



BRIGHAM AND
WOMEN'S HOSPITAL

Estimating and testing interactions in time-to-event analysis

Andrea Bellavia

TIMI study group

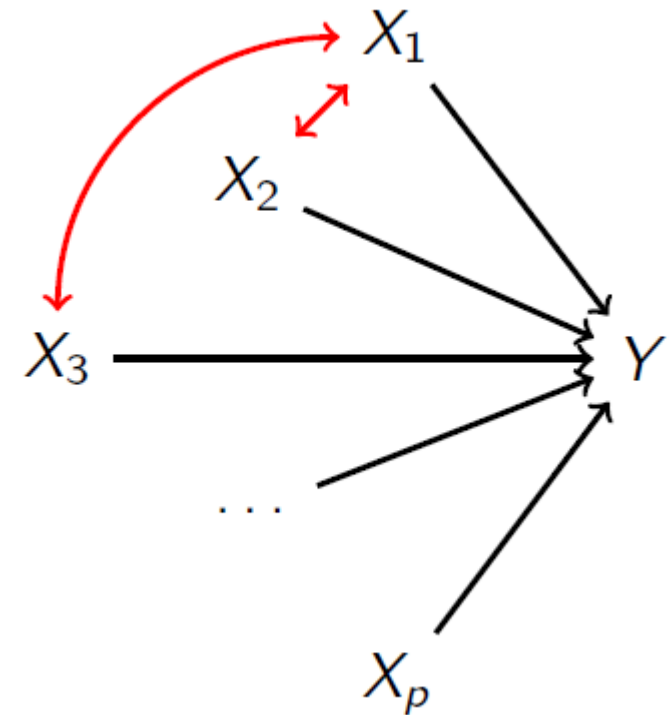
January 19, 2023

Outline

1. The different meanings of the word “interaction”
 - Effect modification vs clinical/biological interaction
 - Multiplicative vs additive interaction
2. How do we assess interactions from data?
3. Estimating/testing/presenting additive and multiplicative interactions in survival analysis:
 - 2 Binary and/or categorical variables
 - 1 Binary and 1 continuous variables
 - 2 Continuous variables
 - More than 2 variables
4. Summary and discussion

1. The different meanings of the word interaction

In the broadest sense, we talk about interaction analysis when we want to evaluate the **joint contribution of 2 or more factors** as they relate to an outcome of interest.



How do factors interact?

Some examples of joint contributions:

- The effect of a given treatment is heterogeneous over patients' characteristics [**effect modification**]
 - e.g. TRT effect stratified by sex
- The simultaneous presence of both factors further enhances the outcome risk/activation [**biological or clinical interaction**]
 - e.g. unhealthy diet + smoking and the risk of stroke
- Two factors operate in a sequential way [**mediation + interaction**] / not discussed today
 - e.g. unhealthy diet, BMI, and the risk of stroke

Effect modification vs interaction

- This distinction is not trivial and has been the subject of several papers, especially in epi literature

> [Epidemiology](#). 2009 Nov;20(6):863-71. doi: 10.1097/EDE.0b013e3181ba333c.

On the distinction between interaction and effect modification

[Tyler J VanderWeele](#) ¹

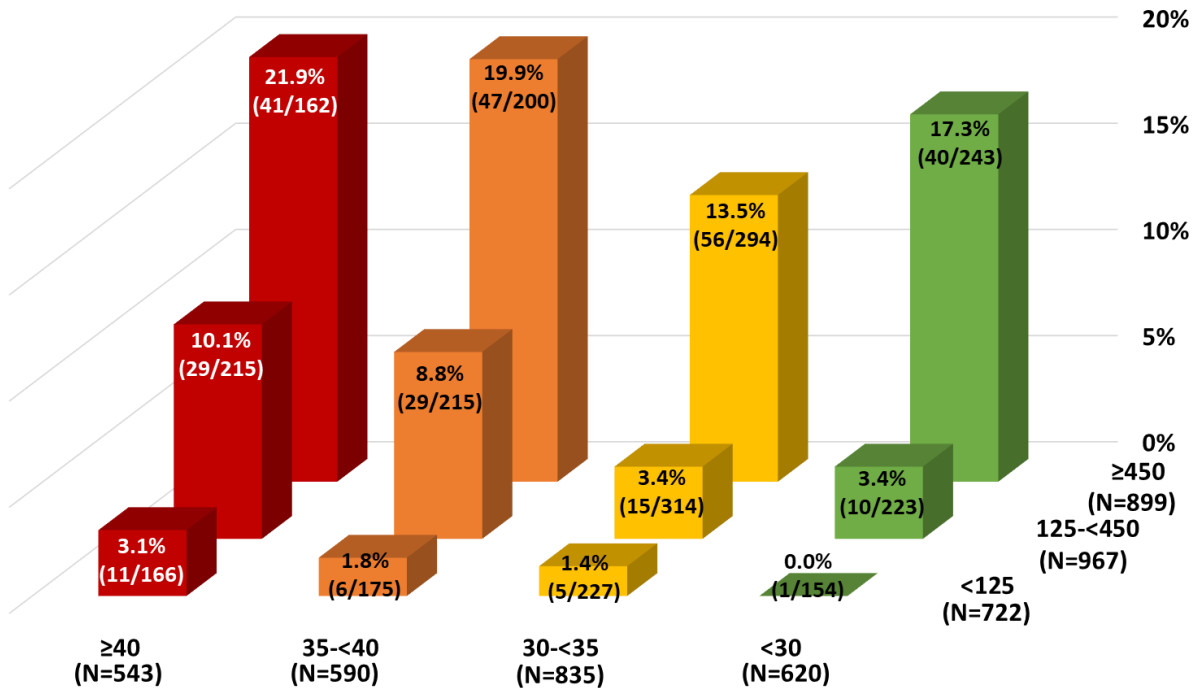
In general:

- **Effect modification**: one primary variable of interest and one (or more) strata variable -> present effect of interest over different strata of the other variables
- **Interaction**: equal interest in all factors at play -> present joint effects

