500 Class 05

https://thomaselove.github.io/500-2024/

2024-02-15

Today's Agenda

- The `toy` and `lindner` examples
- The `dm2200` example: Executing Matching in R
  - Using `Matching` vs. `MatchIt`
- Matching with (or without) a Propensity Score
  - Matching with and without replacement
  - Greedy vs. Caliper vs. Optimal Matches
- Feinstein's Model for Research Architecture
- A Few Words on Extensions to Matching
  - Constrained and "Almost Exact" matching
  - "Fine Balance" in matching
  - Full Matching and Hansen's (2004) Example

Section 1

Our R Examples

The `toy` example

The `toy` example presents methods for doing 1:1 greedy matching without replacement using the `Match` function from the `Matching` package, and for evaluating the balance before and after matching with `cobalt` and with an alternative strategy for obtaining Love plots.

- The example uses 3 Rules I attribute to Rubin (2001) for determining when a sample comparison shows sufficient balance to allow for a reasonable regression model for the outcome.
  - **Please read** Rubin (2001) in advance of Class 6, which will mostly be about that example.
- What to do in terms of a sensitivity analysis is discussed in the final section of that example, and we'll get to that later on.

The `lindner` example

The `lindner` example also does 1:1 greedy matching on the linear propensity score, in this case however there are more treated subjects than controls, so the default greedy approach taken is to not match all treated subjects, since you run out of controls before you get to them all.

- 1:1 nearest neighbor matches are demonstrated, first without replacement (in Task 4) and then **with** replacement (in Task 6).
- The "with replacement" matching has some appealing features, not least of which being that now we can realistically create ATT estimates (since we match all treated subjects to a control, repeating some controls.)
  - The match quality (in terms of a Love plot) is much better with replacement in this case.
  - This does come with a cost, though. Most subjects are matched just a few times, but some are included more than 50 times!

The dm2200 example

This example is solely about matching, and not things like subclassification, weighting and so on. The data are simulated, again, and I'm not really focusing on the outcome models, but rather just on demonstrating how to do the matching, and how to evaluate the quality of the covariate balance.

dm2200: First Four Matches

We demonstrate the use of both the `Matching` package and the `MatchIt` package. First, using `Matching`, which is what `toy` and `lindner` use, we show:

1. 1:1 nearest neighbor matching without replacement
2. 1:2 nearest neighbor matching without replacement
3. 1:3 nearest neighbor matching with replacement
4. 1:1 nearest neighbor matching without replacement within a caliper on the (linear) propensity score

dm2200: Four Additional Matches

Using the `MatchIt` package, we demonstate:

1. 1:1 nearest neighbor matching without replacement
2. 1:1 nearest neighbor matching without replacement within a caliper on the (linear) propensity score
3. 1:1 optimal matching without replacement
4. 1:2 optimal matching without replacement

There are multiple other packages in the world for propensity matching, and `cobalt` describes and supports several of them, but these are the two I have most often used.

- `MatchIt` has some features I really like, in particular, it's easier to work with it in `cobalt`, I think, and it also has one very annoying feature, in that it's hard to get the matched data set from it.

Section 2

Matching Approaches (discussion built on Austin, 2014)

1:1 Greedy Matching

Greedy (or nearest neighbor) matching selects a treated subject and then selects as a matched control subject the untreated subject whose propensity score is closest to that of the treated subject. If multiple untreated subjects are equally close to the treated subject, one of these untreated subjects is selected at random, typically. Options include:

1. Select treated subjects from highest to lowest propensity score.
2. Select treated subjects from lowest to highest propensity score.
3. Select sequentially treated subjects in the order of the best possible match.
    - First treated subject is the one who is closest to an untreated subject.
    - Second treated subject is the one closest to the remaining untreated, etc.
4. Select treated subjects in a random order. Set a fixed random number seed so that the matched sample is reproducible in subsequent analyses.

Results in all treated subjects being matched to a single control.

Greedy Matching with Replacement

Matching *without* replacement means that once an untreated subject has been matched to a treated subject that untreated subject is no longer eligible for further matches to other treated subjects. As a result, each subject can be in at most one matched pair.

Now, in matching *with* replacement, we allow members of the "control" pool to be reused in the matching process.

- The process is somewhat simpler in the nearest neighbor case - just match each treated subject to the closest untreated subject.
- Because untreated subjects are recycled and thus can be included in multiple matched sets, the order in which the treated subjects are selected has no effect on the formation of matched pairs.

Matching 1:k rather than 1:1

Here, we simply try to obtain the k best matching untreated subjects for each treated subject.

- In greedy matching, it is certainly possible that the quality of matches will drop considerably with extra matches, especially near the edges of the distribution of the propensity score.
- 1:k matching is occasionally done with replacement, but of course we still want k unique matched untreated subjects for each treated subject.

Caliper Matching

Match subjects only if they fall within a pre-specified maximum distance (the caliper distance.)

- When using caliper matching, we usually match subjects on the logit of the propensity score using a caliper width as a proportion of the standard deviation of the logit of the propensity score.
- Caliper matching can be combined with other distance metrics (where, for example, a few specific covariates are targeted for more precise matching.)
- Matching with a caliper can be accomplished with or without replacement, and in 1:1 or 1:k settings.

Optimal Matching

The main distinction that matters is between optimal matching approaches and nearest-neighbor (greedy) matching approaches.

