup analyses reveal about differential response to beta-blocker therapy? The Beta-Blocker Heart Attack Trial experience. *Circulation* 1983; 67(6 Pt 2):I98–101. [PubMed: 6133654]
- (17). Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996; 88(3–4):206–207. [PubMed: 8632495]
- (18). Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355(9209):1064–1069. [PubMed: 10744093]
- (19). Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116(1):78–84. [PubMed: 1530753]
- (20). Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21(19):2917–2930. [PubMed: 12325108]
- (21). Stallones RA. The use and abuse of subgroup analysis in epidemiological research. *Prev Med* 1987; 16(2):183–194. [PubMed: 3295858]
- (22). Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. *Am Heart J* 2000; 139(6):952–961. [PubMed: 10827374]
- (23). Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey SG. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001; 5(33):1–56.
- (24). Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57(3):229–236. [PubMed: 15066682]
- (25). Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015; 351:h5651. [PubMed: 26537915]
- (26). Kent DM, Steyerberg EW, van Klaveren D. Personalized evidence-based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018; 363:k4245. [PubMed: 30530757]
- (27). Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ* 2013; 346:e5793. [PubMed: 23386361]
- (28). Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med* 2013; 10(2):e1001381. [PubMed: 23393430]
- (29). Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation* 2015; 131(2):211–219. [PubMed: 25561516]

- (30). Patient-Centered Outcomes Research Institute Methodology Committee. The PCORI Methodology Report. 2019. 2019.
- (31). Kent DM, Paulus JK, D'Agostino RB, Goodman SG, Hayward RA, Ioannidis JP, Patrick-Lake B, Morton S, Pencina M, Raman G, Ross J, Selker HP, Varadhan R, van Klaveren D, Vickers AJ, Wong JB, and Steyerberg EW The PATH Statement Explanation and Elaboration Document [Under review at *Annals of Internal Medicine*].
- (32). Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol* 2016; 1(45):2075–2088.
- (33). van Klaveren D, Balan TA, Steyerberg EW, Kent DM. Models with interactions overestimated the heterogeneity of treatment effect in simulated trials. *Journal of Clinical Epidemiology* 2018.
- (34). Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008; 19(5):640–648. [PubMed: 18633328]
- (35). Basu S, Sussman JB, Rigdon J, Steimle L, Denton BT, Hayward RA. Benefit and harm of intensive blood pressure treatment: Derivation and validation of risk models using data from the SPRINT and ACCORD trials. *PLoS Med* 2017; 14(10):e1002410. [PubMed: 29040268]
- (36). Ternes N, Rotolo F, Heinze G, Michiels S. Identification of biomarker-by-treatment interactions in randomized clinical trials with survival outcomes and high-dimensional spaces. *Biom J* 2017; 59(4):685–701. [PubMed: 27862181]
- (37). Claggett B, Tian L, Castagno D, Wei LJ. Treatment selections using risk-benefit profiles based on data from comparative randomized clinical trials with multiple endpoints. *Biostatistics* 2015; 16(1):60–72. [PubMed: 25122189]
- (38). Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics* 2011; 12(2):270–282. [PubMed: 20876663]
- (39). Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. European Carotid Surgery Trialists' Collaborative Group. *Lancet* 1999; 353(9170):2105–2110. [PubMed: 10382694]
- (40). Kent DM, Hayward RA, Griffith JL, Vijan S, Beshansky JR, Califf RM et al. An independently derived and validated predictive model for selecting patients with myocardial infarction who are likely to benefit from tissue plasminogen activator compared with streptokinase. *Am J Med* 2002; 113(2):104–111. [PubMed: 12133748]
- (41). Kent DM, Ruthazer R, Selker HP. Are some patients likely to benefit from recombinant tissue-type plasminogen activator for acute ischemic stroke even beyond 3 hours from symptom onset? *Stroke* 2003; 34(2):464–467. [PubMed: 12574561]
- (42). Yusuf S, Diener HC, Sacco RL, Cotton D, Ounpuu S, Lawton WA et al. Telmisartan to prevent recurrent stroke and cardiovascular events. *N Engl J Med* 2008; 359(12):1225–1237. [PubMed: 18753639]
- (43). Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med* 2013; 369(3):245–254. [PubMed: 23863051]
- (44). Sussman JB, Kent DM, Nelson JP, Hayward RA. Improving diabetes prevention with benefit based tailored treatment: risk based reanalysis of Diabetes Prevention Program. *BMJ* 2015; 350:h454. [PubMed: 25697494]
- (45). Lipkovich I, Dmitrienko A, D'Agostino BR Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med* 2017; 36(1):136–196. [PubMed: 27488683]
- (46). Paulus JK, Raman G, Rekkas A, Koethe B, Tanprasertsuk J, Lutz JS et al. White Paper, Appendix 1: Methods and Results of Evidence Review Committee Search The Predictive Approaches to Treatment Effect Heterogeneity (PATH) Statement. Washington, D.C.: Patient-Centered Outcomes Research Institute (PCORI); 2018.
- (47). Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; 2(3):i–88.