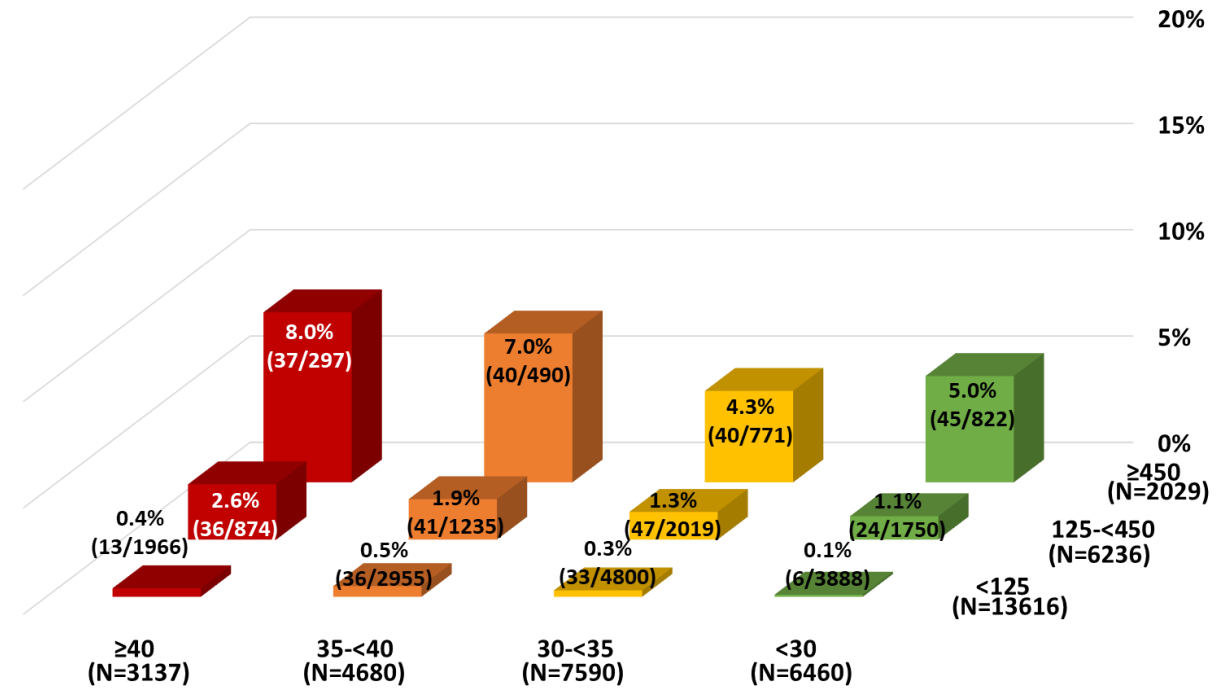
- **Effect modification** is addressed with **stratified analysis**
- We are not interested in the effect of the strata itself. This is often established and/or non-modifiable (e.g. disease history)
- The tests of interest are on each within-strata effects, and on the between-strata difference
- On the other hand, for an **interaction** analysis we are interested in the **joint effect** as it decomposes into main+main+interaction

Example from last Thursday's presentation

Prior HF



No Prior HF

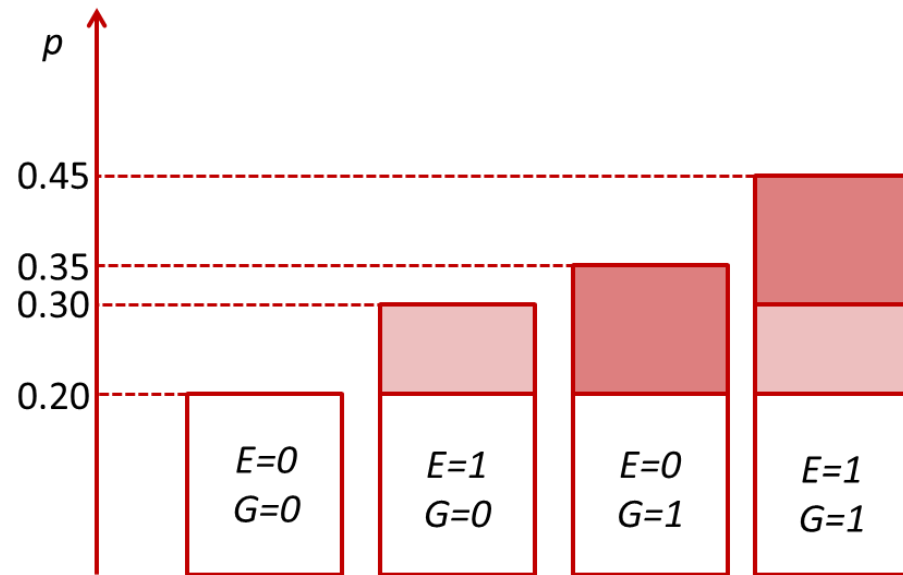


[Slide from S. Patel's presentation on 1/12. HHF rates over combined categories of NT-proBNP and BMI, stratified by prior HF]

