Subgroup analysis: a look at the SEAMOS approach

SEAMOS: Standardised Effects Adjusted for Multiple Overlapping Subgroups

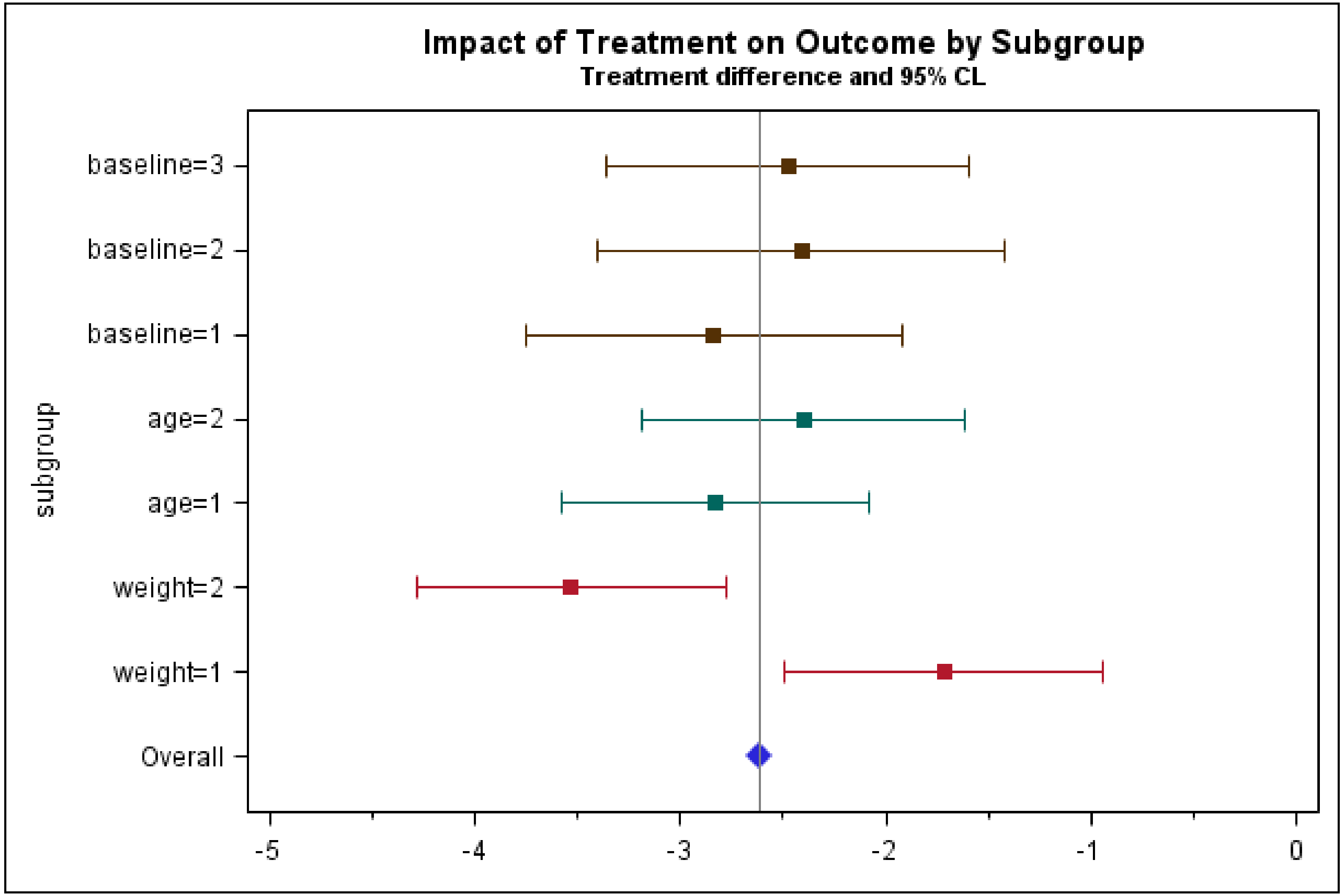
Michael O'Kelly, Centre for Statistics in Drug Development, IQVIA

Acknowledgements

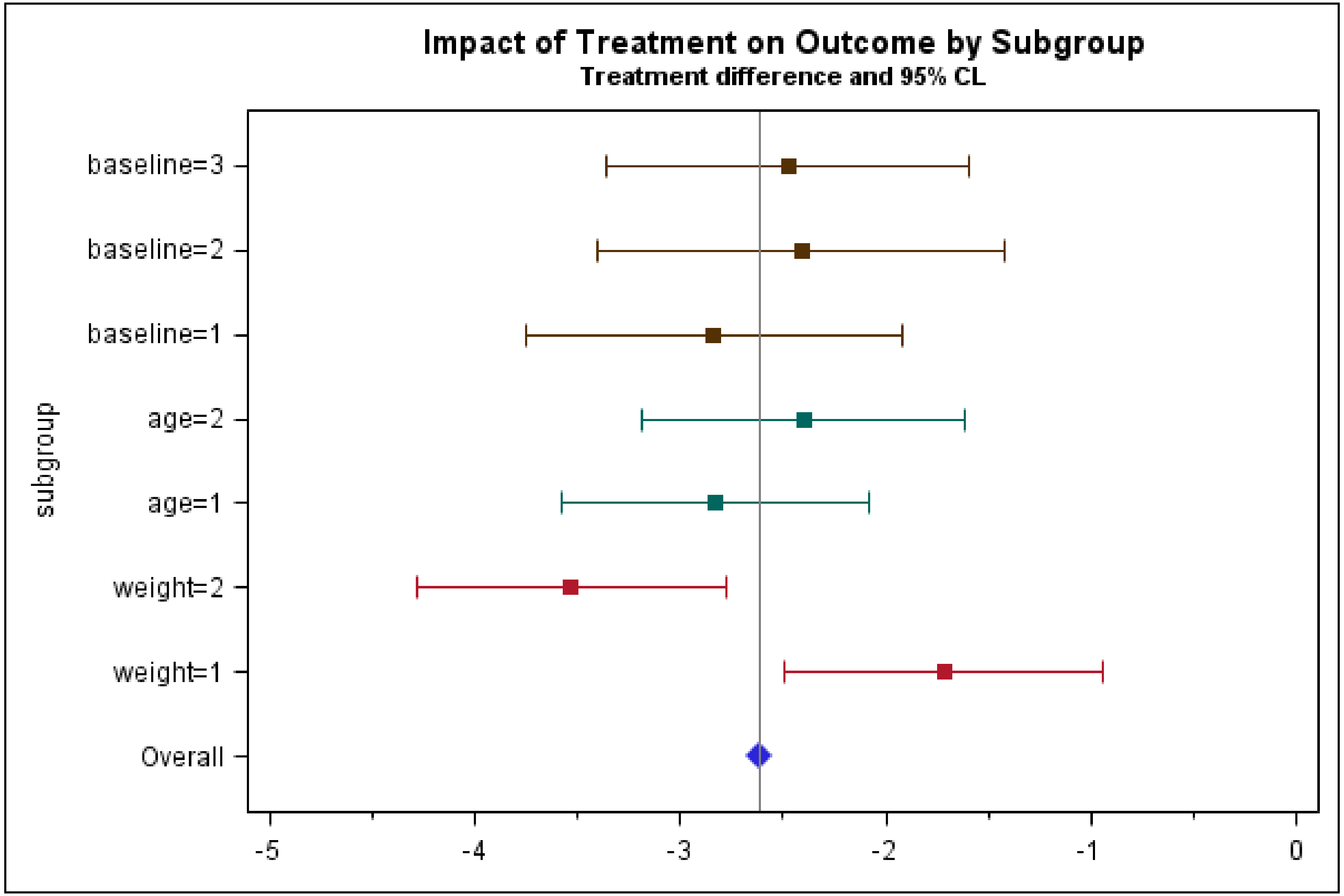
- For their ideas, advice, insights, comments and discussion:
  - Tom Parke, Berry Consultants
  - Ilya Lipkovich, Eli Lilly
  - David Svensson, AstraZeneca
  - Jonathan Bartlett, University of Bath

The problem: a standard clinical-trial output

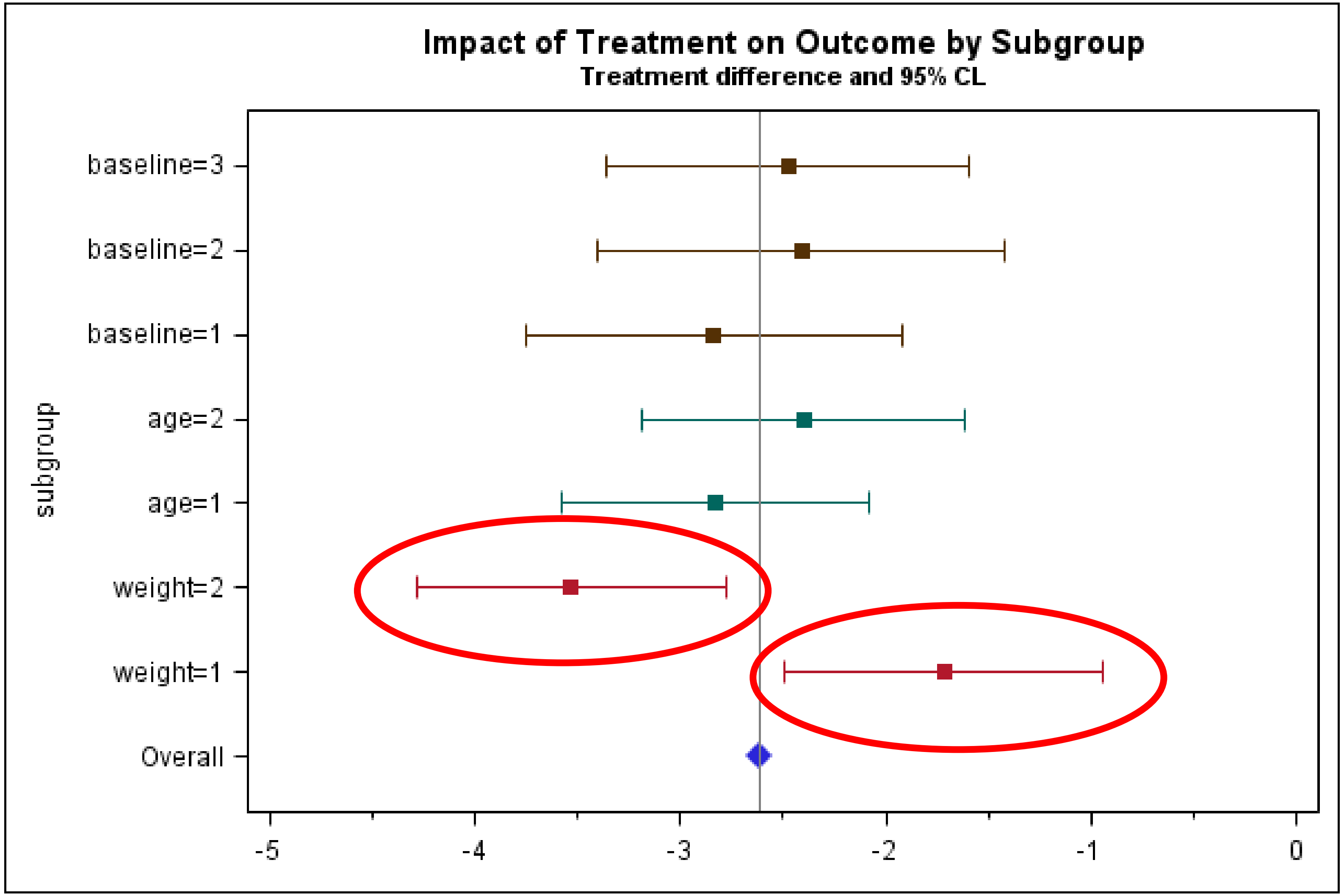
The problem


Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
subgroup
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
-5
-4
-3
-2
-1
0

“Move along, nothing to see here”?