- (48). van Klaveren D, Vergouwe Y, Farooq V, Serruys PW, Steyerberg EW. Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. *J Clin Epidemiol* 2015; 68(11):1366–1374. [PubMed: 25814403]
- (49). Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ Cardiovasc Qual Outcomes* 2014; 7(1):163–169. [PubMed: 24425710]
- (50). Abadie A, Chingos MM, West MR. Endogenous stratification in randomized experiments (12 2013). Working Paper No.w19742. 2013. Available from: <http://ssrn.com/abstract=2370198>.
- (51). Lesko CR, Henderson NC, Varadhan R. Considerations when assessing heterogeneity of treatment effect in patient-centered outcomes research. *J Clin Epidemiol* 2018; 100:22–31. [PubMed: 29654822]
- (52). Instrument for assessing the credibility of effect modification analyses (ICEMAN) in a meta-analysis of randomized controlled trials [Under review at *Annals of Internal Medicine*].
- (53). Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312(7023):71–72. [PubMed: 8555924]
- (54). Tian L, Alizadeh AA, Gentles AJ, Tibshirani R. A Simple Method for Estimating Interactions between a Treatment and a Large Number of Covariates. *J Am Stat Assoc* 2014; 109(508):1517–1532. [PubMed: 25729117]
- (55). Foster JC, Taylor JM, Ruberg SJ. Subgroup identification from randomized clinical trial data. *Stat Med* 2011; 30(24):2867–2880. [PubMed: 21815180]
- (56). Chen G, Zhong H, Belousov A, Devanarayan V. A PRIM approach to predictive-signature development for patient stratification. *Stat Med* 2015; 34(2):317–342. [PubMed: 25345685]
- (57). Wang R, Schoenfeld DA, Hoepfner B, Evins AE. Detecting treatment-covariate interactions using permutation methods. *Stat Med* 2015; 34(12):2035–2047. [PubMed: 25736915]
- (58). Kovalchik SA, Varadhan R, Weiss CO. Assessing heterogeneity of treatment effect in a clinical trial with the proportional interactions model. *Stat Med* 2013; 32(28):4906–4923. [PubMed: 23788362]
- (59). Kunzel S, Sekkon J, Bickel P, Yun B. Meta-learners for estimating heterogeneous treatment effects using machine learning. *arXiv preprint arXiv:1706.03461* 2018.
- (60). Schuler A, Shah N, Baiocchi M. General-purpose model selection when estimating individual treatment effects. *ARXIV* 2018.
- (61). Califf RM, Woodlief LH, Harrell FE Jr., Lee KL, White HD, Guerci A et al. Selection of thrombolytic therapy for individual patients: development of a clinical model. GUSTO-I Investigators. *Am Heart J* 1997; 133(6):630–639. [PubMed: 9200390]
- (62). Fox KA, Poole-Wilson P, Clayton TC, Henderson RA, Shaw TR, Wheatley DJ et al. 5-year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the British Heart Foundation RITA 3 randomised trial. *Lancet* 2005; 366(9489):914–920. [PubMed: 16154018]
- (63). Yeh RW, Secemsky EA, Kereiakes DJ, Normand SL, Gershlick AH, Cohen DJ et al. Development and Validation of a Prediction Rule for Benefit and Harm of Dual Antiplatelet Therapy Beyond 1 Year After Percutaneous Coronary Intervention. *JAMA* 2016; 315(16):1735–1749. [PubMed: 27022822]
- (64). Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science* 2015; 349(6245):255–260. [PubMed: 26185243]
- (65). Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther* 2017; 102(6):924–933. [PubMed: 28836267]
- (66). Decker C, Garavalia L, Garavalia B, Gialde E, Yeh RW, Spertus J et al. Understanding physician-level barriers to the use of individualized risk estimates in percutaneous coronary intervention. *Am Heart J* 2016; 178:190–197. [PubMed: 27502869]
- (67). Selker HP. Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving hospital care. *Ann Intern Med* 1993; 118(10):820–822. [PubMed: 8470856]