- Optimal matching forms matched pairs so as to minimize the average within-pair difference in propensity scores.
- Optimal matching is rarely the first way I run an analysis (it's a bit slow, especially with large matching problems) but this problem is disappearing as smarter people and more effective computers emerge.

Double Robust Approaches

Nothing is stopping us from using regression adjustment along with matching. It's not unusual to consider the incorporation of the linear propensity score, or an important set of prognostic covariates in a setting where we are analyzing propensity-matched subjects.

Peter Austin's (2014) Comparison

A comparison of 12 algorithms for matching on the propensity score

Peter C. Austin[a,b,c*†]

Propensity-score matching is increasingly being used to reduce the confounding that can occur in observational studies examining the effects of treatments or interventions on outcomes. We used Monte Carlo simulations to examine the following algorithms for forming matched pairs of treated and untreated subjects: optimal matching, greedy nearest neighbor matching without replacement, and greedy nearest neighbor matching without replacement within specified caliper widths. For each of the latter two algorithms, we examined four different sub-algorithms defined by the order in which treated subjects were selected for matching to an untreated subject: lowest to highest propensity score, highest to lowest propensity score, best match first, and random order. We also examined matching with replacement. We found that (i) nearest neighbor matching induced the same balance in baseline covariates as did optimal matching; (ii) when at least some of the covariates were continuous, caliper matching tended to induce balance on baseline covariates that was at least as good as the other algorithms; (iii) caliper matching tended to result in estimates of treatment effect with less bias compared with optimal and nearest neighbor matching; (iv) optimal and nearest neighbor matching resulted in estimates of treatment effect with negligibly less variability than did caliper matching; (v) caliper matching had amongst the best performance when assessed using mean squared error; (vi) the order in which treated subjects were selected for matching had at most a modest effect on estimation; and (vii) matching with replacement did not have superior performance compared with caliper matching without replacement. © 2013 The Authors. ***Statistics in Medicine*** **published by John Wiley & Sons, Ltd.**

Keywords: propensity score; matching; computer algorithms; optimal matching; Monte Carlo simulations; propensity-score matching

- You'll find this article on our Sources page.

Austin's conclusions re: 12 Algorithms

- Larger numbers of matched pairs (from complete optimal or complete greedy matches) yields more precise estimates than smaller numbers of matched pairs (say, when a caliper is used and only some treated subjects are matched.)
- Caliper matching often yields better "balance" and less biased estimates as compared to other algorithms.
- So we have a bias - variance tradeoff in our estimation strategies, but in terms of MSE, caliper matching usually performs pretty well.
- In terms of ordering of treated subjects for greedy matching or caliper matching, random selection is competitive with other options.
- Optimal matching is pretty comparable to nearest neighbor matching with random selection order, and in fact, it's not clearly any better than that approach.

Section 3

Feinstein's Model for Research Architecture (expanded by Neal Dawson)

7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation

Assembly

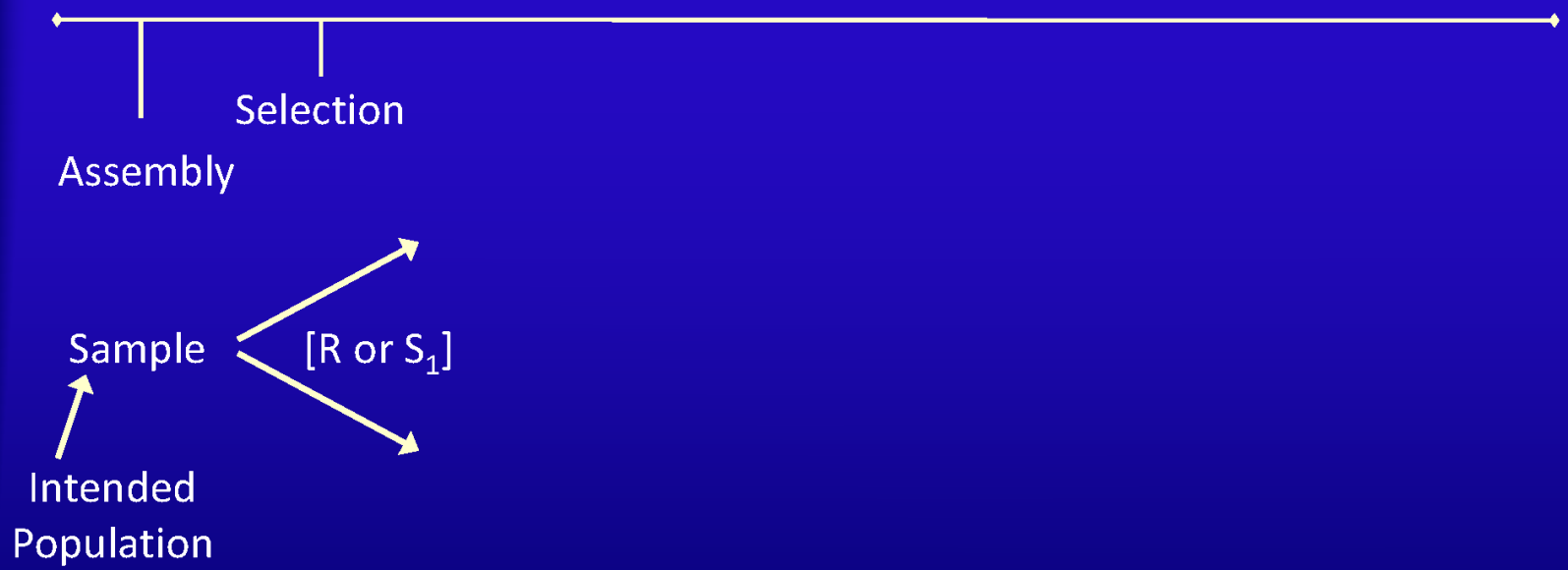
Sample

Intended Population

- Possibility of distorted assembly – sample doesn't reflect the population to which the results will be generalized.