Additive vs multiplicative interaction

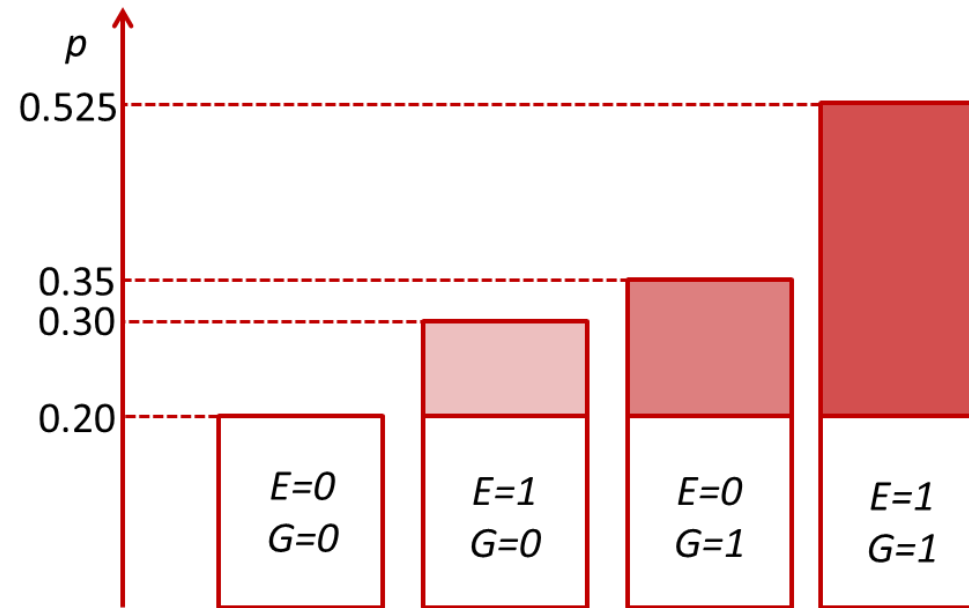
Another important distinction is the scale on which interaction is measured:

Is the combined effect larger than the **sum** of the 2 main effects?



Perfect additivity / no interaction on the additive scale

Is the combined effect larger than the **product** of the 2 main effects?



Perfect multiplicativity / no interaction on the multiplicative scale

e.g. G =smoking, E =obesity

- Additive and multiplicative interactions, however, can underline different conceptual models, and, potentially, different interventions
- Whether one is interested on one or the other scale will depend on **biological / clinical assessment**

Very interesting reading on the underlying biology of additive and mult. interactions: Diaz-Gallo et al. PLOS ONE, 2021

And on the relevance of additive interaction for assessing intervention's effects in public health: VanderWeele & Knol, A Tutorial on Interaction

- Especially when dealing with time-to-event endpoints, we are often interested in either hazard (multiplicative scale), absolute risks (additive scale), or both
- Ideally, we would always assess both scales and both interactions. Most clinical trials, however, only report measures and tests for multiplicative interaction
- The goal of today's presentation is to detail how (if possible) to estimate, present, and test for interaction or effect modification on either scales

2. Statistical interaction

In general terms, we assess whether there is an interaction between E and G in predicting Y by including a product term $E*G$ in the statistical model

$$f(Y) = \beta_0 + \beta_1 E + \beta_2 G + \beta_3 E*G$$

- β_1 and β_2 are the main effects of E and G
- β_3 describes the additional change in $f(Y)$ when both E and G are present
- $\beta_3 = 0$ implies absence of interaction; $\beta_3 > 0$ positive interaction; $\beta_3 < 0$ negative interaction
- The p-value associated with $\beta_3 = 0$ can be used as p-value for interaction

Test for effect modification

When interested in EM, the test for between-strata difference correspond to the test for interaction (product term)

The difference is in how we present the results:

- Interaction: use the estimates from the model coefficients to **predict $f(Y)$ at all levels of E and G**
- Effect modification: Fit **additional regression models over strata of G**, and present the main effect of E over strata, with tests for within-strata trends

Key concept in statistical interaction

Whether we are testing additive or multiplicative interaction depends on the scale of the underlying model

- The product term (and associated test) from a **linear model** (e.g. linear regression) estimates (and test) **additive interaction**
- The product term (and associated test) from a **log-linear model** (e.g. logistic/cox) estimates (and test) **multiplicative interaction**

3. Time-to-event endpoints

With survival data we are often assessing the effects of a given predictors in terms of **both**:

- **hazard** (hazard ratios -> multiplicative scale)
- **risk** (CIF/absolute risks -> additive scale)

The common modeling choice is the Cox model, hence the predominance of interaction tests only on the hazard (multiplicative) scale

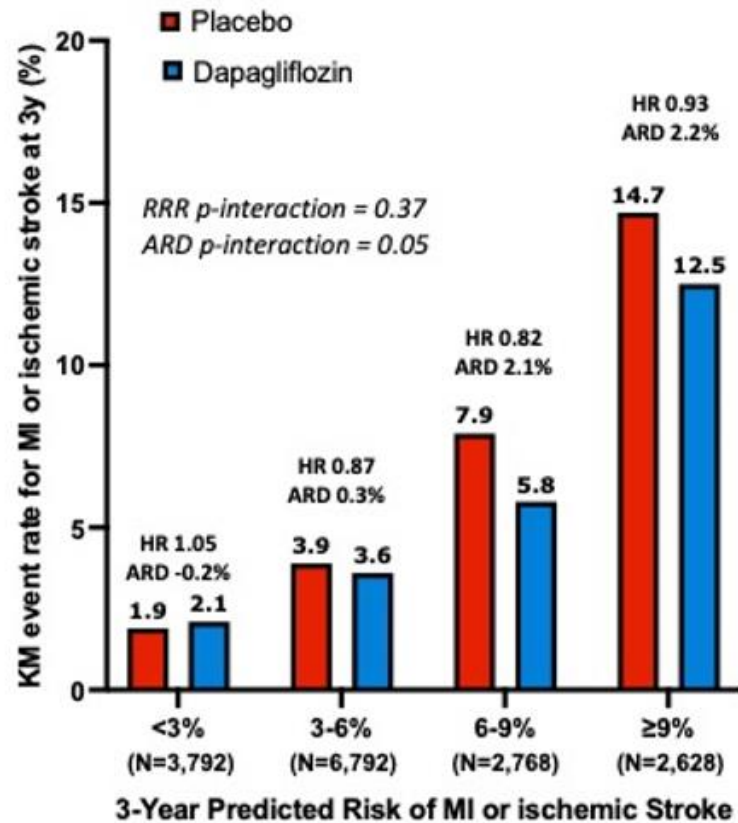


Figure from Berg et al. under review.

When designing a study, it is important to clearly identify the type of interaction/EM analyses of interest

There could be situations where an interaction is only observed on a given scale. In these cases, the clinical interpretation is critical.