- (68). Krumholz HM, Ross JS, Gross CP, Emanuel EJ, Hodshon B, Ritchie JD et al. A historic moment for open science: the Yale University Open Data Access project and medtronic. *Ann Intern Med* 2013; 158(12):910–911. [PubMed: 23778908]
- (69). Dahabreh IJ, Kent DM. Can the learning health care system be educated with observational data? *JAMA* 2014; 312(2):129–130. [PubMed: 25005647]
- (70). Vickers AJ, Scardino PT. The clinically-integrated randomized trial: proposed novel method for conducting large trials at low cost. *Trials* 2009; 10:14. [PubMed: 19265515]
- (71). van Staa TP, Klungel O, Smeeth L. Use of electronic healthcare records in large-scale simple randomized trials at the point of care for the documentation of value-based medicine. *J Intern Med* 2014; 275(6):562–569. [PubMed: 24635449]
- (72). Fiore LD, Lavori PW. Integrating Randomized Comparative Effectiveness Research with Patient Care. *N Engl J Med* 2016; 374(22):2152–2158. [PubMed: 27248620]
- (73). Institute of Medicine. Redesigning the Clinical Effectiveness Research Paradigm: Innovation and Practice-Based Approaches: Workshop Summary. 2010.
- (74). Ioannidis JP, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med* 2009; 1(7):7ps8.
- (75). Goodman SN. Probability at the bedside: the knowing of chances or the chances of knowing? *Ann Intern Med* 1999; 130(7):604–606. [PubMed: 10189332]
- (76). Kent DM, Shah ND. Risk models and patient-centered evidence: should physicians expect one right answer? *JAMA* 2012; 307(15):1585–1586. [PubMed: 22511683]
- (77). Stern RH. Individual risk. *J Clin Hypertens (Greenwich)* 2012; 14(4):261–264. [PubMed: 22458749]

Box A.**Consensus Criteria - When is a risk modeling approach to RCT analysis likely to be of most value?**

1. When there is a well-established overall treatment effect
 - A. Subgroup results (including risk-based subgroup results) from overall null trials should be interpreted cautiously.
2. When the benefits and harms/burdens of a given intervention are finely balanced (i.e. of similar magnitude on average), increasing the sensitivity of the treatment decision to risk prediction.
3. When treatments are associated with a non-trivial amount of serious harm or burden, increasing the importance of careful patient selection.
4. When several large, randomized, well-conducted clinical trials of contemporary interventions are available and appropriate for pooling in individual patient meta-analysis, to provide improved statistical power and broader variation in baseline outcome risk.
5. When substantial, identifiable heterogeneity of risk in the trial population is anticipated
 - A. When there are validated risk models and well established risk factors
 - B. When there is substantial case mix heterogeneity in the trial population
6. When there is strong preliminary evidence that a prediction model is clinically useful for treatment selection, or when there are models in current use for treatment selection (i.e., validation is a high priority)
7. When the clinical variables in the proposed models are routinely available in clinical care

Box B.**Consensus Guidance on Risk Modeling Approaches to Identify HTE****General**

1. Reporting RCT results stratified by a risk model is encouraged when overall trial results are positive to better understand the distribution of effects across the trial population.
2. Predictive approaches to HTE require close integration of clinical and statistical reasoning and expertise.