Or is there a true subgroup effect hidden in the forest?



Objective, paradoxical for data visualisation session

- Can we find a statistic or statistics to help us assess this visual representation of subgroups?

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - Search out promising subgroups?
  - Understand patterns linking treatment effect to population attributes?

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - Search out promising subgroups?
  - Understand patterns linking treatment effect to population attributes?
- Complications
  - Correlated subgroups
  - Numbers of categories in a subgroup
  - Unequal N per subgroup

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - Search out promising subgroups?
  - Understand patterns linking treatment effect to population attributes?

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - Search out promising subgroups? [machine learning, e.g. SIDES, etc.]
  - Understand patterns linking treatment effect to population attributes? [multivariate modelling]

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - Search out promising subgroups?
  - Understand patterns linking treatment effect to population attributes?

Massive statistical literature on subgroup identification

- Objective
  - Check for homogeneity?
  - SEAMOS: Standardised Effects Adjusted for Multiple Overlapping Subgroups
  - Idea, focussed specifically on assessment of forest plots:
    - › Identify an overall measure of existence of subgroup in the plot
    - › Permute the subjects along with their subgroup attributes, preserving the treatment effect (i.e. treatment group and outcome are not permuted)
    - › Reject null hypothesis quantified by overall measure of existence of subgroup
    - › SEAMOS: overall measure is largest standardised subgroup treatment effect
  - Dane et al., (2018) Subgroup analysis: White paper of the EFSPI/PSI working group on subgroup analysis, *Pharmaceutical Statistics.*

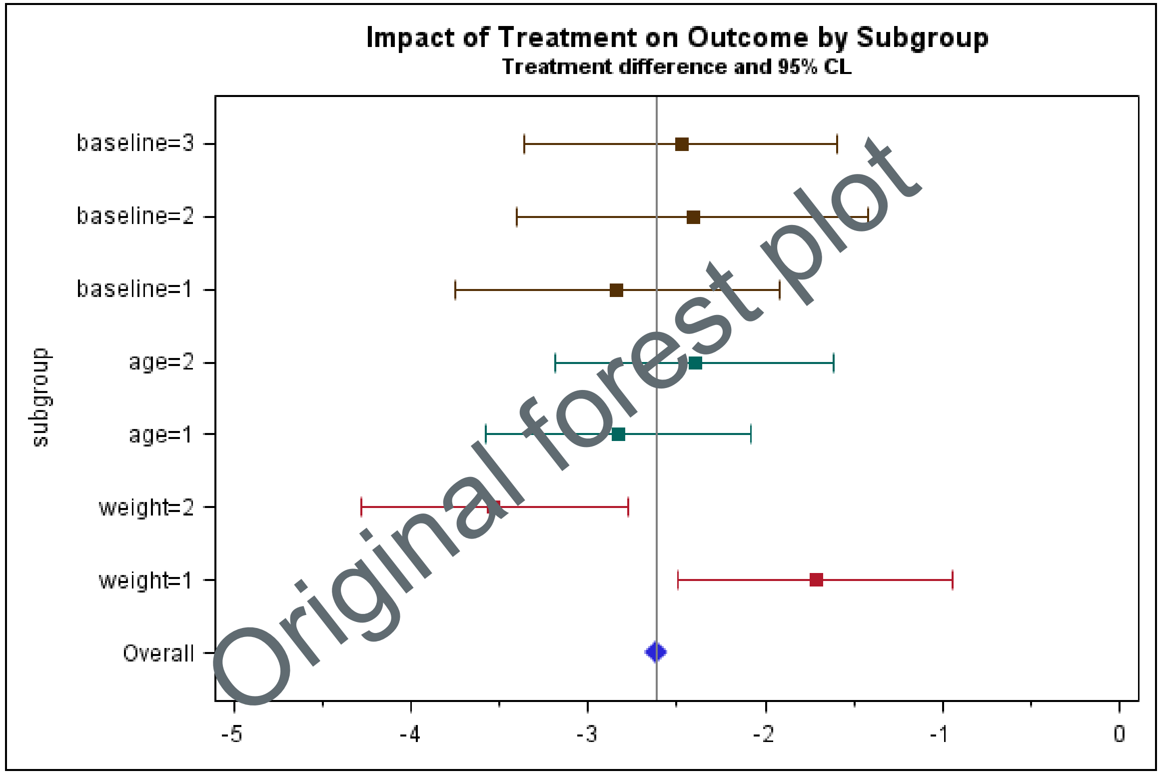
SEAMOS

- SEAMOS: overall measure is largest standardised subgroup treatment effect, $\max(\Delta_{ij})$, where

$$\Delta_{ij} = \frac{\bar{\delta} - \delta_{ij}}{SE_{\delta_{ij}}}$$

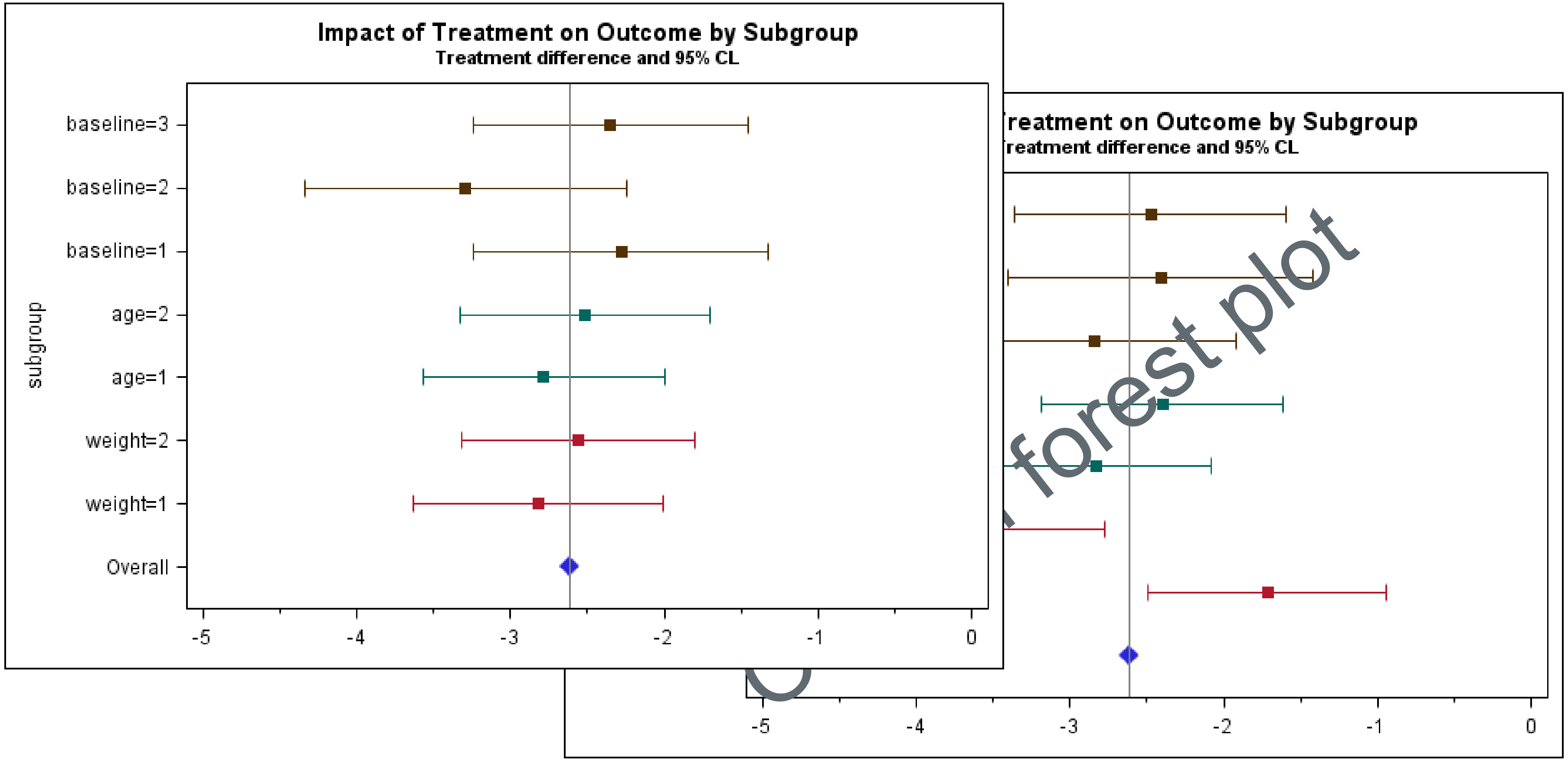
- …and SE is the standard error, δ_{ij} is the estimate of treatment difference for category *j* of subgroup *i*, and $\bar{\delta}$ is the overall estimate of treatment difference ignoring subgroup.
- The permuting of the subgroup information in q=1, 2…Q permuted datasets allows access to a "null-hypothesis" scenario, percentiles of whose $\max(\Delta_{ijq})$ can be used to reject the null hypothesis of no subgroup as quantified by $\max(\Delta_{ij})$

Like this….

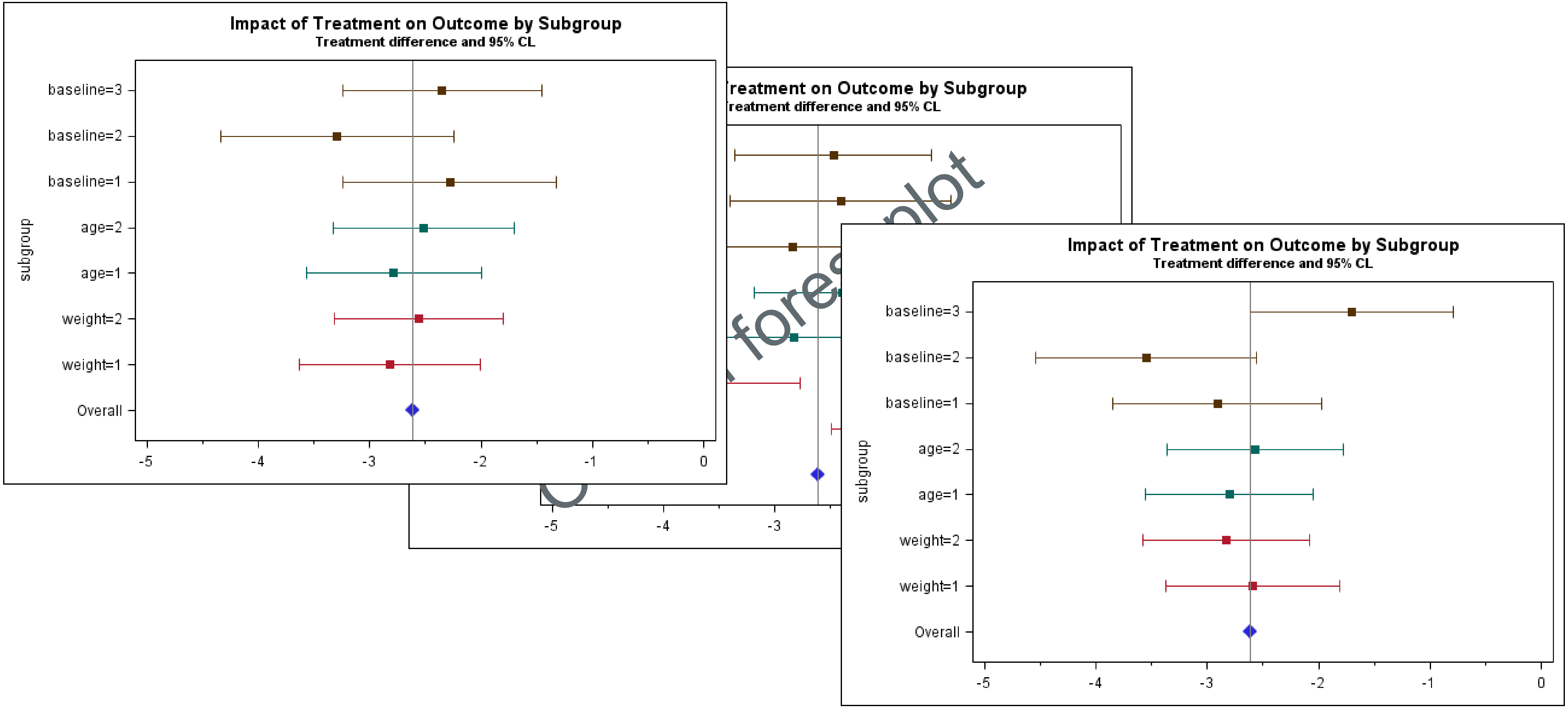


Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
subgroup
-5
-4
-3
-2
-1
0
Original forest plot

Like this….