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

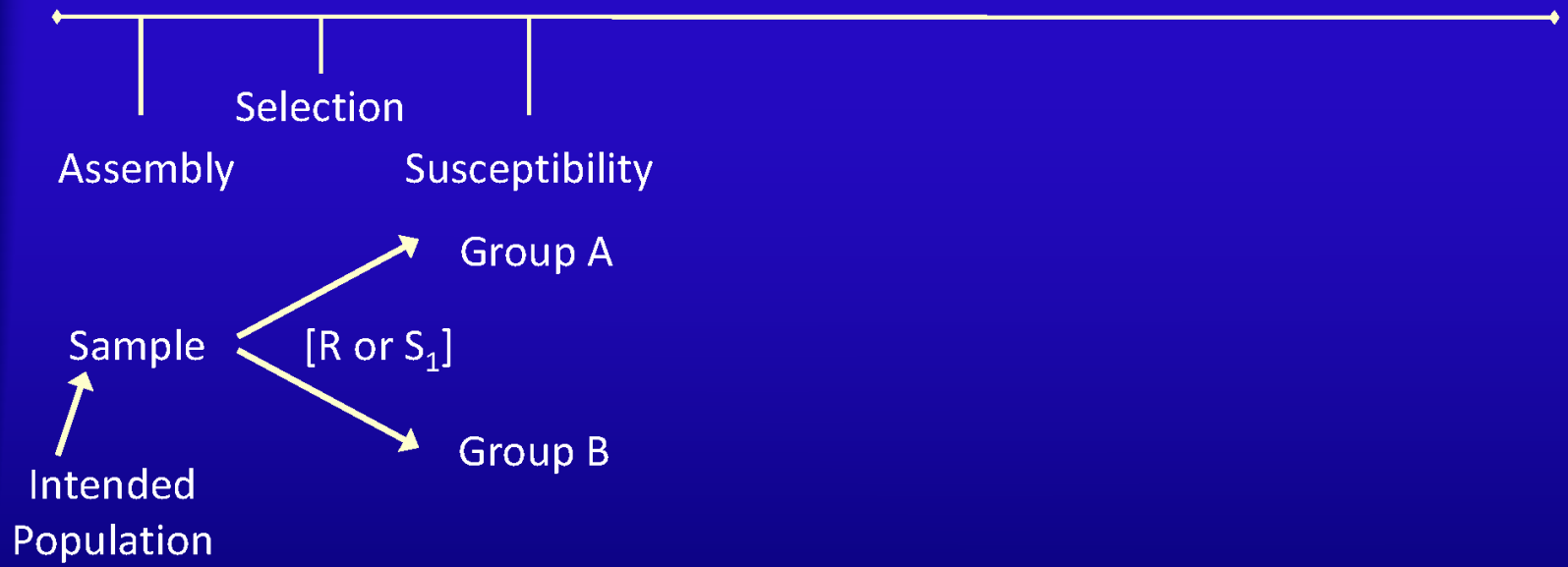
7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation



- Selection Bias – who receives the exposure? Basis: (possibly unmeasured) covariates linked to outcomes? Why randomize?

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

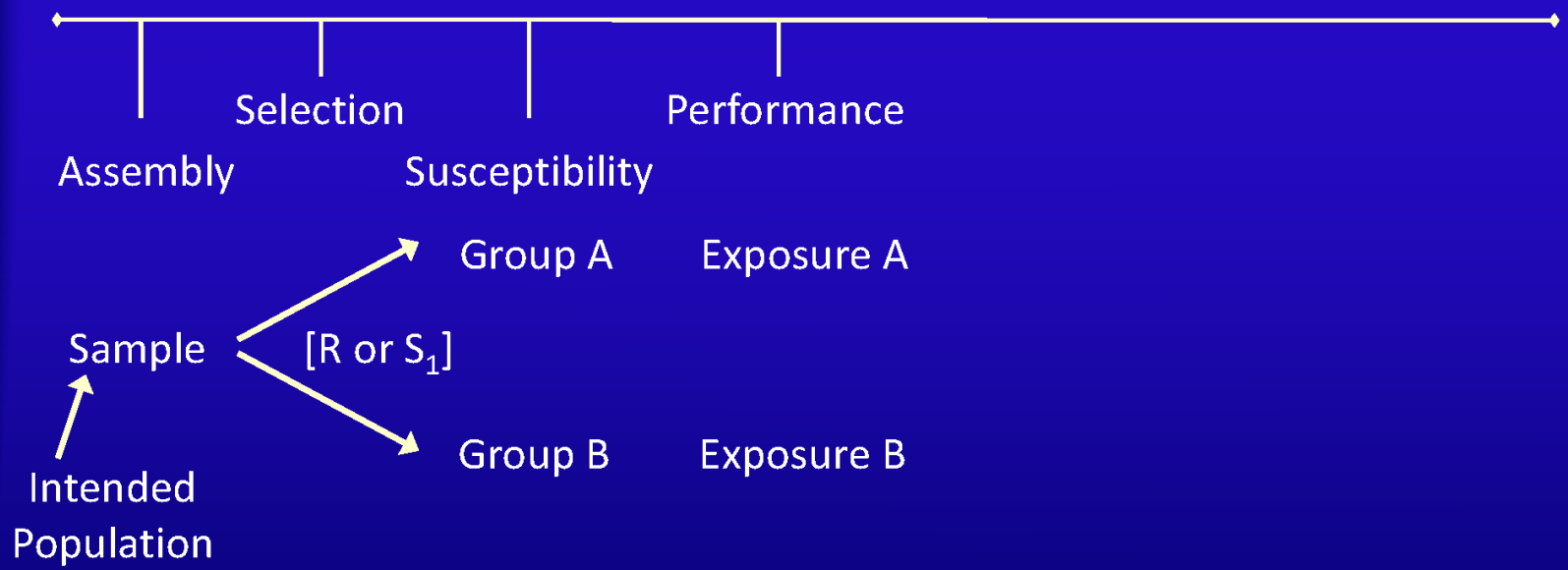
7 Key Aspects of Research Architecture*: Making Fair Judgments about Causation