Identify or Develop a Model

3. When available, apply a high-quality, externally-developed, compatible risk model to stratify trial results.
4. When a high-quality, externally-developed model is unavailable, consider developing a model using the entire trial population to stratify trial results; avoid modeling on the control arm only.
5. When developing new risk models or updating externally-developed risk models, pre-specify the analytic plan prior to examination of trial data and follow guidance for best practice for prediction model development.

Apply the Model and Report Results

6. Report metrics for model performance for outcome prediction on the RCT, including measures of discrimination and calibration (when appropriate).
7. Report distribution of predicted risk (or the risk score) in each arm of the trial, and in the overall study population.
8. Report outcome rates and both relative and absolute risk reduction across risk strata.
9. When there are important treatment-related harms, these harms should be reported in each risk stratum to support strata-specific evaluation of benefit-harm trade-offs.
10. To test the consistency of the relative treatment effect across prognostic risk, a continuous measure of risk (e.g., the logit of risk) may be used in an interaction term with treatment group indicator.

Box C.**Consensus Statements on Caveats and Considerations Before Moving to Clinical Practice**

1. Clinical interpretation of HTE should stress differences in the absolute treatment effects across risk groups: the statistical significance of effect modification on the relative scale should not be conflated with the clinical significance of absolute treatment effect estimates.
2. External validation and calibration of risk prediction is important for translation of risk-specific treatment effects into clinical practice.
3. Clinical implementation may be supported by translating multivariable risk-based subgroup analysis into models yielding continuous treatment effect predictions to avoid artefactual discontinuities in estimation at the quantile boundary of an outcome risk group.

Box D.**Consensus Statements on Considerations and Caveats in Effect Modeling for HTE**

1. When *highly credible* relative effect modifiers have been identified, they should be incorporated into prediction models using multiplicative treatment-by-covariate interaction terms.
 - A. Credibility should be evaluated using rigorous multidimensional criteria (such as described in the ICEMAN tool) and should not rely solely on statistical criteria (such as p-value thresholds).
2. Avoid one-variable-at-a-time null hypothesis testing or stepwise selection (e.g., backward selection, forward selection) strategies to select single variable relative effect modifiers.
3. Avoid the use of regression methods that do not take into account model complexity when estimating coefficients (e.g., “conventional” unpenalized maximum-likelihood regression) when one or more treatment by covariate interaction terms are included in a treatment effect model.
4. Avoid evaluating models that predict treatment benefit using only conventional metrics for outcome risk prediction (e.g., based on discrimination and calibration of outcome risk prediction).