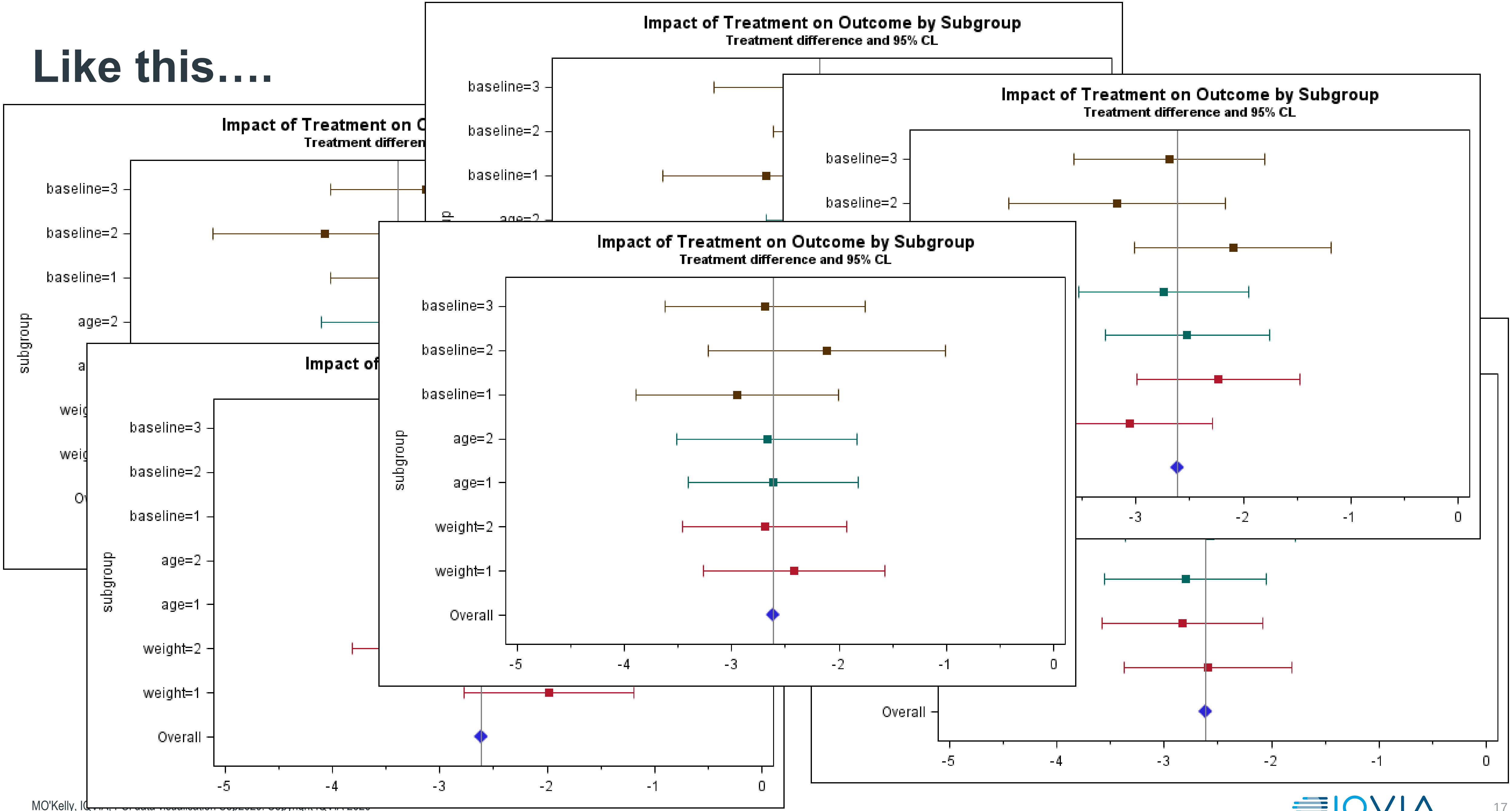

Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
subgroup
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
-5
-4
-3
-2
-1
0
reatment on Outcome by Subgroup
reatment difference and 95% CL
forest plot

Like this….



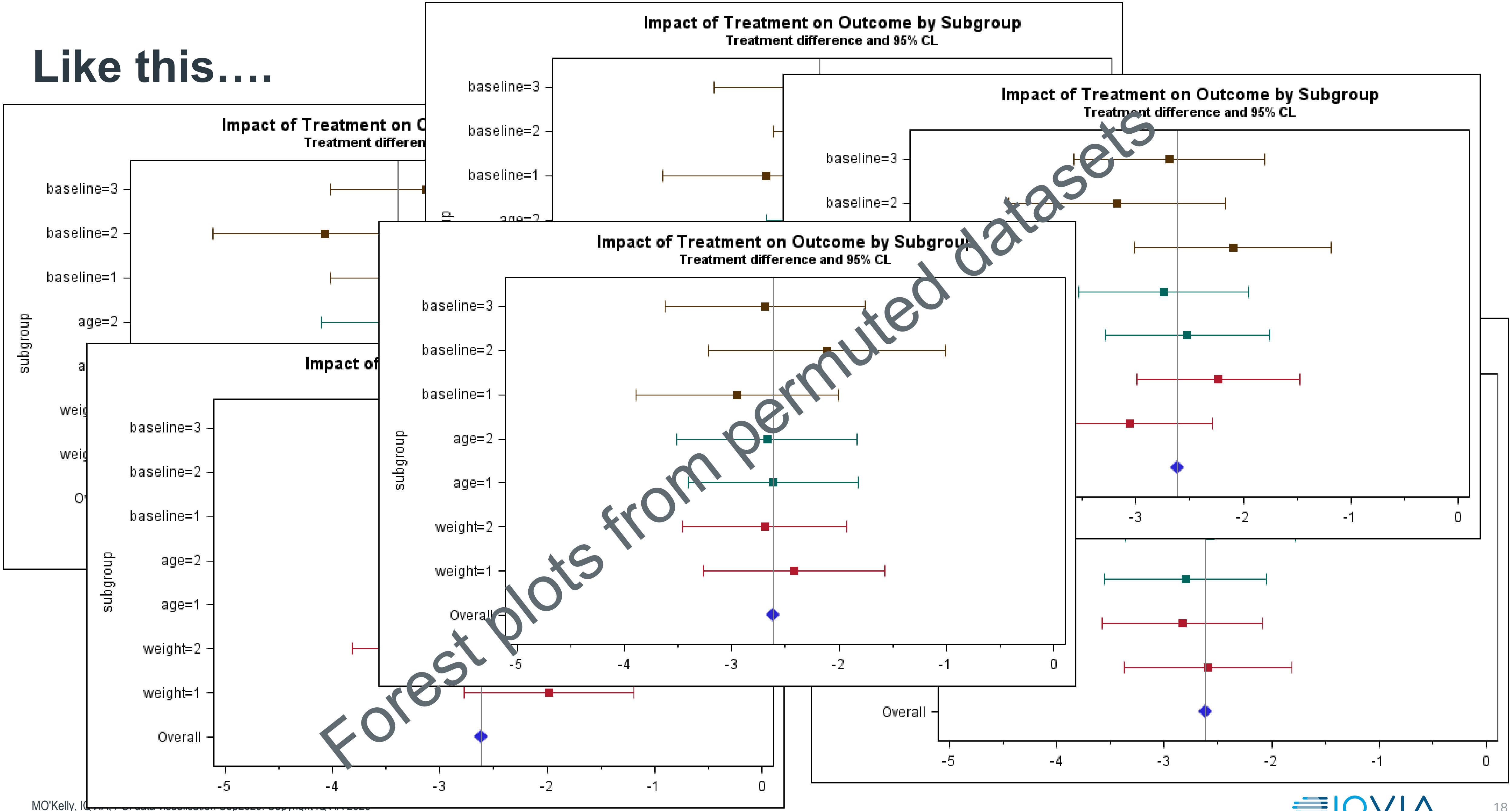
Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
subgroup
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
-5
-4
-3
-2
-1
0
Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
subgroup
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
-5
-4
-3
-2
-1
0

Like this….



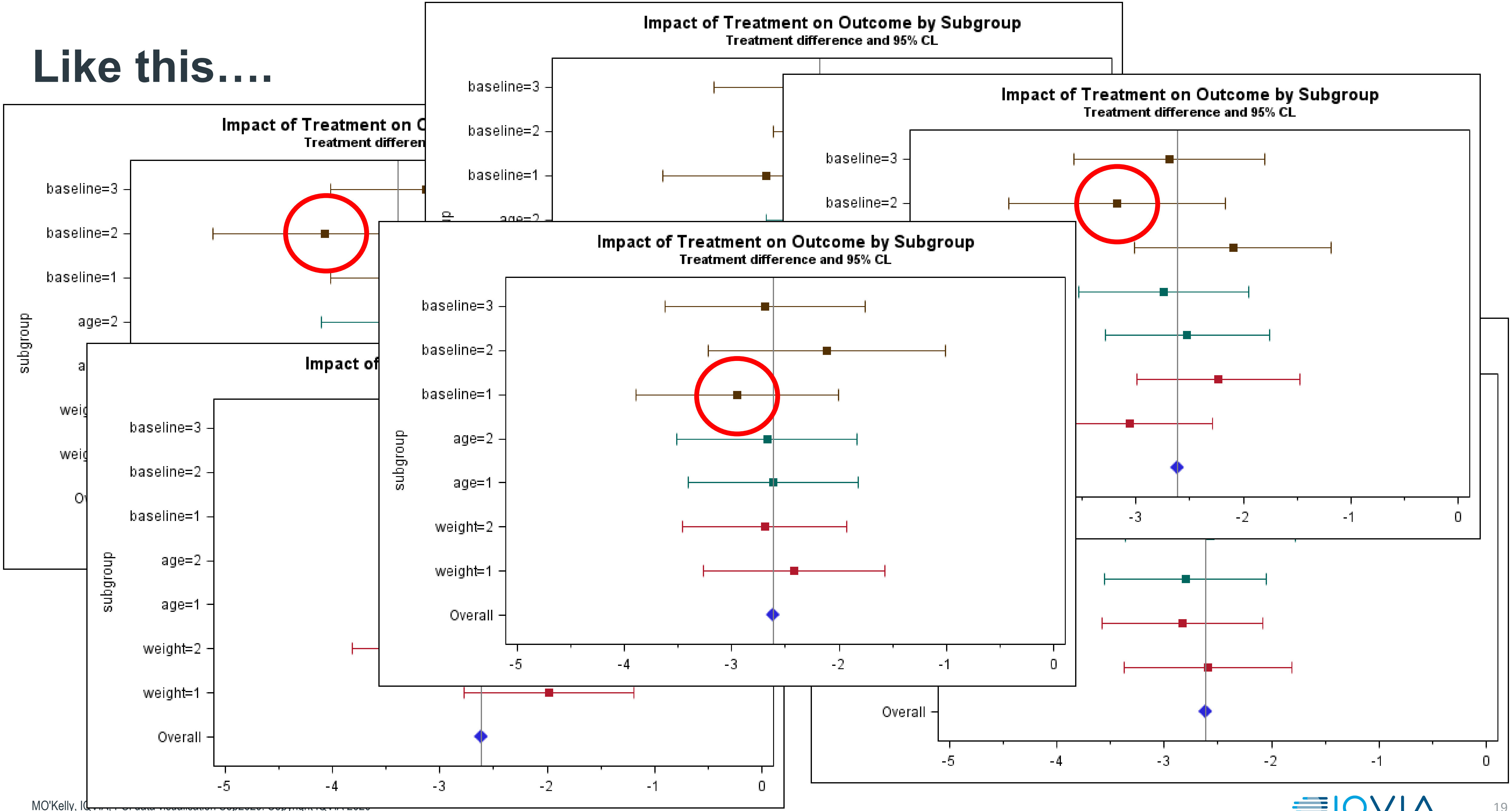
Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
subgroup
-5
-4
-3
-2
-1
0

Like this….