Based on the figure on the left, would you say that there is an interaction?

- We focus here on interaction assessment from the Cox model
- Current literature comprehensively covers the setting of 2 binary covariates. We are working on extending methodological framework and software tools to more complex scenarios

	Multiplicative		Additive	
	Estimate and display	CI / test	Estimate and display	CI / test
2 binary and/or categorical covariates	Established	Established	Established	Options available - framework not established
1 binary and 1 continuous	Established	Established	TIMIstats WiP	TIMIstats WiP
1 binary and 1 continuous, with splines	TIMIstats WiP	TIMIstats WiP	TIMIstats WiP	TIMIstats WiP
2 continuous, possibly with splines	Established	\	\	\
More than 2 covariates	Options available - framework not established	Options available - framework not established	\	\

Data for illustration

- Risk score for atherothrombotic events in patients with t2 diabetes (Berg et al under review)
- Secondary analysis: assessment of the interaction between the developed risk score and treatment (dapa/placebo) in DECLARE participants
- Interest in both hazard rates and absolute risks of MI/IS at 3 years
- Manuscript includes results with categorical version of predicted risk (4 groups). Other settings are presented here only for illustrative purpose

Binary and categorical covariates

- Multiplicative interaction between 2 binary covariates is estimated and tested using the product term

```
coxph(formula = surv(days2miistr, miistrfu) ~ trt * pred2a,  
      ties = "breslow", x = TRUE, y = TRUE)
```

```
n= 8895, number of events= 610
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
trt	-0.23460	0.79089	0.13260	-1.769	0.0769