- Are there importantly different expectations at baseline, for the eventual outcomes? Susceptibility reflects covariate differences.

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

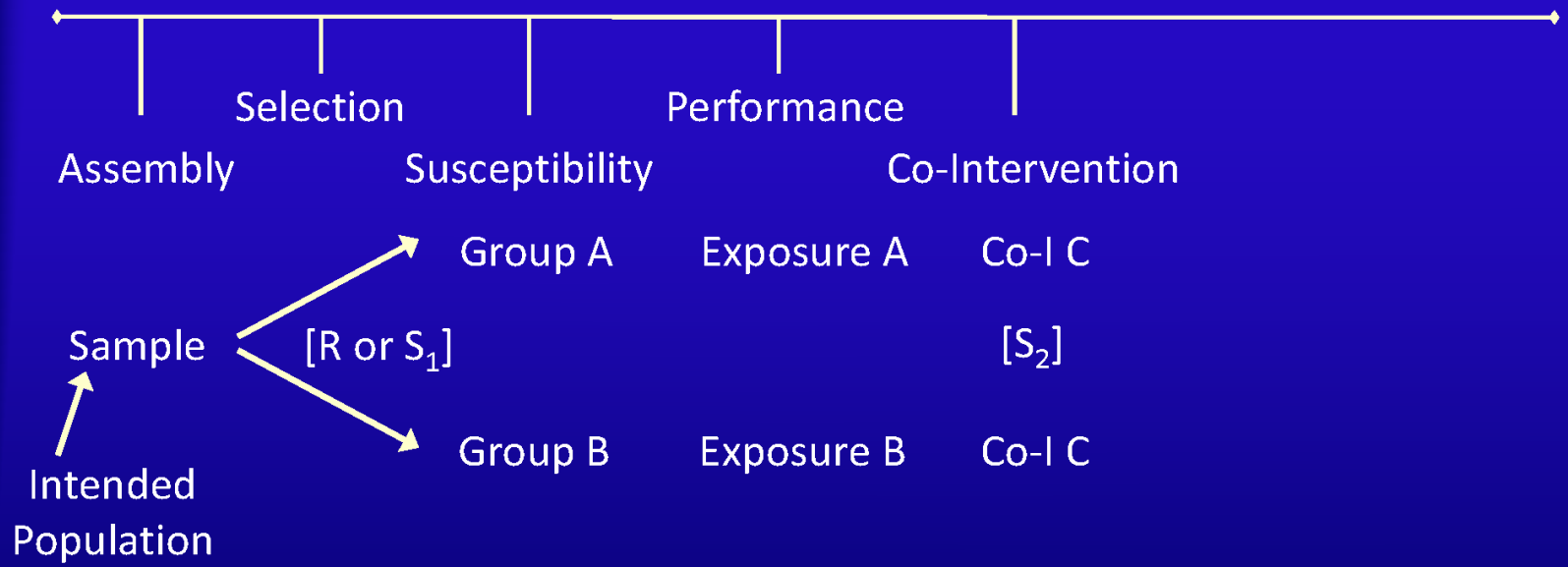
7 Key Aspects of Research Architecture*: Making Fair Judgments about Causation



- Are exposures applied with the same proficiency? How "well" do pts receive the exposures (dosage schedules, compliance)?