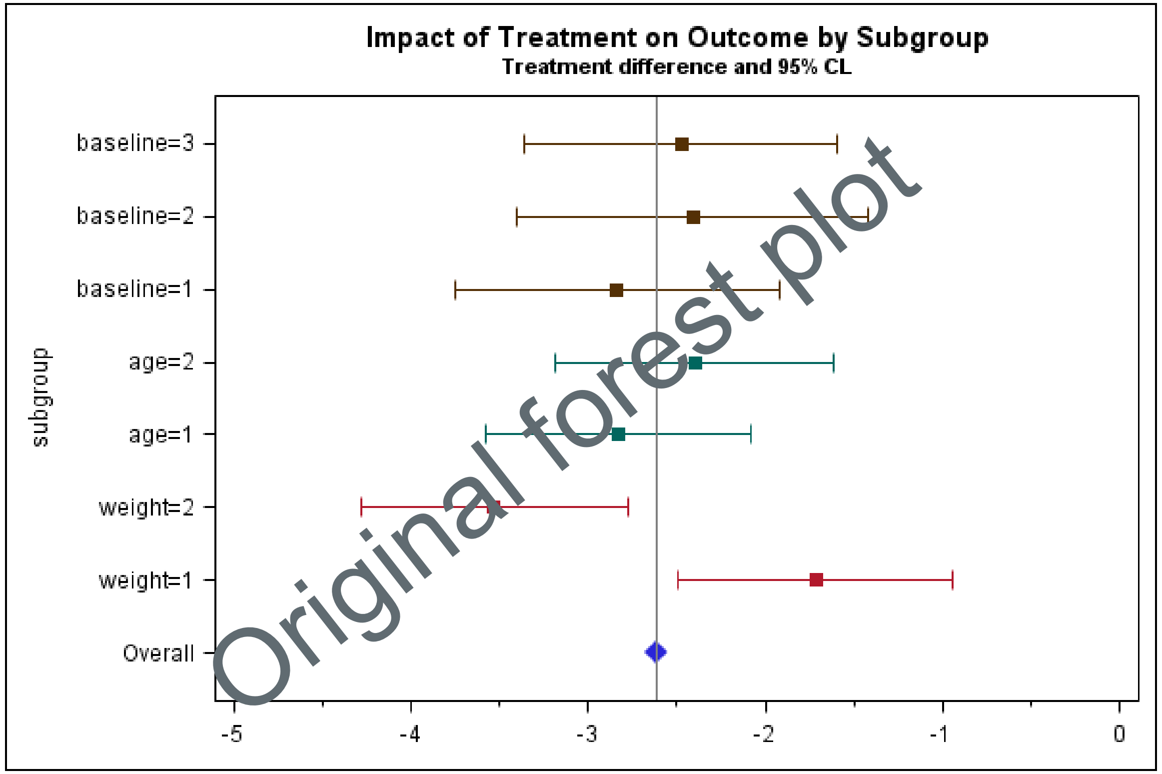
Forest plots from permuted datasets
Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
subgroup
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
-5
-4
-3
-2
-1
0

Like this….



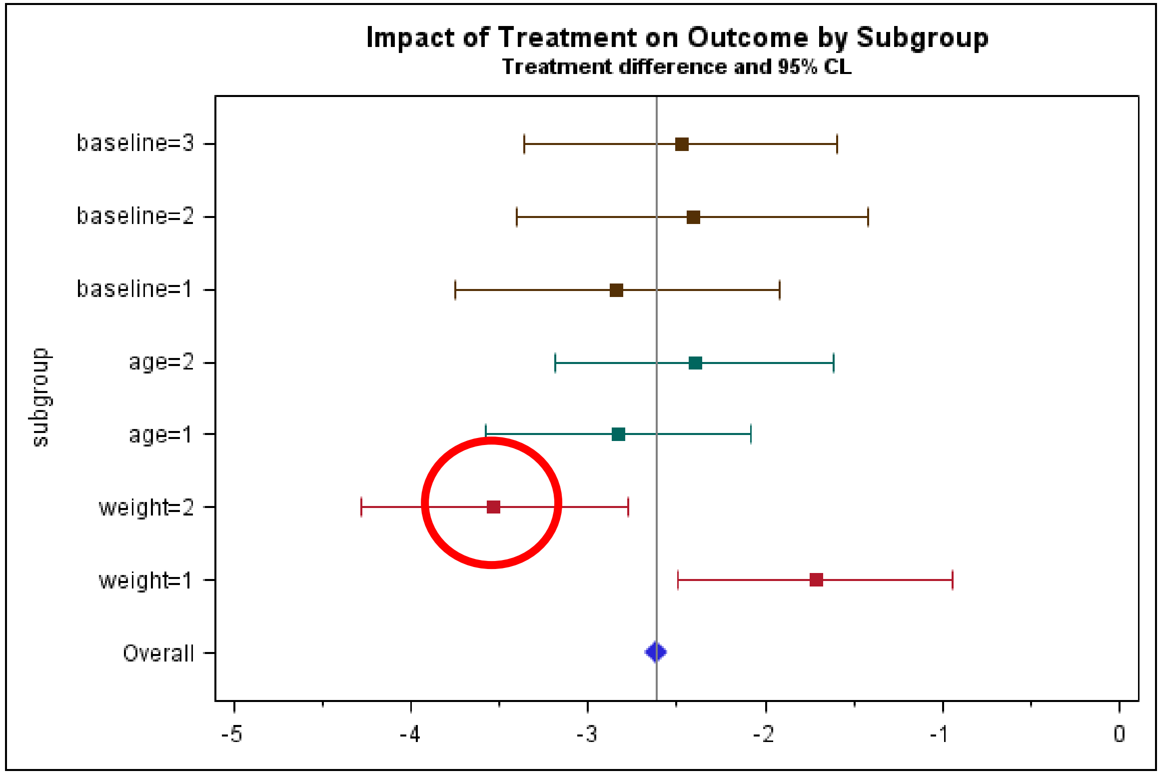
Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
subgroup
-5
-4
-3
-2
-1
0

Like this….

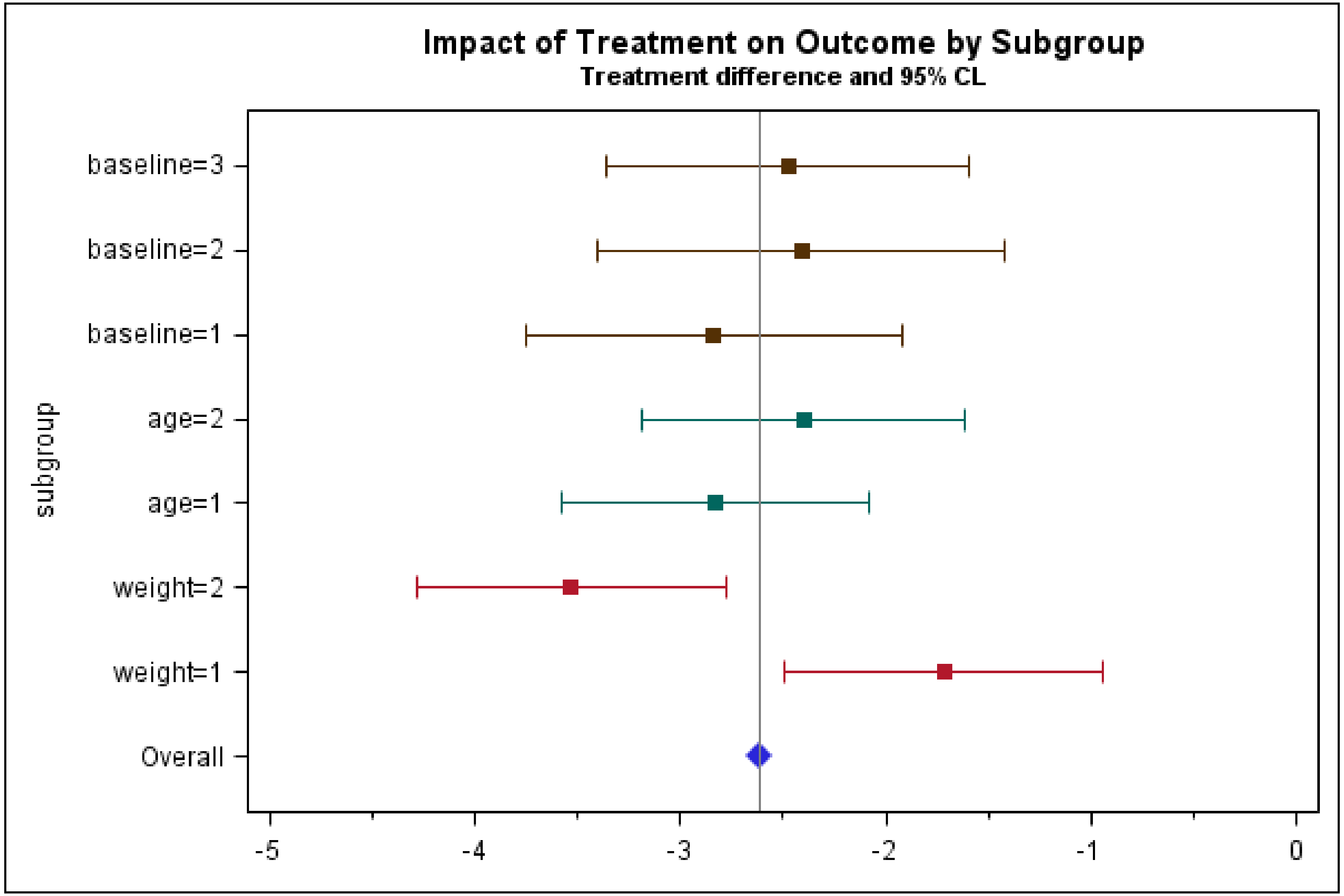


Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
subgroup
-5
-4
-3
-2
-1
0
Original forest plot