pred2a>=12%	0.78637	2.19540	0.11192	7.026	2.12e-12
trt:pred2a>=12%	-0.02275	0.97750	0.16821	-0.135	0.8924

- Coefficients can be used to predict absolute risks at a given time point (here 3 years)

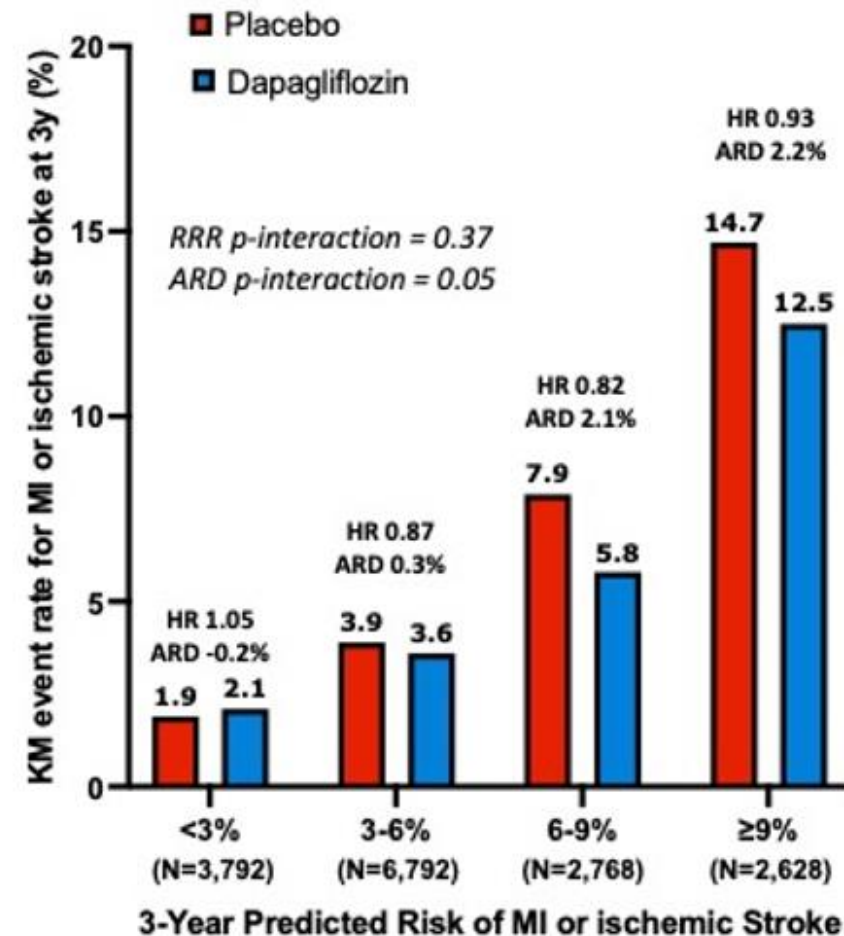
	DAPA	Placebo
Low risk	5%	11%
High risk	7%	14%

- The additive interaction is

$$AR_{11}-AR_{10}-AR_{01}-AR_{00} = \\ 0.14-0.11-0.07+0.05 = 0.01$$

- The case with 2 binary covariates is the only one where formal model-based tests exist (e.g. RERI). Details, if interested, in Li & Chambless 2007. In this example, $p=0.099$ (note that p on the hazard scale=0.89)

- The same reasoning can be extended to categorical variables with more than 2 groups.
- In this case, the actual product terms do not have a clear interpretation -> display absolute risks and HRs over combined groups



Statistical testing becomes more burdensome

- The p for interaction on the **hazard scale** can be calculated with the **joint test** of the different combinations of product term (similar to the joint test of all dummies for a categorical covariate)
- A p for interaction on the **additive risk scale can not be calculated from the model***
- Options we commonly use include tests for homogeneity trend in the predicted ARs across the strata variable, and inverse-variance weighted least-square model to account for strata imbalance
- P-values from these tests are not coming from the original Cox model. A direct comparison with the (model-based) p on the hazard scale should be avoided

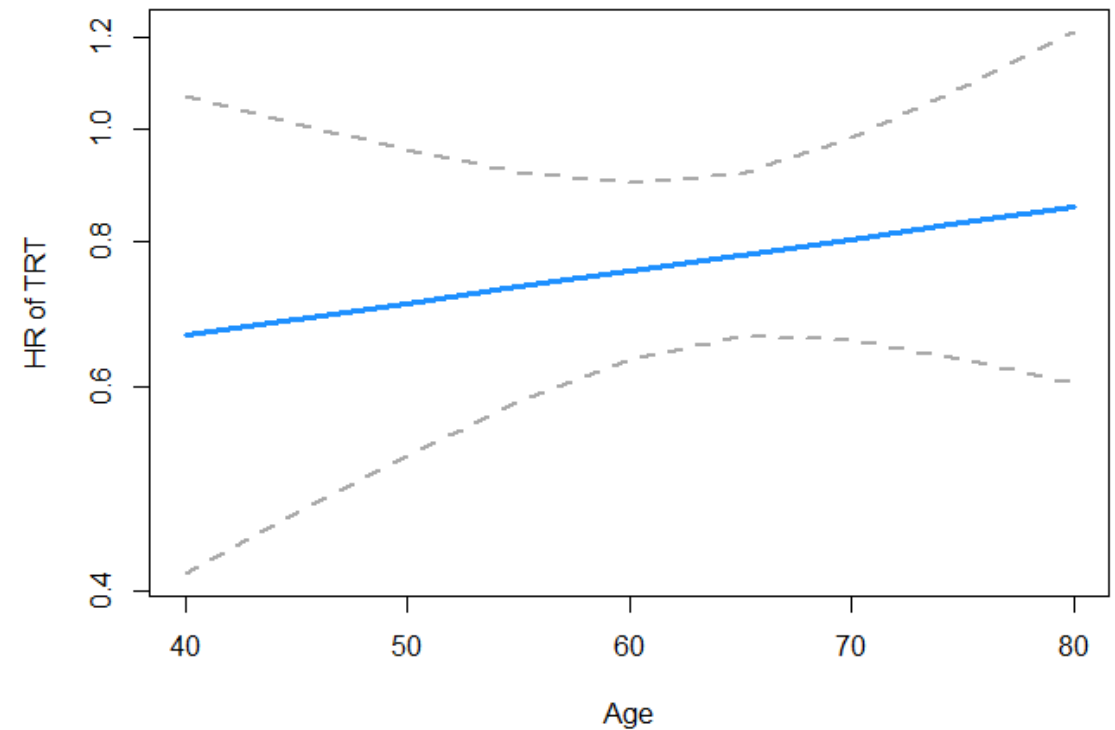
* tests like RERI have been extended to categorical and one continuous covariate (no splines) only for logistic regression

Binary and continuous covariates

- From a modeling perspective, integrating a continuous covariate is straightforward