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

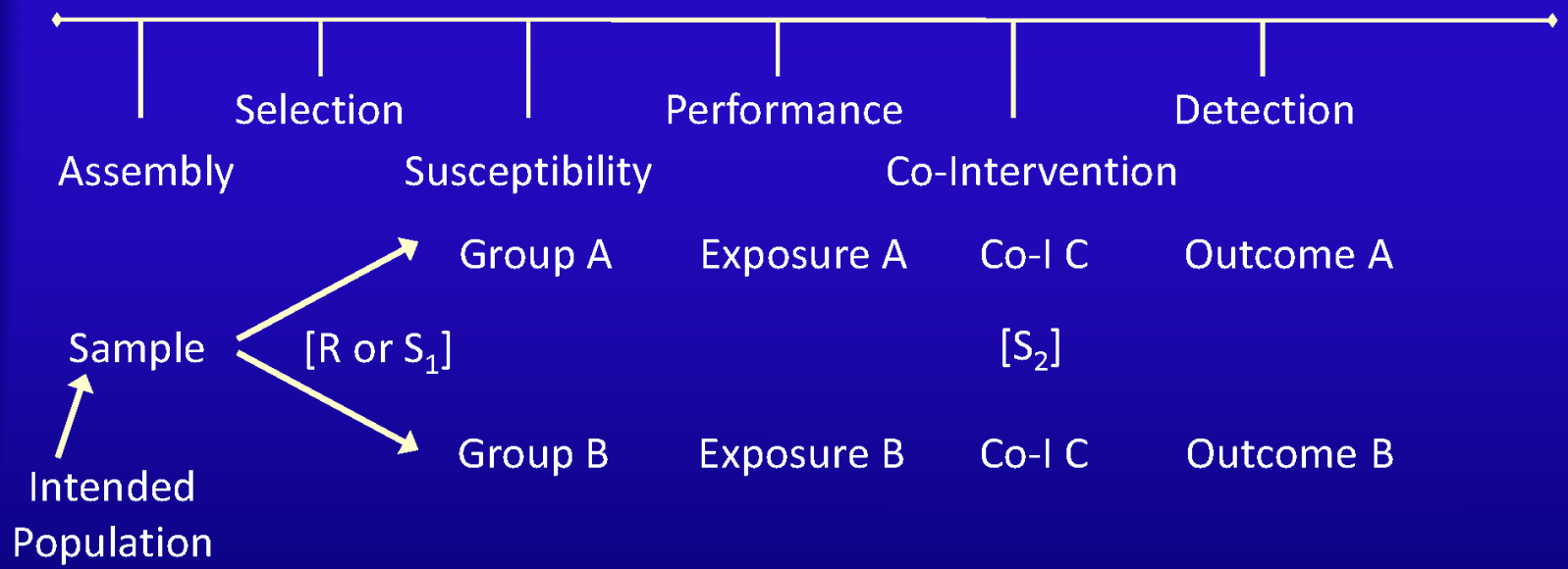
7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation



- Additional selection opportunities – co-interventions (beyond exposure of interest) may influence likelihood of outcomes.

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation

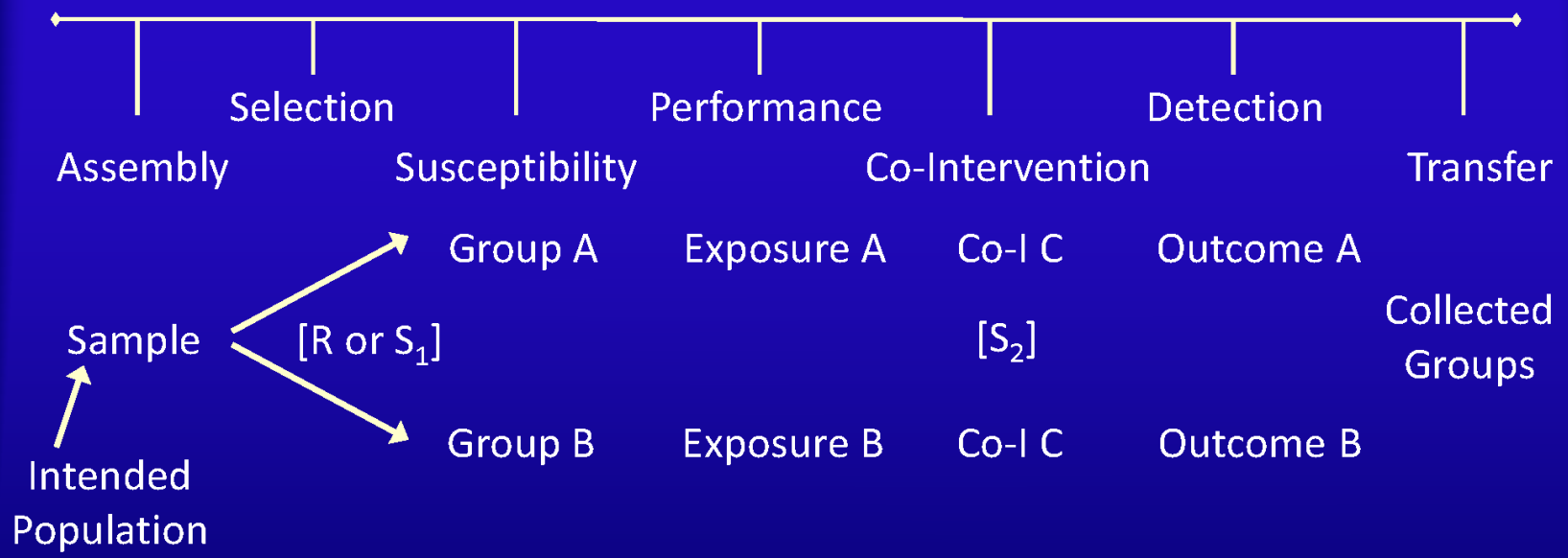


- Is process for determining outcomes applied unequally? Differences in surveillance, diagnostic testing, or interpretation?