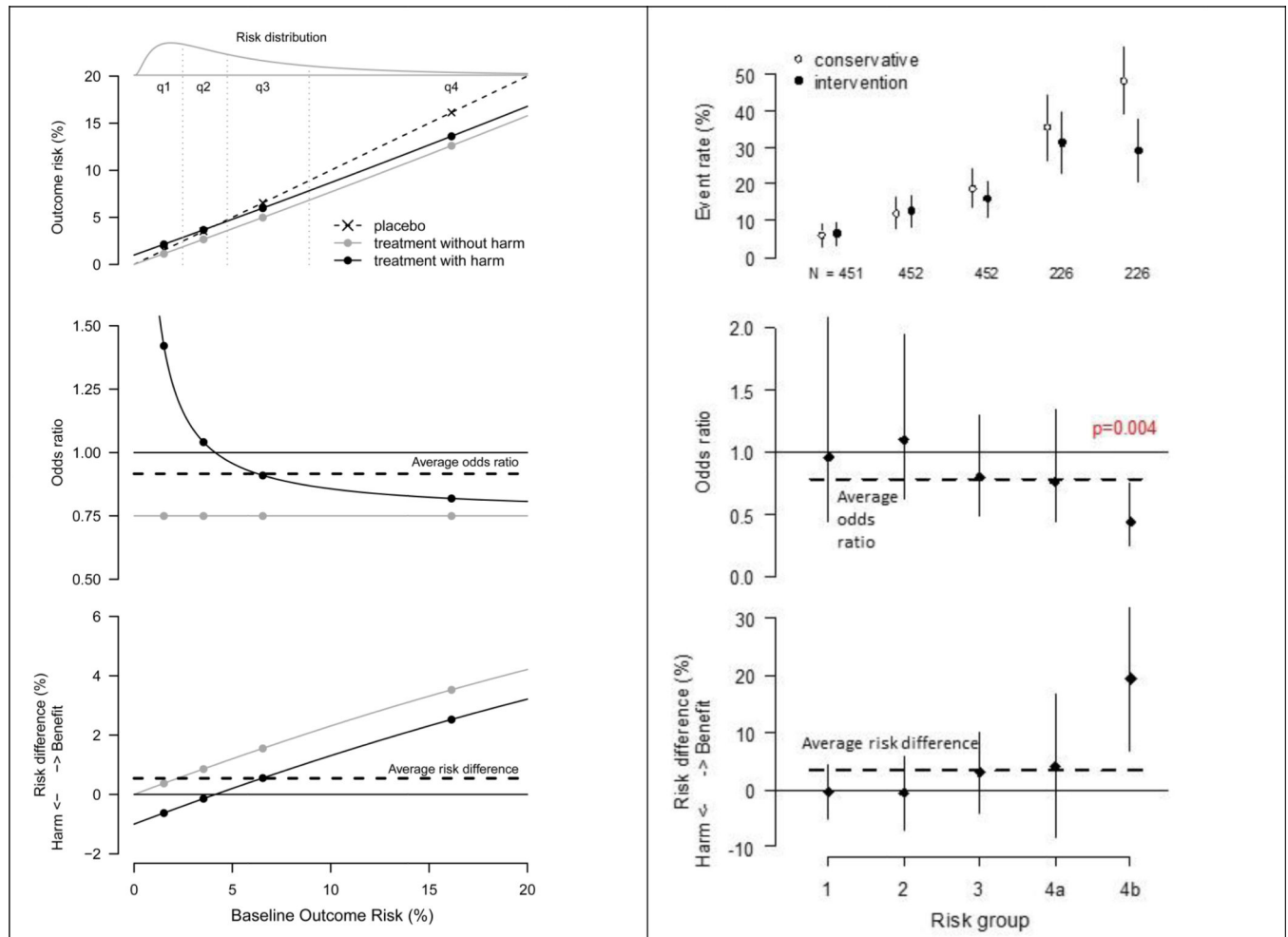


Figure 1. Schematized and Actual Risk-based Heterogeneous Treatment Effects

A. Schematic results in a trial for a hypothetical intervention which lowers the odds of an outcome by 25% but with an absolute treatment-related harm of 1%

This figure schematically depicts outcome risks for a trial testing a hypothetical intervention with an odds ratio of 0.75 but with an absolute treatment-related harm of 1% (shown in the top panel). Observed odds ratios (middle panel) and risk differences (bottom panel) are shown. Overall trial results are dependent on the average risk of the enrolled trial population. When the average risk is ~7% (as above), a well-powered study would detect a positive overall treatment benefit (shown by the horizontal dashed line in the middle and bottom panels). However, a prediction model with a C-statistic of 0.75, generates the risk distribution at the top of the figure. A treatment-by-risk interaction emerges (middle panel). Whether or not this interaction is statistically significant, examination of treatment effects on the absolute risk difference scale (bottom panel) reveals harm in the low risk group and very substantial benefit in the high risk group, both of which are obscured by the overall summary results. Conventional one-variable-at-a-time subgroup analyses are typically inadequate to disaggregate patients into groups that are sufficiently heterogeneous for risk such that benefit-harm trade-offs can misleadingly appear to be consistent across the trial population. Although the figure here shows idealized relations between risk and treatment