- The product term itself, however, is not very intuitive anymore

	coef	exp(coef)	se(coef)	z	Pr(> z)
trt	-0.660240	0.516727	0.617506	-1.069	0.285
AGE	0.009778	1.009826	0.006365	1.536	0.124
trt:AGE	0.006316	1.006336	0.009620	0.657	0.511

- Better to show the main effect of the binary covariate over levels of the continuous one
- Power for interaction is lower than power for main effects. Even in RCTs, be careful in using the 0.05 threshold. The figure tells a different story (there are specific ages where the TRT effect is significant)



- To better capture the actual TRT effect change, we can relax the linear interaction assumption using splines
- The test for interaction is the joint test of all product term coefficients (even lower power)

	coef	exp(coef)	se(coef)	z	Pr(> z)
trt	-1.00629	0.36557	1.14889	-0.876	0.38109
rct:AGE, 3)AGE	-0.02733	0.97304	0.01291	-2.118	0.03419
rct:AGE, 3)AGE'	0.05022	1.05150	0.01568	3.203	0.00136
trt:rct:AGE, 3)AGE	0.01296	1.01304	0.02027	0.639	0.52261
trt:rct:AGE, 3)AGE'	-0.01136	0.98870	0.02397	-0.474	0.63546

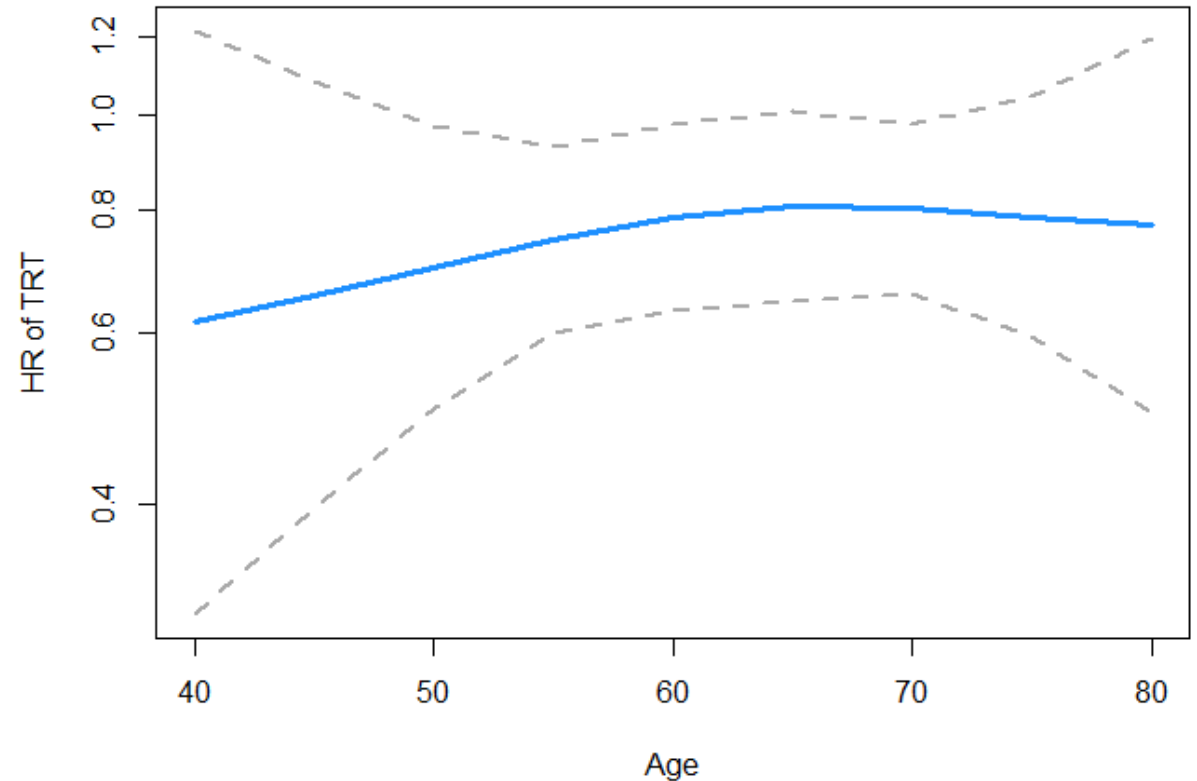
Hypothesis:

trt:rct:AGE, 3)AGE = 0
trt:rct:AGE, 3)AGE' = 0

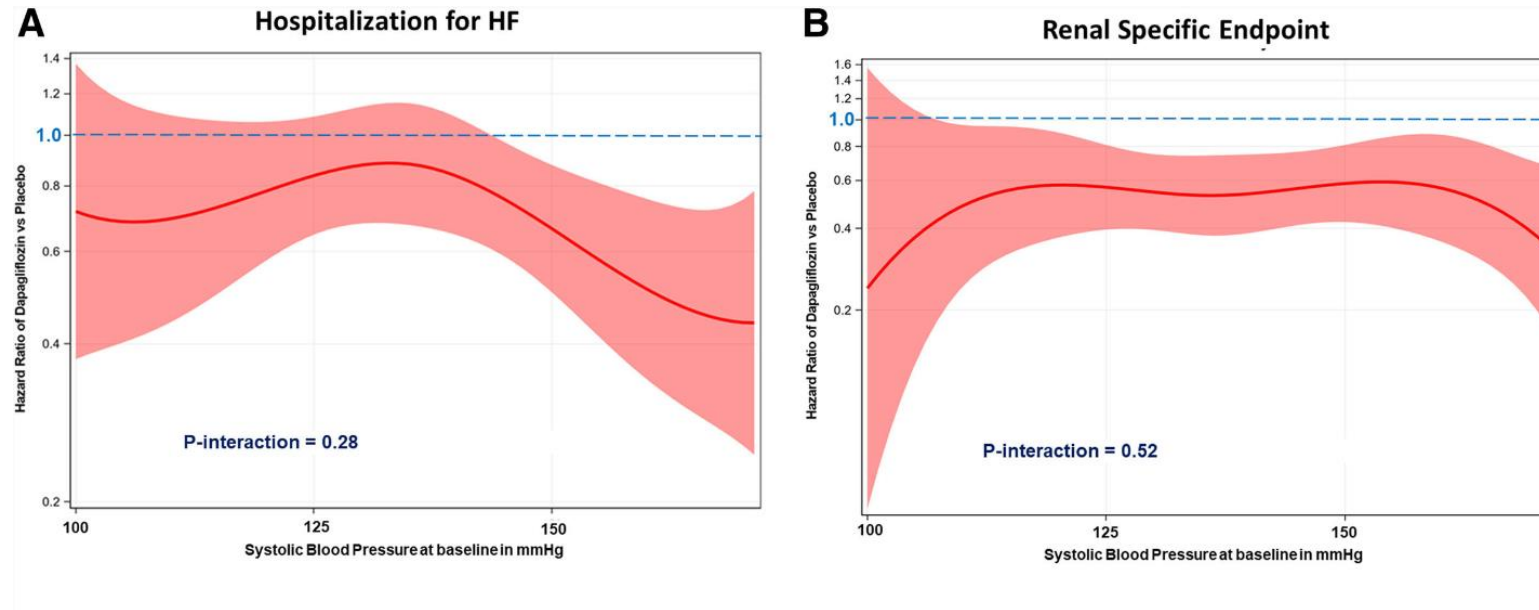
Model 1: restricted model

Model 2: Surv(days2miistr, miistrfu) ~ trt * rct:AGE, 3)

	Res.Df	Df	Chisq	Pr(>Chisq)
1	8892			
2	8890	2	0.4628	0.7934



Generating those plots is a very complicated coding task, and literature/software documentation is scarce. Kudos to TIMI stats colleagues who have developed SAS macros over the years



e.g. use of TIMI SAS macro in Furtado et al. 2022, Circulation / dapa*SBP interaction

And, more recently, extensions for R

Package ‘interactionRCS’

October 13, 2022