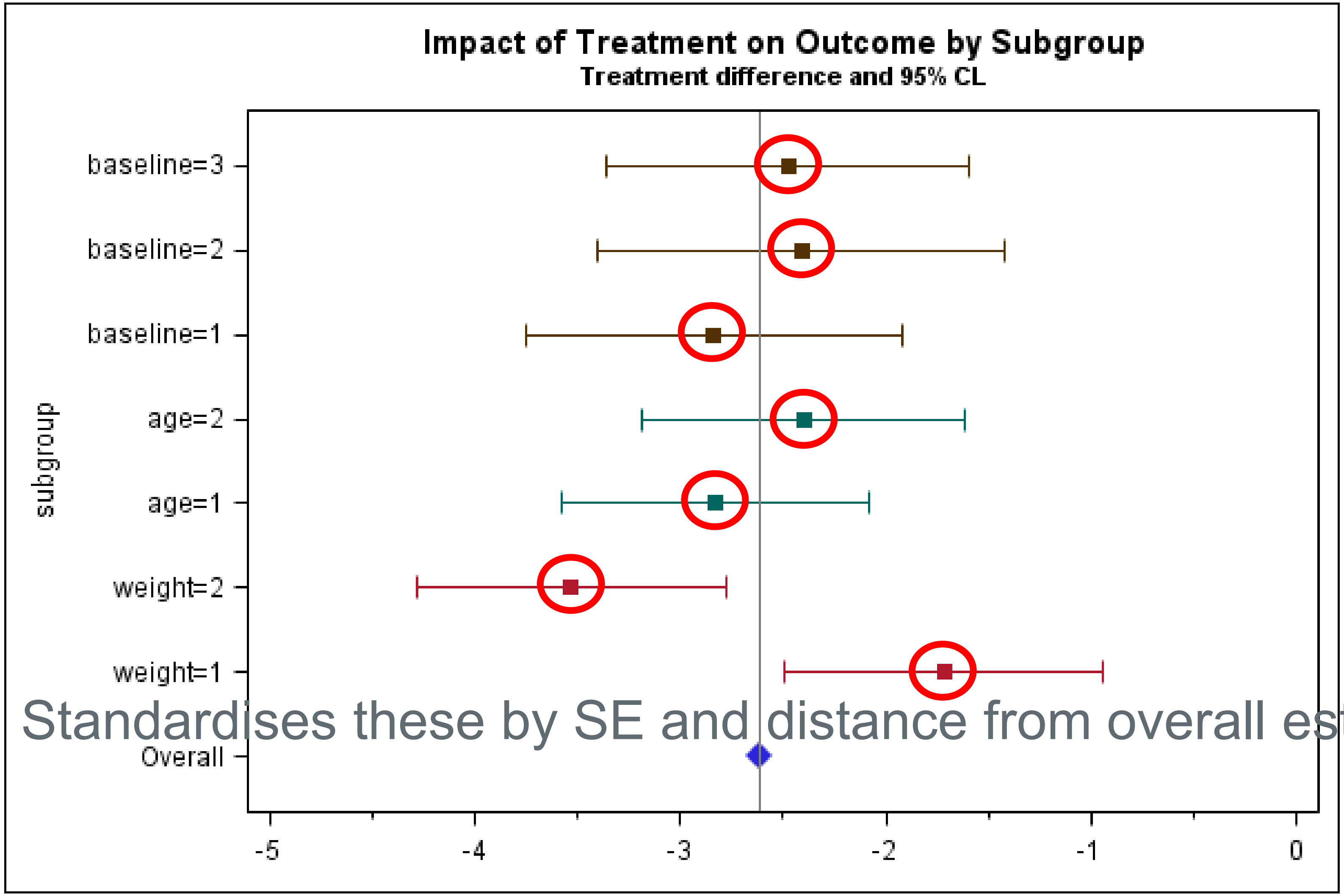
Like this….


Impact of Treatment on Outcome by Subgroup
Treatment difference and 95% CL
baseline=3
baseline=2
baseline=1
age=2
age=1
weight=2
weight=1
Overall
subgroup
-5
-4
-3
-2
-1
0