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

8

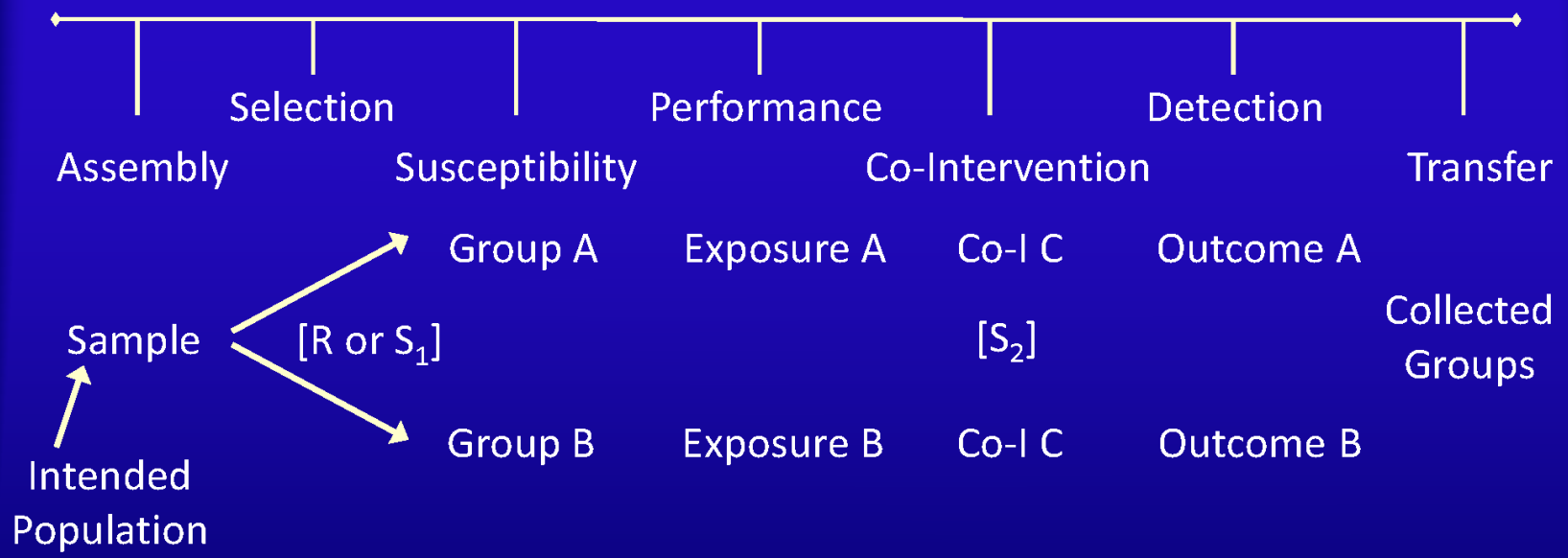
7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation



- Comparison of members of original cohorts of A and B – dropouts, in-study exclusions, crossovers, dealing with missing data…

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

7 Key Aspects of Research Architecture*:
Making Fair Judgments about Causation



- Goal: Comparability of groups who did and did not receive the exposure (except for the actual receipt of the exposure)

*Adapted by Neal Dawson from Alvan Feinstein's intellectual model (5 key aspects)

Section 4

A Published Example of My Early Propensity Score Work

Ahmed et al. 2012

Outcomes in younger and older adults with chronic advanced systolic heart failure: A propensity-matched study

Mustafa I. Ahmed [a], Marjan Mujib [a], Ravi V. Desai [a], Margaret A. Feller [a], Casey Daniel [a], Inmaculada B. Aban [a], Thomas E. Love [b], Prakash Deedwania [c], Bertram Pitt [d], Wilbert S. Aronow [e], Ali Ahmed [a,f,*]

[a] *University of Alabama at Birmingham, Birmingham, AL, USA*
[b] *Case Western Reserve University, Cleveland, OH, USA*
[c] *University of California-San Francisco, Fresno, CA, USA*
[d] *University of Michigan, Ann Arbor, MI, USA*
[e] *New York Medical College, Valhalla, NY, USA*
[f] *VA Medical Center, Birmingham, AL, USA*

- doi:10.1016/j.ijcard.2010.09.006
- M.I. Ahmed et al. *International Journal of Cardiology* 154 (2012) 128–133.

Abstract of Ahmed et al.

A B S T R A C T

Background: Older age is an independent predictor of all-cause mortality in patients with mild to moderate heart failure (HF). Whether older age is also an independent predictor of mortality in patients with more advanced HF is unknown.

Methods: Of the 2707 Beta-Blocker Evaluation of Survival Trial (BEST) participants with ambulatory chronic HF (New York Heart Association class III/IV and left ventricular ejection fraction <35%), 1091 were elderly (≥65 years). Propensity scores for older age, estimated for each of the 2707 patients, were used to assemble a cohort of 603 pairs of younger and older patients, balanced on 66 baseline characteristics.

Results: All-cause mortality occurred in 33% and 36% of younger and older matched patients respectively during 4 years of follow-up (hazard ratio {HR} associated with age ≥65 years, 1.05; 95% confidence interval {CI}, 0.87–1.27; $P = 0.614$). HF hospitalization occurred in 38% and 40% of younger and older matched patients respectively (HR, 1.01; 95% CI, 0.84–1.21; $P = 0.951$). Among 603 pairs of unmatched and unbalanced patients, all-cause mortality occurred in 28% and 36% of younger and older patients respectively (HR, 1.34; 95% CI, 1.10–1.64; $P = 0.004$) and HF hospitalization occurred in 34% and 40% of younger and older unmatched patients respectively (HR, 1.24; 95% CI, 1.03–1.50; $P = 0.024$).