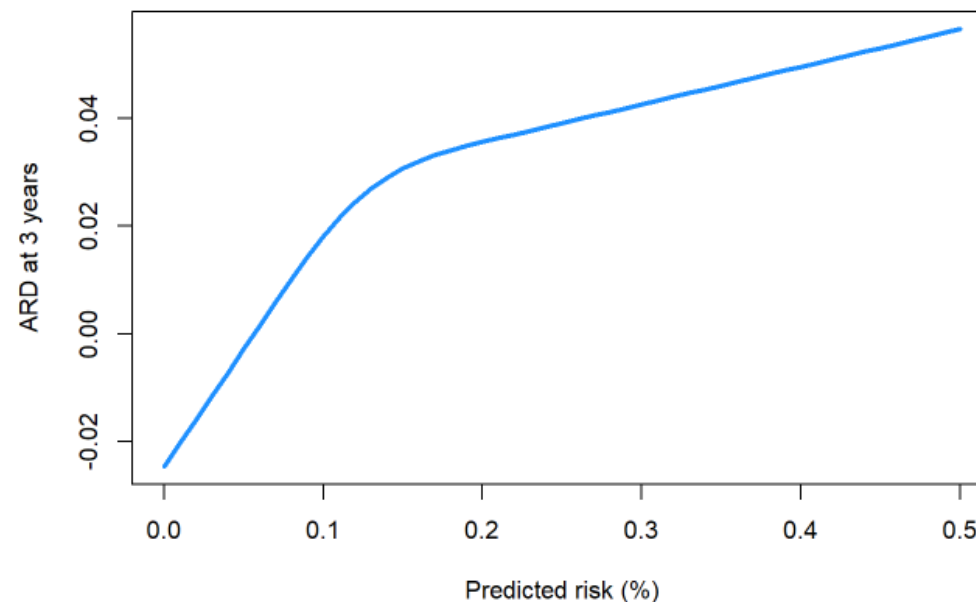
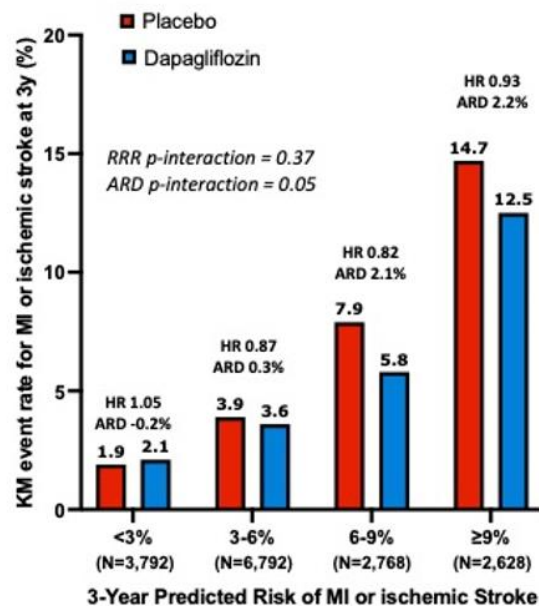
Type Package

Title Calculate Estimates in Models with Interaction

Version 0.1.0

Binary and continuous / additive scale

- How to translate binary/continuous interactions on the additive absolute risk scale has not been addressed, and software tools are not available
- Our ongoing research is addressing this topic, also integrating flexible modeling with splines



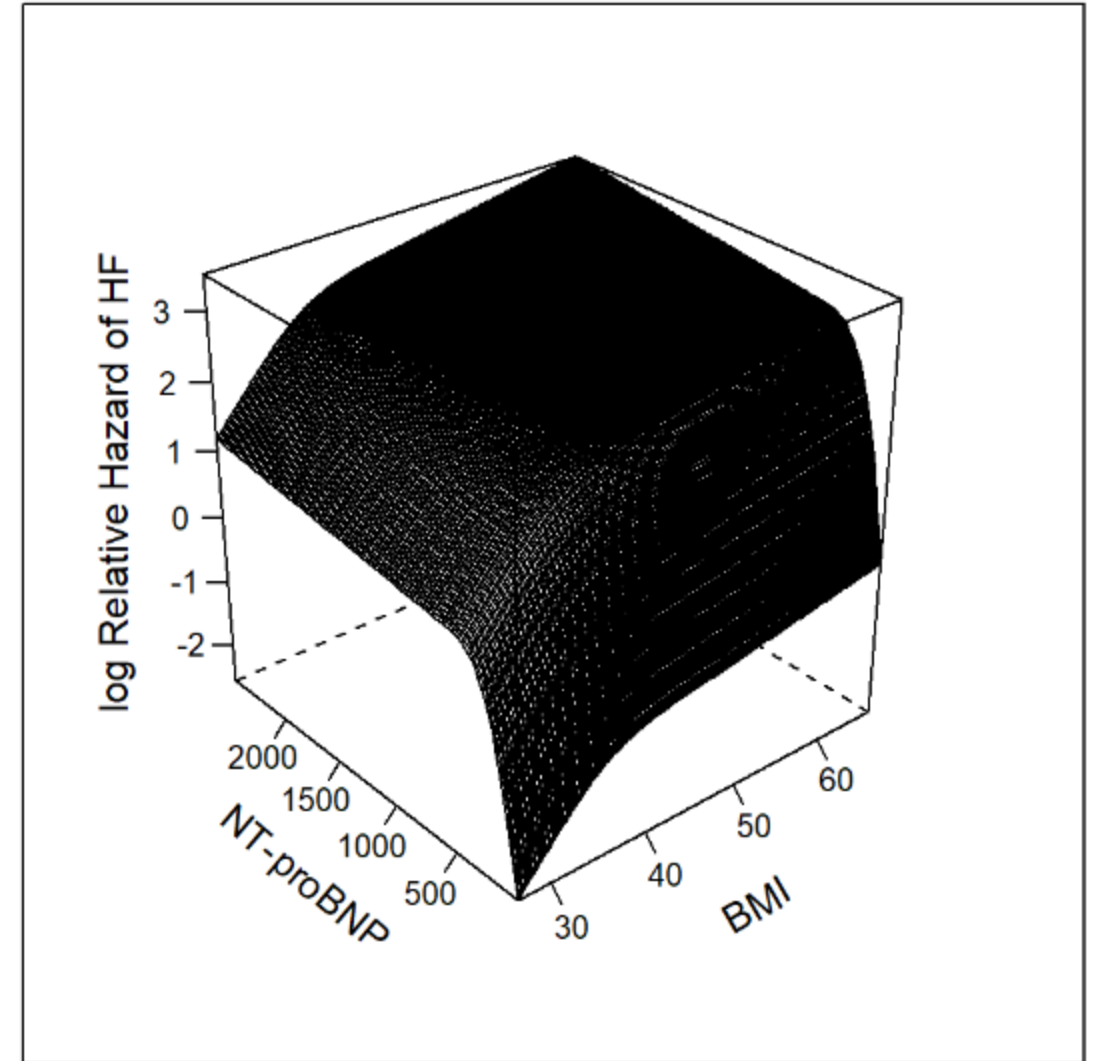
TIMI stats WiP

- We extended the theoretical framework and confirmed its validity through illustrative examples and simulation studies
- Work in progress to identify the best approach for confidence intervals (extension of previous approach / bootstrap / delta method)
- Currently integrating this approach to the existing macros and R package. This will provide a comprehensive SAS/R tool for flexible interactions fully developed within TIMI stats
- Upcoming conference presentation and manuscript drafting in progress

2 continuous covariates

- In this particular setting it is critical to distinguish whether we are interested in EM (effect of one covariate over levels of the second one), or actual interaction (joint effect)
- In the first case, we can use the previous approach (plotting the HR/ARD of 1-SD unit increase in the primary covariate, over levels of the second one)
- The joint effect is more complicated. A 3-dimensional plot is probably the best approach. When using splines for both covariates, this is also known as **surface plot**

- Very complicated to include confidence intervals
- The p for interaction is the joint test of all splines combinations (very low power -> even more caution)



Example of surface plot from the analyses presented last Thursday