Actual SEAMOS approach, takes observed forest plot…

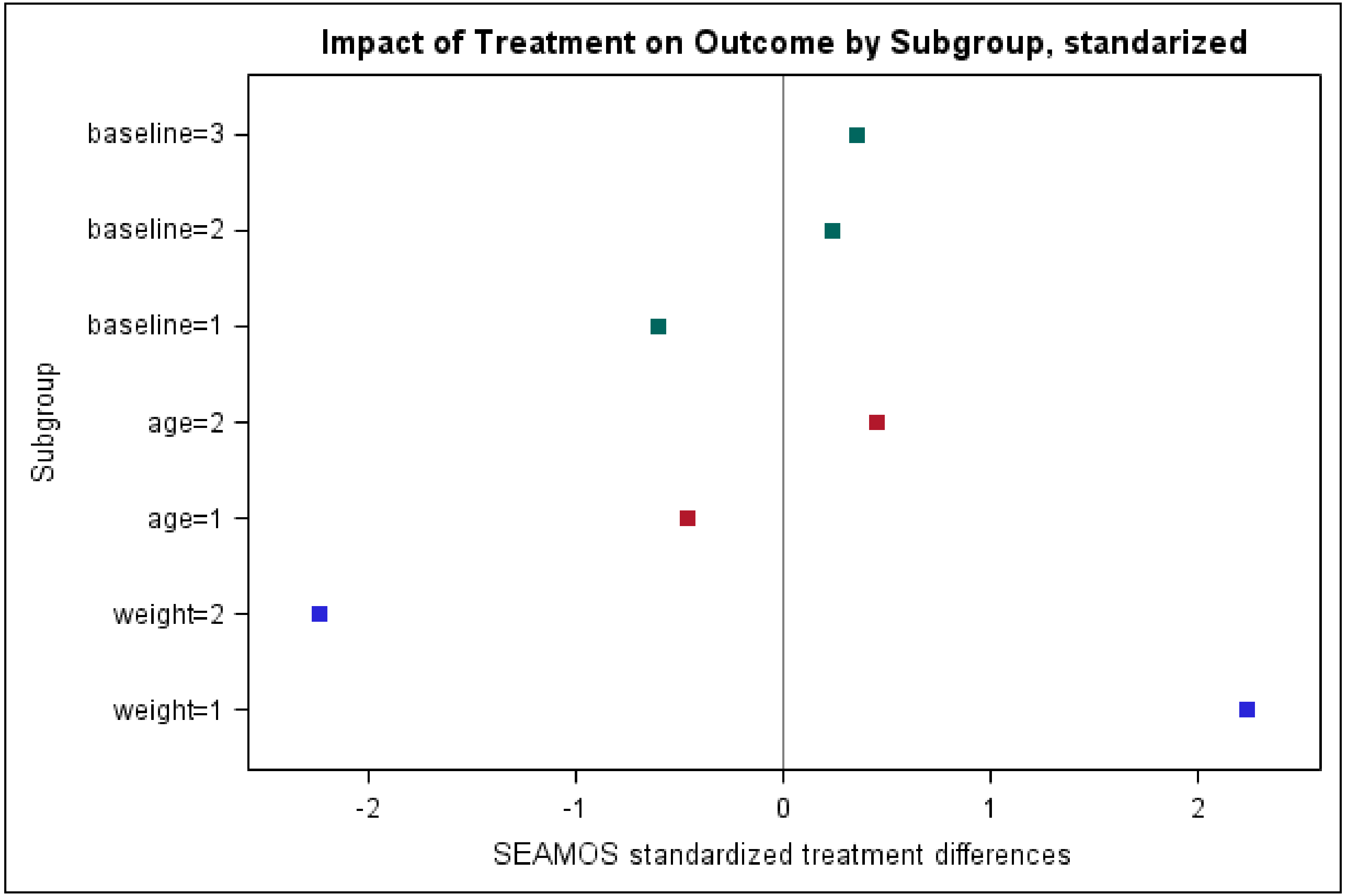


Actual SEAMOS approach, takes observed forest plot…

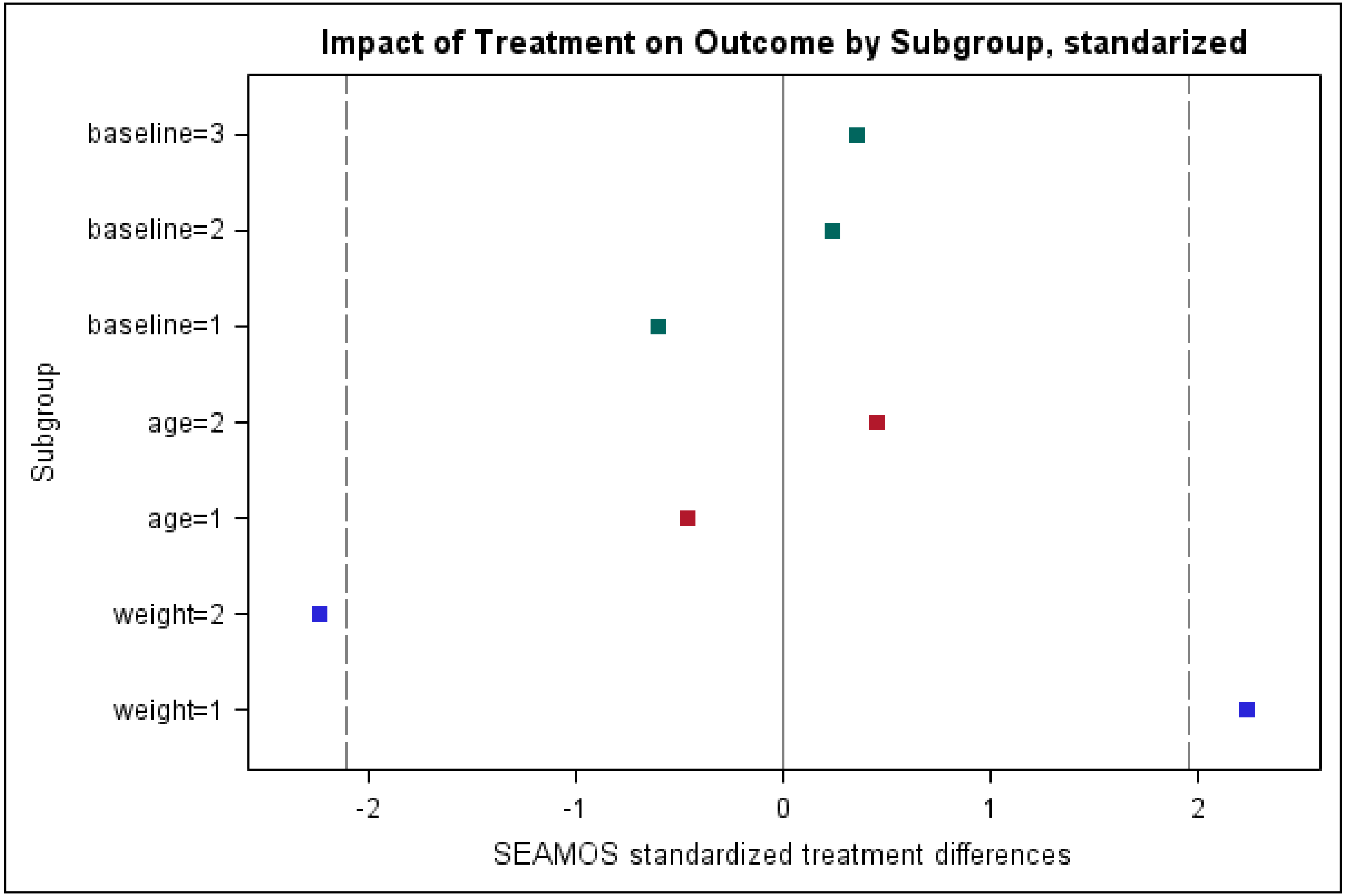


Standardises these by SE and distance from overall estimate, to get…

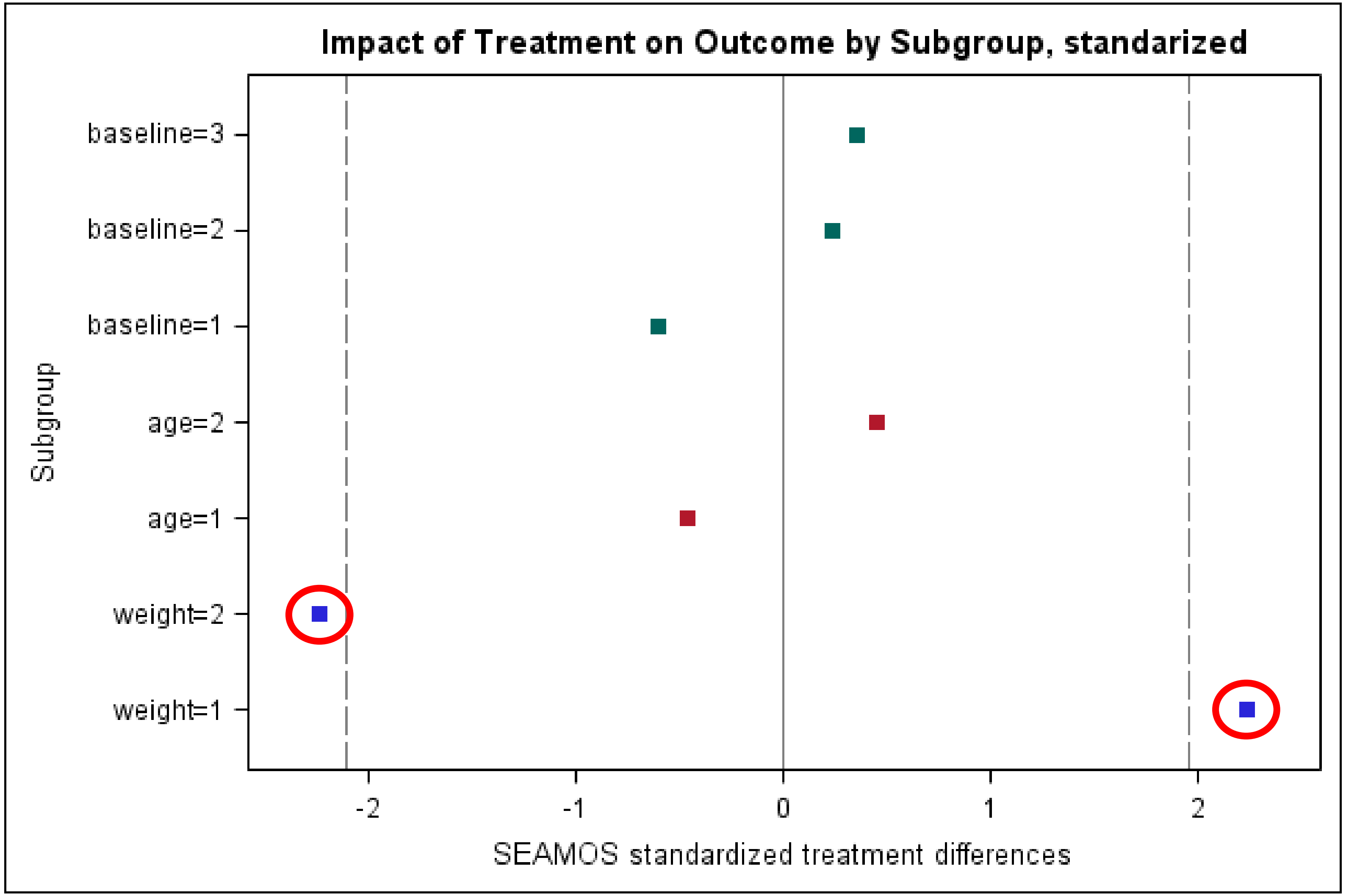
Actual SEAMOS standardised estimates



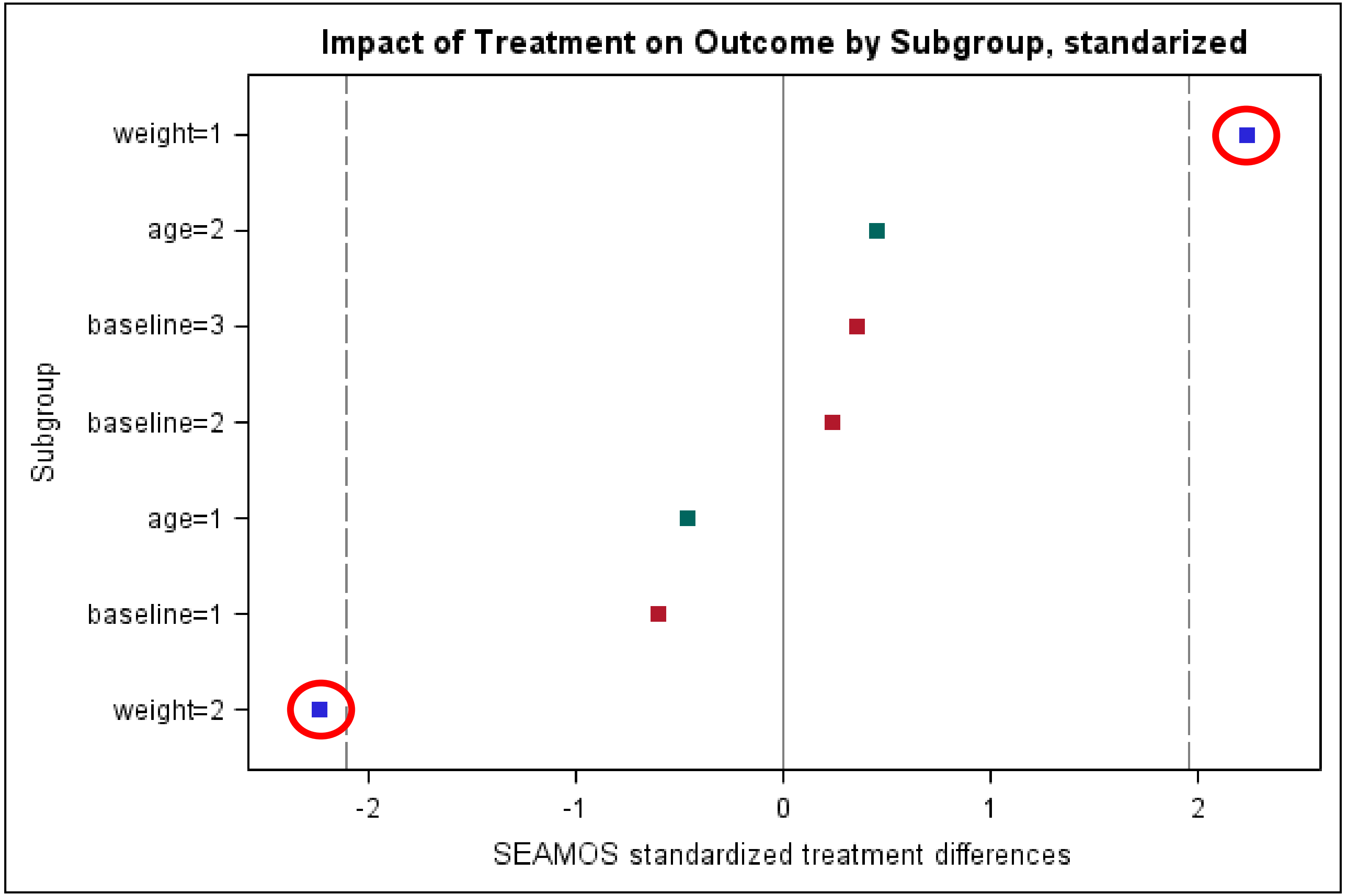
SEAMOS estimates with limits = 2.5% and 97.5% percentiles of permuted SEAMOS estimates



SEAMOS estimates with limits = 2.5% and 97.5% percentiles of permuted SEAMOS estimates



SEAMOS standardized differences are often presented ordered by difference



Objective of this research

- Dane et al. (2018) assess SEAMOS Type I error and power via simulated null-hypothesis and “true” subgroup scenarios
- Objective of this presentation:
  - Identify difficulties of SEAMOS’s model based approach, if any.
  - Further investigate nuances of Type I error
    - How numbers of *true* subgroups vs. numbers of subgroups *assessed*, affect Type I error and power
  - Compare the max(Δ_{ij}) statistic with alternative measures of existence of subgroup
    - Standard Type 3 sums of squares (SS) for the subgroup, Bonferroni corrected
    - Type 3 SS vs. Type 3 SS from permuted datasets

Simulations

- 1000 clinical trials simulated, each with 200 subjects per arm
- Outcome simulated as Y~ N(μ, σ^2)

$$Y = \beta_t T_t + \sum_{i=1,j=1}^{i=3,j=K_i} \beta_{ij} X_{ij} + \sum_{i=1,j=1,t=1}^{i=3,j=K_i,t=2} \beta_{ijt} X_{ij} T_t + e$$

- …where Y is the response, the β are coefficients for three subgroups X_i, i=1, 2, 3 (age (2 categories), weight (2 categories) and baseline value of Y (3 categories), i.e. K_1=2, K_2=2, and K_3=3; and T_t, t=1, 2 represents treatment group
- …with the β ε {0,1} depending upon scenario, and e ~ i.i.d N(0, 2^2)
- For some scenarios weight and age were correlated.
- Each simulated dataset permuted 200 times when estimating SEAMOS statistic.

Scenarios simulated, methods assessed

- Scenarios
  - Null hypothesis scenarios
    - With and without correlations $\rho=0.5$ between weight and age
  - Scenario with true subgroup(s)
    - Weight and age are true subgroups, but not baseline
      - Again, with and without correlations $\rho=0.5$ between weight and age
- Measures of existence of subgroup
  - SEAMOS's max(Δ_{ij})
  - Max(Type 3 SS) per subgroup, evaluated vs. permuted version of this statistic
  - Standard Type 3 SS (not permuted) with Bonferroni correction.
- One, two and three subgroups assessed.

Note on use of the permutation approach

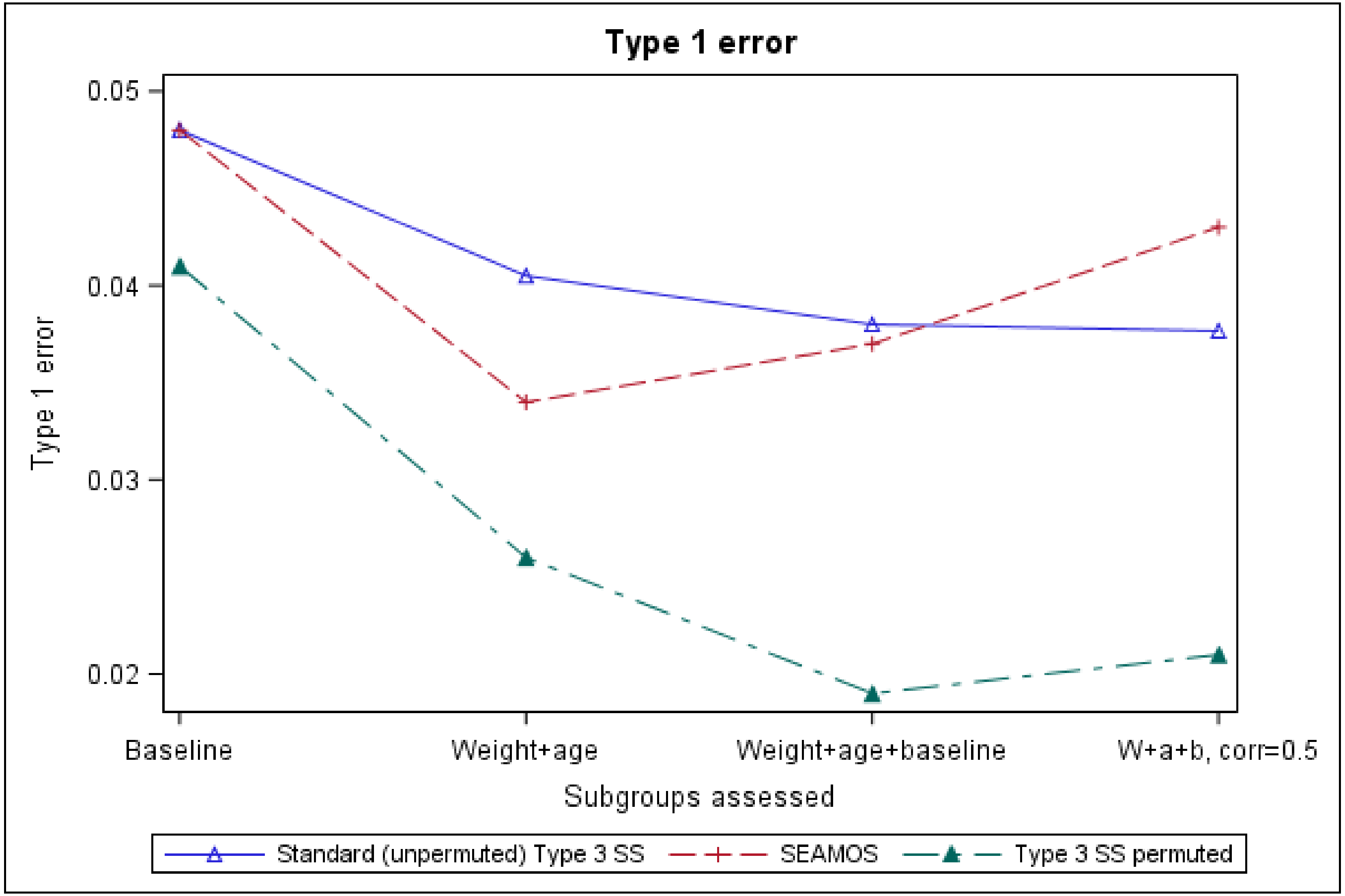
- In our scenarios, candidate subgroup variables are expected to be associated with outcome (i.e. expected to have a “main effect”).
- When a subgroup main effect is in the planned model, the permuted-dataset analysis cannot make use of this factor. Compared to the original non-permuted analysis, residual variance in permuted analysis is inflated: the Δ_{ij} and $\max(\Delta_{ij})$ in permuted datasets will tend to be smaller than that in the original forest plot.
  - =>Inflation of Type I error in identification of subgroups.
  - Solution: preserve the *original* subgroup variable for main subgroup effect, but using *permuted* subgroup variable for treatment-by-subgroup interaction.
  - Work ongoing on this by PSI Subgroups SIG.
- General principle: permuted test statistic should be estimated on same basis as original test statistic, except for existence of subgroup-by-treatment interaction.