effects, these relations will be sensitive to how risk is described (i.e. what variables are in the risk model). Baseline risk is logit normal distributed with $\mu=-3$ and $\sigma=1$ (the log odds are normally distributed). Figure adapted from Kent DM et al. JAMA 2007.³

B. Stratified results of the Randomized Intervention Trial of unstable Angina (RITA)-3⁶²

The RITA-3 trial (N=1810) tested early intervention versus conservative management of non-ST-elevation acute coronary syndrome. Results for the outcome of death or non-fatal myocardial infarction at 5 years are shown above, stratified into equal-sized risk quarters using an internally-derived risk model; the highest risk quarter is sub-stratified in halves (groups 4a and 4b). Event rates with 95% confidence intervals (top panel), odds ratios (middle panel), and risk difference (bottom panel) are displayed. The risk model is comprised of the following easily obtainable clinical characteristics: age, sex, diabetes, prior MI, smoking status, heart rate, ST depression, angina severity, left bundle branch block, and treatment strategy. As in the schematic diagram to the left, the average treatment effect seen in the summary results (horizontal dashed line in middle and bottom panels) closely reflect the effect in patients in risk quarter 3, while fully half of patients (q1 and q2) receive no treatment benefit from early intervention. Absolute benefit (bottom panel) in the primary outcome was very pronounced in the eighth of patients at highest risk (4b). A statistically significant risk-by-treatment interaction* can be seen when results are expressed in the odds ratio scale (middle panel). Such a pattern can emerge if early intervention is associated with some procedure-related risks that are evenly distributed over all risk groups, eroding benefit in low risk but not high risk patients, as illustrated schematically in Figure 1A.

*The interaction p value is from a likelihood ratio test for adding an interaction between the linear predictor of risk and treatment assignment (one degree of freedom).

Table 1:

Treatment Effect is Mathematically Dependent on the Control Event Rate

Measure	Definition
Absolute Risk Difference	CER-TER
Relative Risk Reduction	$1 - \frac{TER}{CER}$
	CER
Odds Ratio	$\frac{TER/(1-TER)}{CER/(1-CER)}$
	CER/(1-CER)

CER=control event rate

TER=treatment event rate

TABLE 2.**Equations Corresponding to Risk Modeling and Effect modeling Approaches**

<u>Risk Modeling</u>	
A multivariable regression model f that predicts the risk of an outcome based on risk predictors x_i is identified or developed:	
Equation 1	$risk = f(\alpha + \beta_1 * x_1 + \dots + \beta_p * x_p)$
Variation in the treatment effect across risk can be tested statistically on the relative scale through the interaction between a linear predictor of risk ($lp = \beta_1 * x_1 + \dots + \beta_p * x_p$) and treatment:	
Equation 2	$risk = f(\alpha + \beta_{tx} * tx + \beta_{lp} * lp + \delta_{lp} * lp * tx)$
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 the statistical significance of this interaction effect, δ).	
When relative effects across risk strata appear constant, a model with a constant treatment effect may suffice:	
Equation 3	$risk = f(\alpha + \beta_{tx} * tx + \beta_1 * x_1 + \dots + \beta_p * x_p)$, 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.
<u>Effect Modeling</u>	
A regression model f is developed on RCT data with inclusion of risk predictors x_i , a treatment assignment variable tx , and potential inclusion of treatment interaction terms ($x_i * tx$):	
Equation 4	$risk = f(\alpha + \beta_{tx} * tx + \beta_1 * x_1 + \dots + \beta_p * x_p + \delta_1 * x_1 * tx + \dots + \delta_p * x_p * tx)$