Conclusion: Significant bivariate associations suggest that older age is a useful marker of poor outcomes in patients with advanced chronic systolic HF. However, lack of significant independent associations suggests that older age per se has no intrinsic effect on outcomes in these patients.

From the Introduction

> *The majority of heart failure (HF) patients are 65 years, and most deaths and HF-related hospitalizations in HF patients occur in this patient group. We have previously demonstrated that in a propensity-matched cohort of ambulatory patients with mild to moderate chronic HF, older age (≥ 65 years) was associated with increased mortality but not hospitalization.*

The objective of the current study was to examine the independent effect of older age on outcomes in chronic advanced systolic HF patients using a propensity-matched design.

Data Source and Subjects

> *This study was conducted using retrospective analysis of public-use copies of the Beta-Blocker Evaluation of Survival Trial (BEST) datasets obtained from the National Heart, Lung, and Blood Institute (NHLBI).*
>
> *BEST was a multicenter randomized controlled trial of the beta-blocker bucindolol in chronic systolic HF.*

Patients with advanced systolic HF were enrolled from 90 different sites across the United States and Canada. All patients had New York Heart Association class III or IV symptoms and a left ventricular ejection fraction below 35%.

Exposure and Outcomes

Exposure: We categorized (2707) patients into two age groups: younger (< 65 years) and older (≥ 65 years; n $= 1091$ or 40%).

Primary outcomes were all-cause mortality and HF hospitalization. Secondary outcomes included cardiovascular mortality, HF mortality and all-cause hospitalization. All outcomes were centrally adjudicated.

Greedy Matching Approach

> *Using a greedy matching protocol described elsewhere in detail, we were able to match 603 of the 1091 older patients with 603 patients* < 65 *years old who had similar propensity scores.*

We began by using a non-parsimonious multivariable logistic regression model to estimate propensity score for age ≥ 65 years for each of the 2707 participants. In the model, an age ≥ 65 years was used as the dependent variable and all clinically relevant baseline characteristics (see Fig. 1, in two slides) were included as covariates.

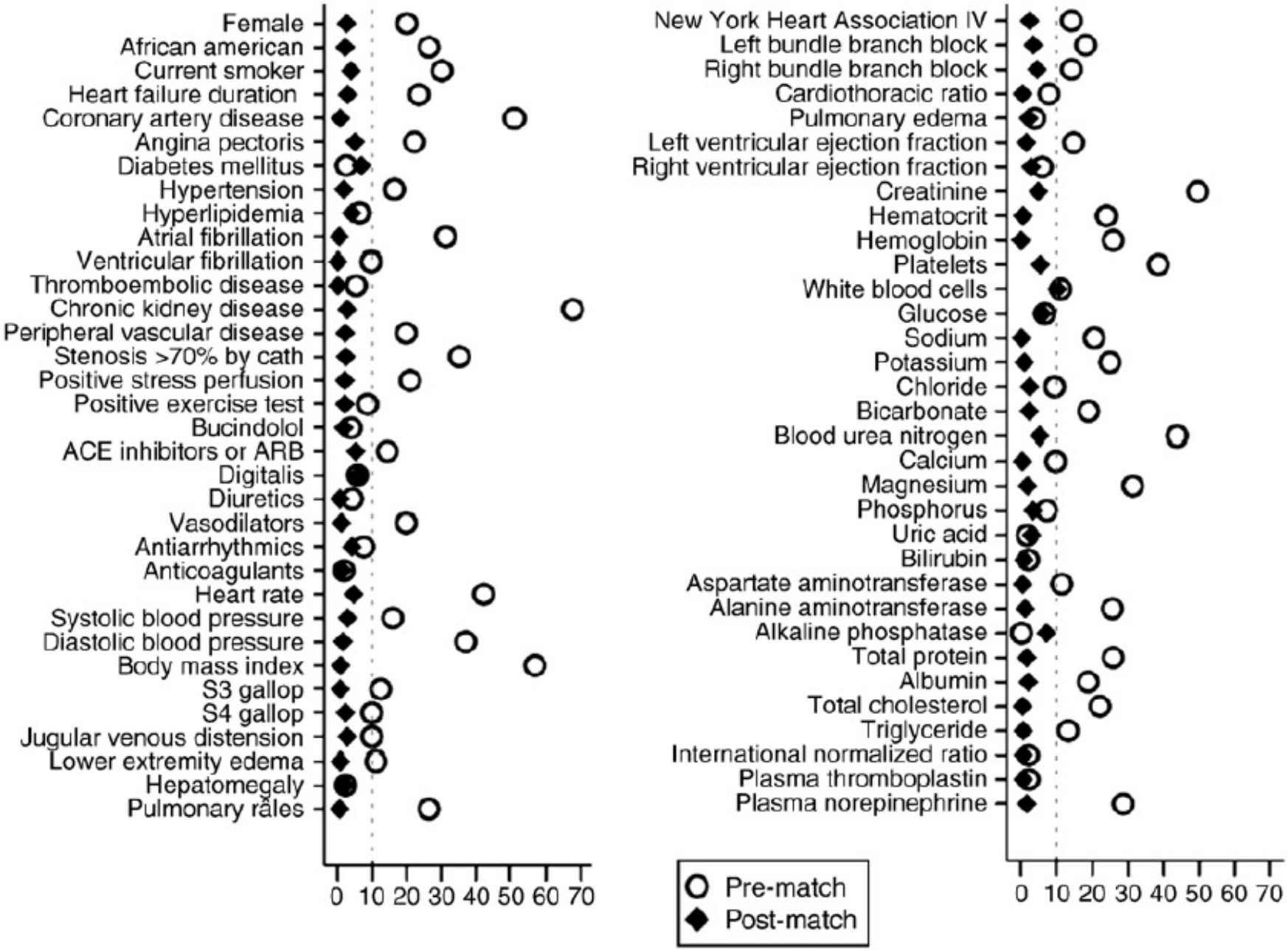
- The matching process used an SPSS macro published in the second edition of Levesque R, ed. *A Guide for SPSS and SAS Users* 2005.