Model used in the analysis of the simulated data sets

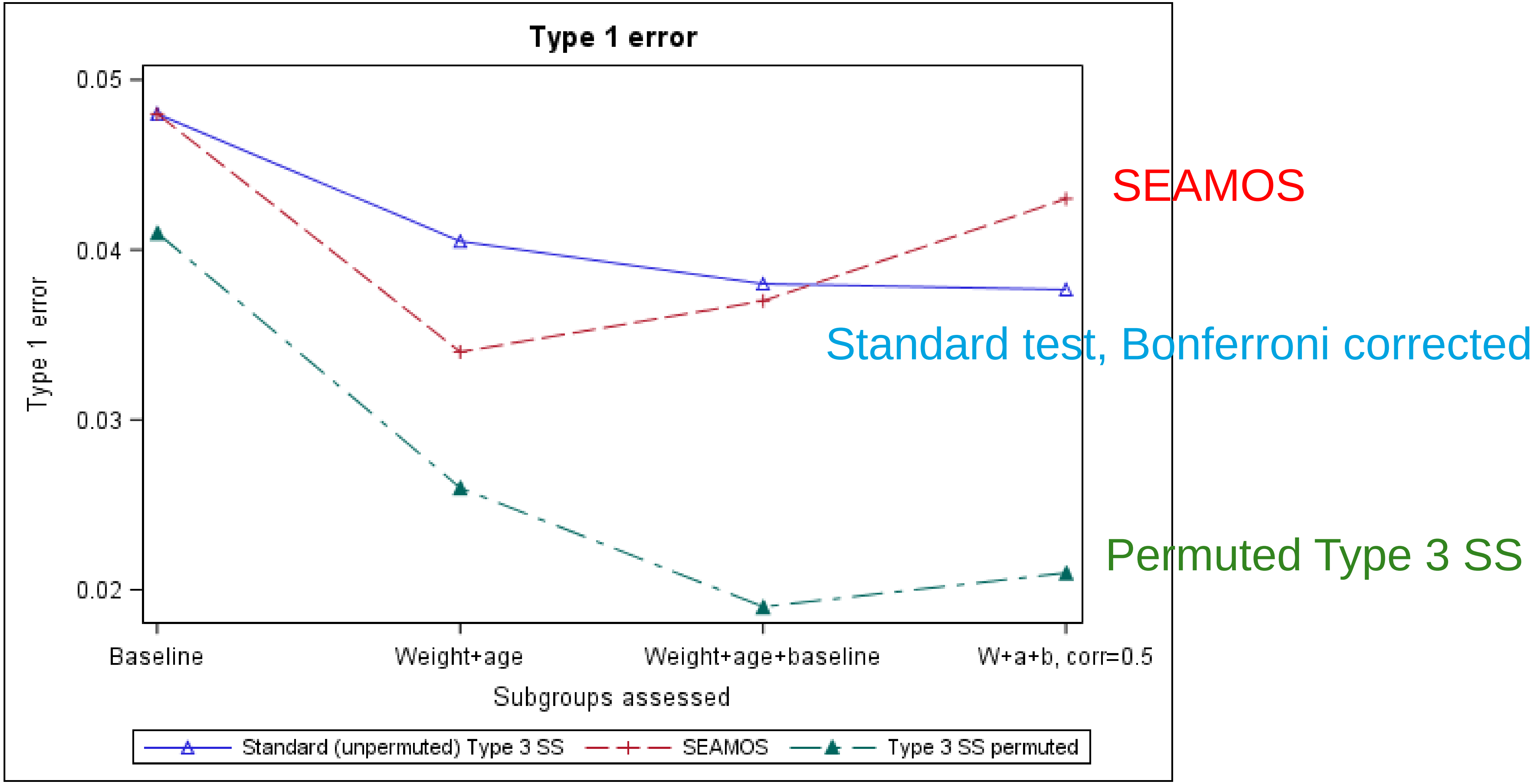
- For each candidate subgroup, test statistic was estimated using the following model

$$Y = \beta_0 + \beta_t T_t + \sum_{i=1,j=1}^{i=3,j=K_i} \beta_{ij} X_{ij} + \sum_{i=1,j=1,t=1}^{i=3,j=K_i,t=2} \beta_{ijt} X_{ij} T_t + \mathrm{e}$$

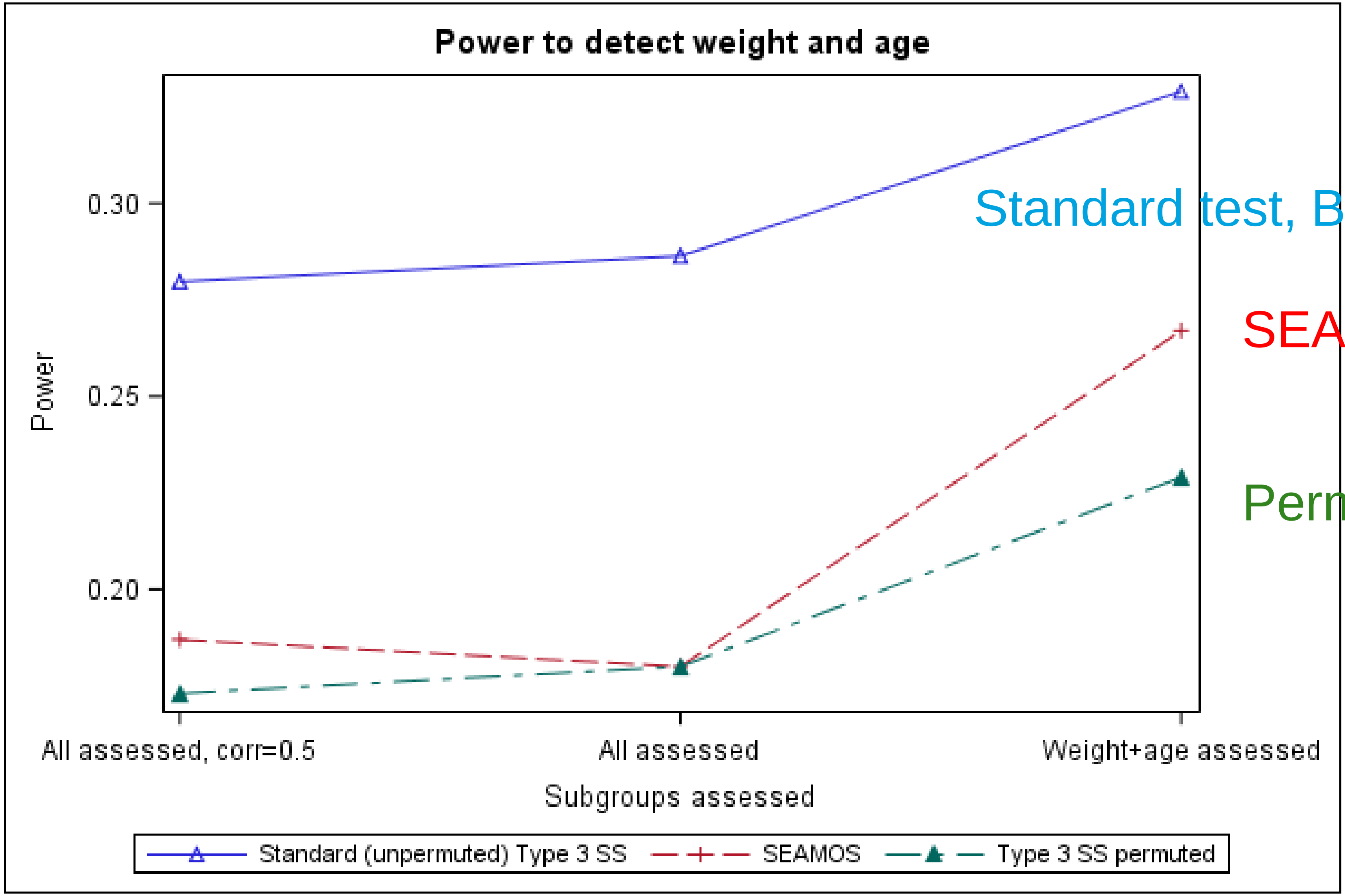
Results: Type 1 error, nominally two-sided 5% or equivalent