- Approaches to translate this setting to the absolute risk scale are not available
- Creating categories of both covariates is a valid alternative

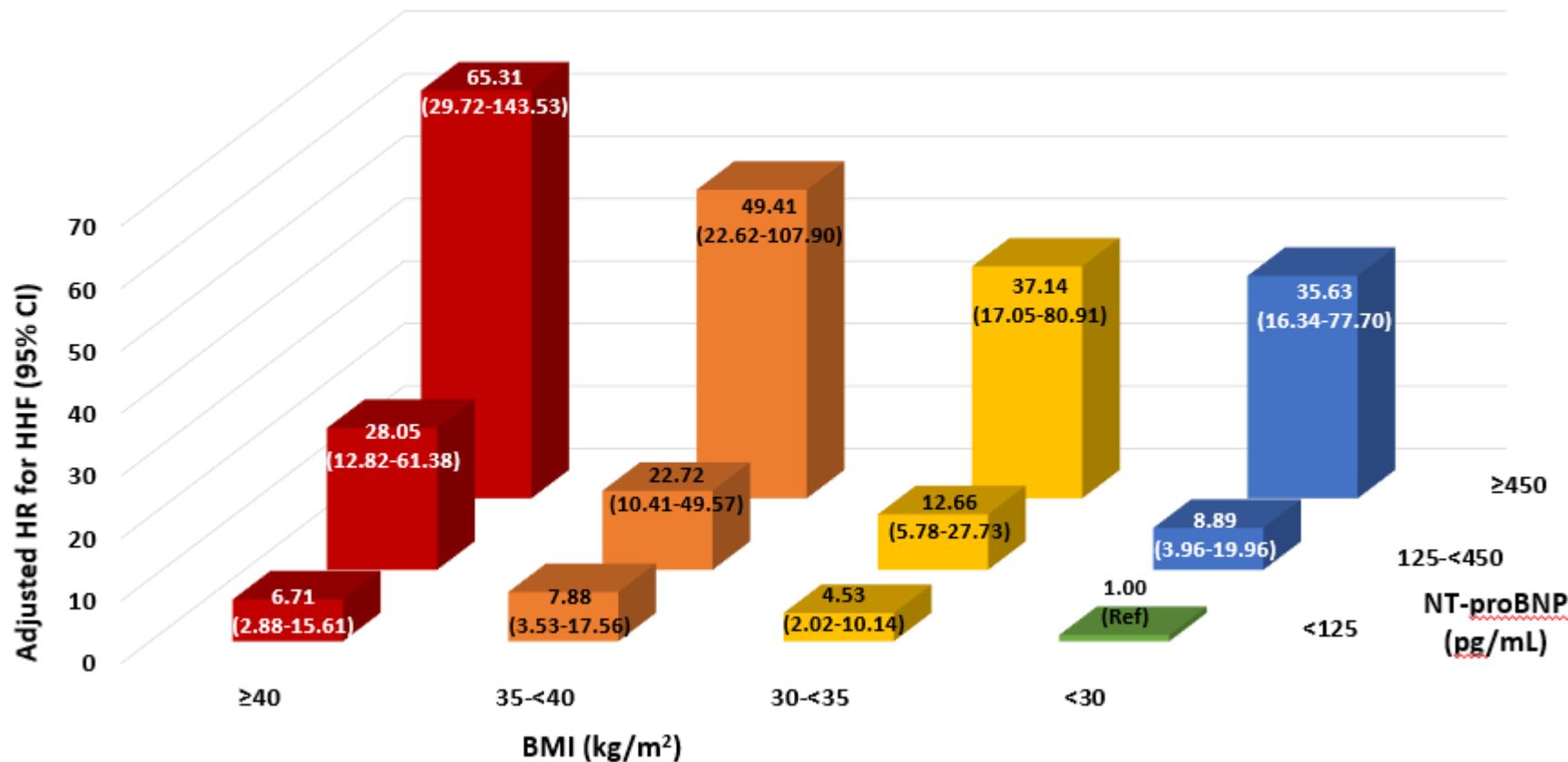


Figure from 1/12 Thursday meeting presentation

3 or more variables

- 2-way interaction only touches the surface of the real-world complexity
- Example: joint effect of BMI and NT-proBNP over levels of previous HF (slide 6). How are the 3 factors jointly contributing to the event of interest?
- First task is to clarify what factors are potential EMs (stratification)
- Obtaining overall p-values is a very low-power not-recommended procedure

- Assessing high-dimensional interactions is a key topic with big data (e.g. proteomics). Current recommendation is for interaction screening in a 2-stage procedure
- Extensions on interpretable additive interactions in this setting has not been addressed

4. Summary and discussion

- Assessing interactions is not a simple task
- Difference between Effect Modification and biological/clinical interaction
- Difference between additive and multiplicative interaction
 - Is there a specific biological/clinical model to guide the choice?
 - In some settings (e.g. TRT interactions in clinical trials) the relationship can be straightforward, but in most setting it is not (e.g. BMI*NT-proBNP example)
- Especially with continuous plots, favor CIs over p-values. Tests for interaction are (increasingly) low-power procedures and difficult to compare across settings
- We are on the forefront of the research on this topic. Still, there are several unaddressed topics and little applications in real data. Looking forward to potential applications for improvements and new developments