Results: Type 1 error, nominally two-sided 5% or equivalent



Results, power

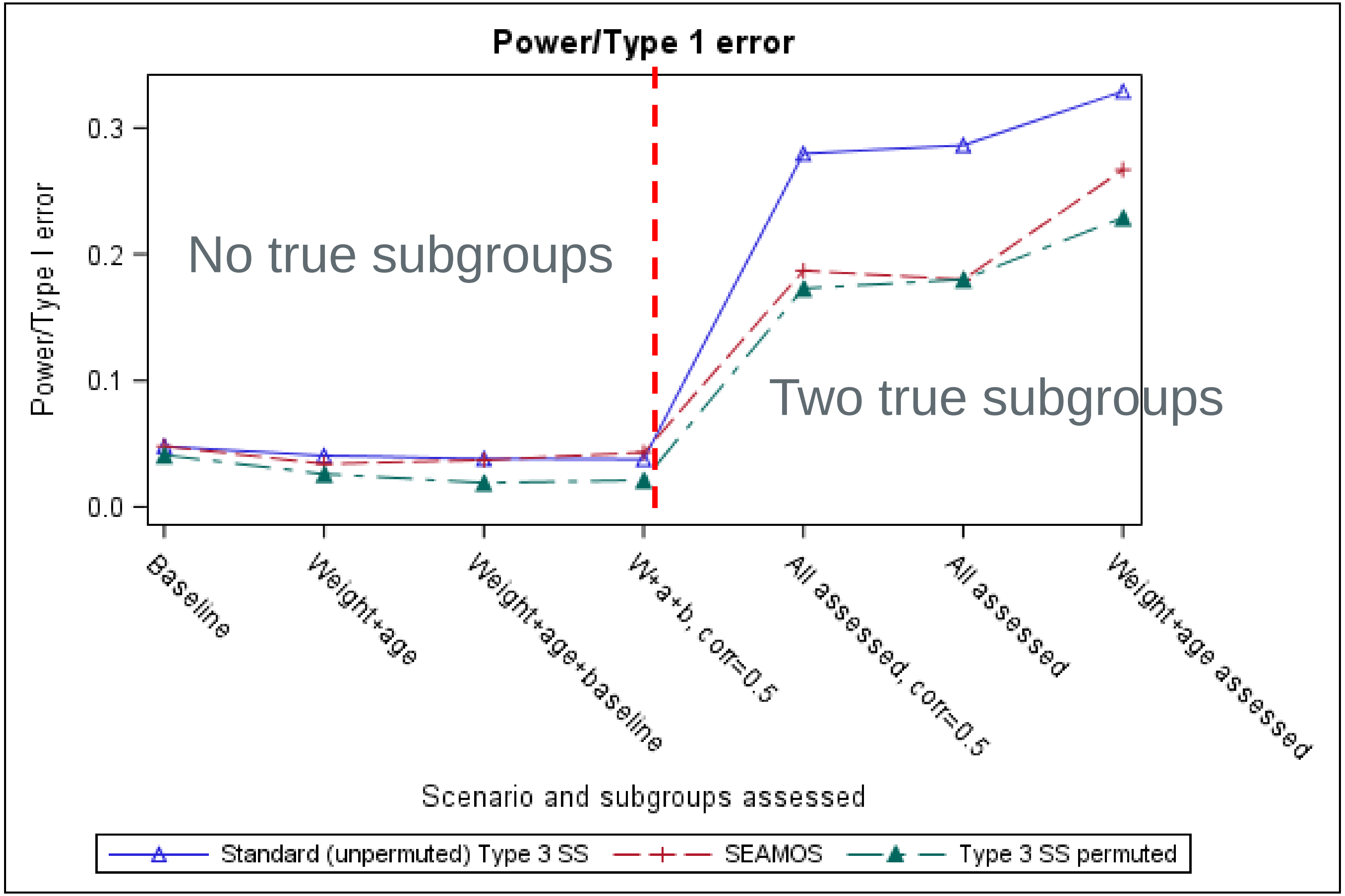


Standard test, Bonferroni corrected

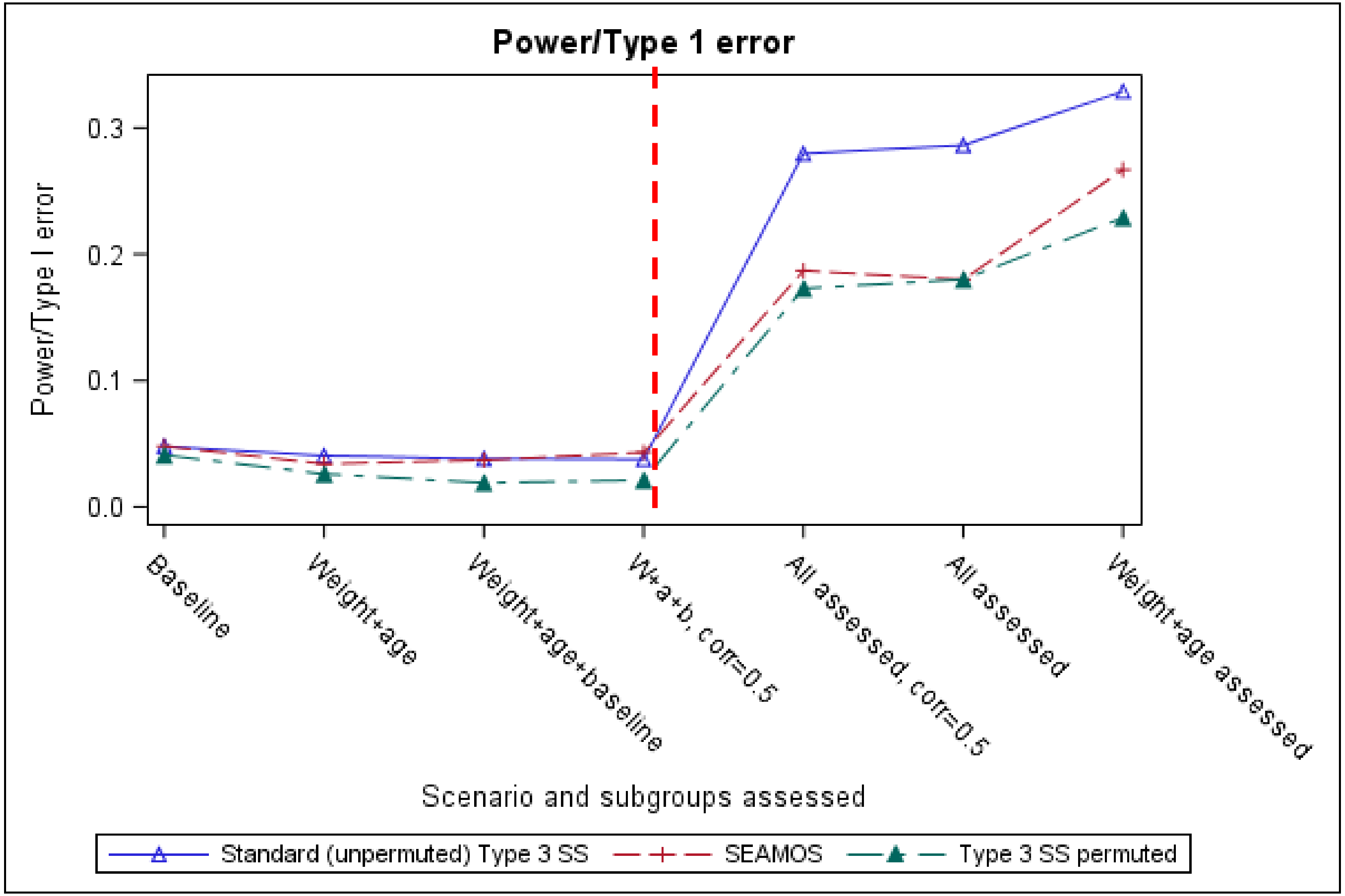
SEAMOS

Permuted Type 3 SS

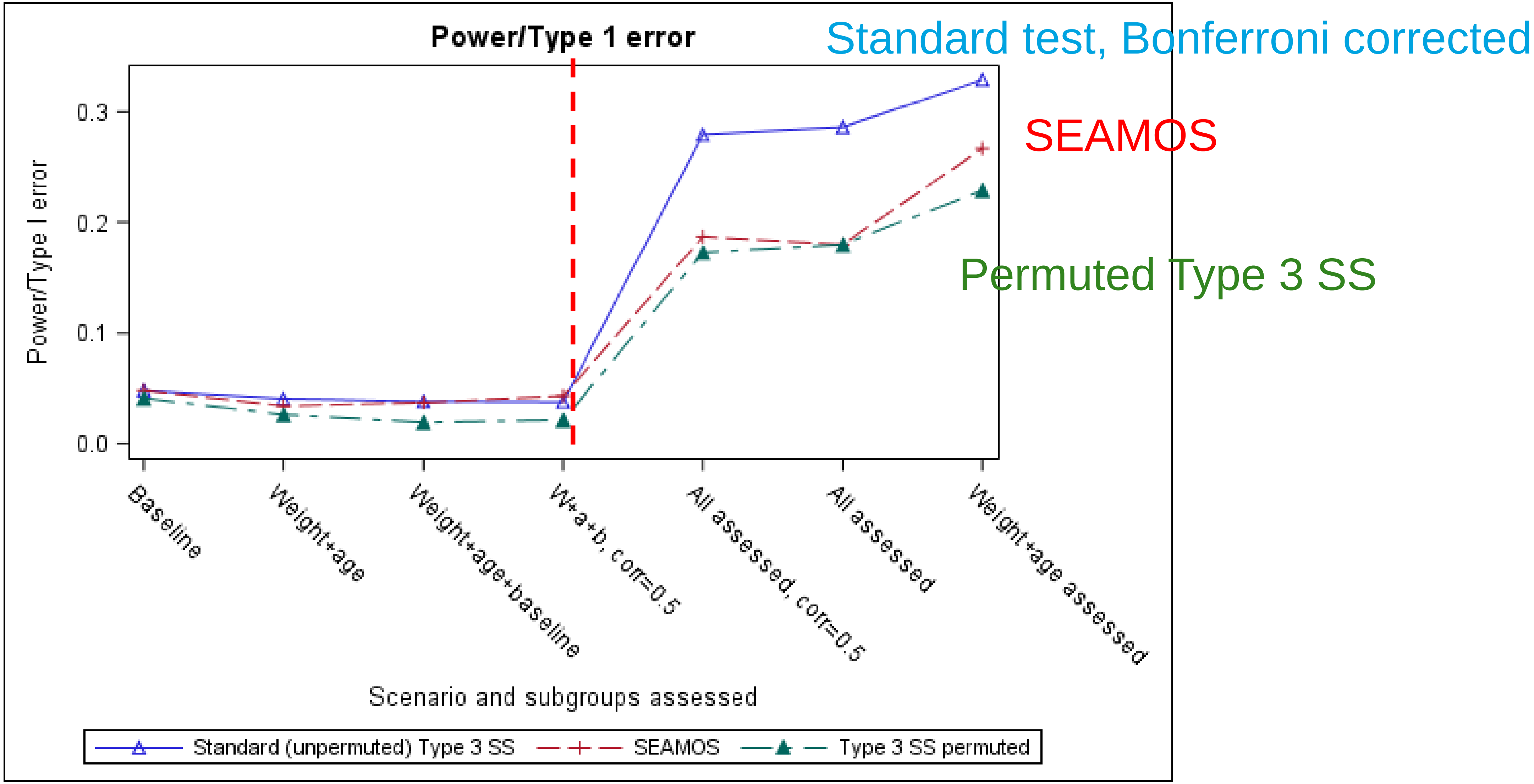
Results, combined



Results, combined



Results, combined



Limitations in our research

- This research was confined to a particular objective with regard to subgroup identification: assessment of forest plots.
- A relatively small number of subgroups were simulated/assessed.
- Scenarios where subgroups existed had relatively large proportion of true subgroups.
- Subgroups were assumed to have a main effect
- Did not compare the approaches with, e.g.,

  - bootstrap-based approach also described by Dane at al. (2018);

  - machine-learning approaches such as SIDES.
- Detection of existence of *any* subgroup in the forest plot was what was evaluated - did not evaluate whether the max(Δ_{ij}) came from a true subgroup.

Some learnings

- Standard Type 3 SS test for subgroups (with Bonferroni correction) controls “false positive” findings, even with correlated subgroups (as in Dane et al (2018)).
- Standard Type 3 SS test appears more sensitive than rival permutation-based approaches. (This test was less sensitive in the scenarios of Dane et al. (2018)).
- When using permutation based approaches, it is prudent to include unpermuted main effect of the subgroup when modelling the permuted datasets.
  - In our simulations, failing to do so inflated Type I error, where there is a subgroup main effect and main subgroup effect is in the model.
- For all approaches and scenarios assessed, presence/absence of correlation among subgroups has little association with changes in Type I error or power.
- As expected, the inclusion of extra “non-true” subgroups in the search erodes power for all approaches (e.g. 19% vs.27% for SEAMOS.

Some learnings

- SEAMOS is reasonably straightforward to implement in SAS.
  - Code created for this research is sharable, but user-friendly SAS macro has not been created.
- Permutation-based methods can be heavy on IT resources when used in simulations.
  - Hence our results limited to 200 permutations for each of the 1000 simulations.

Further research required

- Relative power of “standard” global interaction test (i.e. Type III SS test) vs. SEAMOS-type tests seems to vary depending on scenario – standard global test performed relatively well in the scenarios simulated here, compared to performance in scenarios of Dane et al. (2018).
  - Further work needed to identify characteristics of scenarios that suit standard subgroup approaches vs. SEAMOS-type approaches when assessing forest plots.

Questions?