Propensity Matching Results

Before matching, older patients were more likely to have coronary artery disease, atrial fibrillation and chronic kidney disease than younger patients.

After matching, absolute standardized differences between age groups were $< 10\%$ for all measured covariates (with the exception of white blood cell count which was 10.2%) with most values $< 5\%$ demonstrating substantial covariate balance across the groups (Love plot, next slide.)

Love Plot (Figure 1)

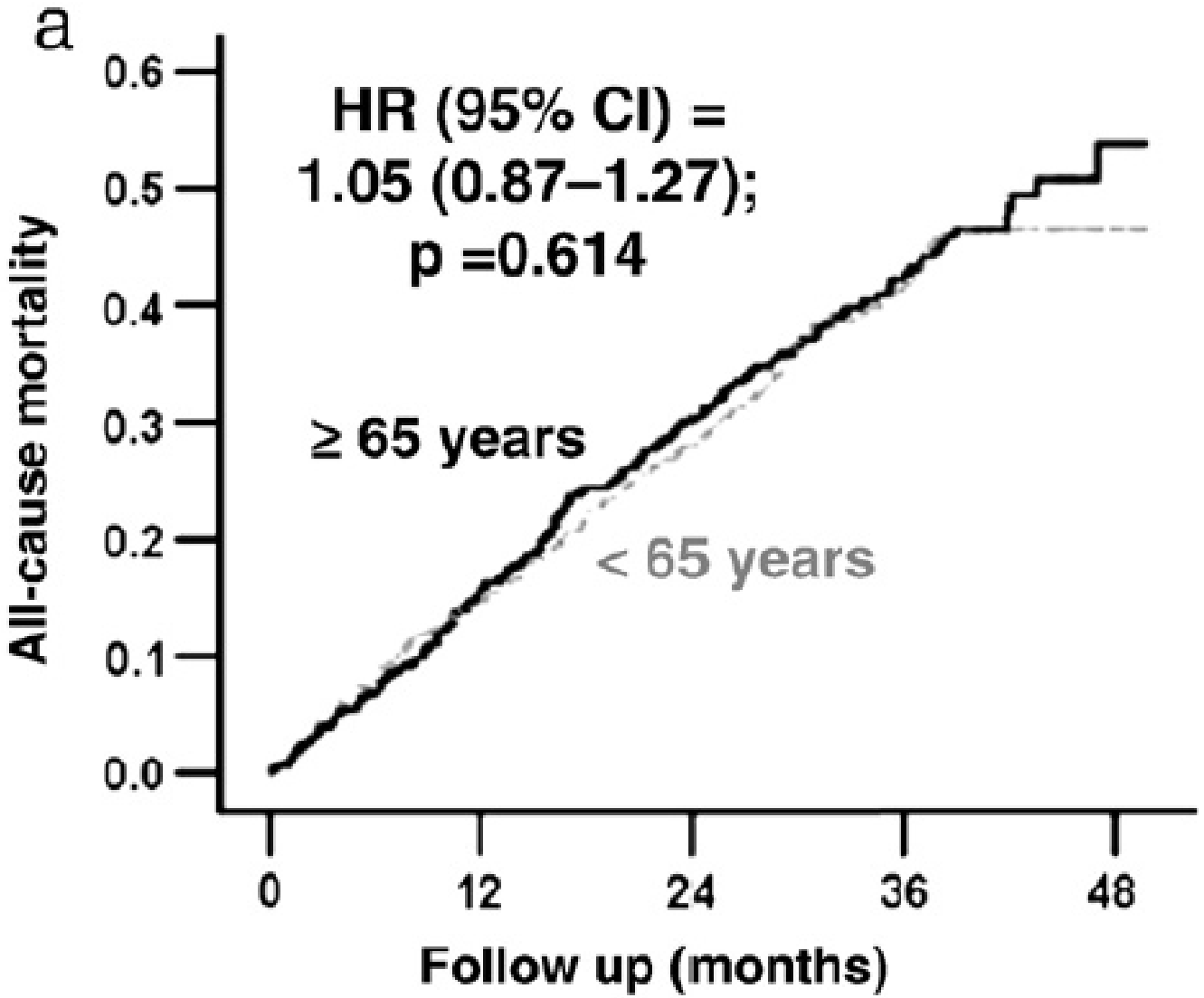


All-Cause Mortality Effect?

All-cause mortality occurred in 33% (202/603) and 36% (215/603) of matched younger and older patients respectively during 4 years of follow-up.

The hazard ratio (HR) for all-cause mortality when older patients are compared to younger patients was HR = 1.05, 95% CI: 0.87 - 1.27.

Kaplan-Meier Plot for All-Cause Mortality



Number of patients at risk

<65 years	603	460	289	117	16

Table 2

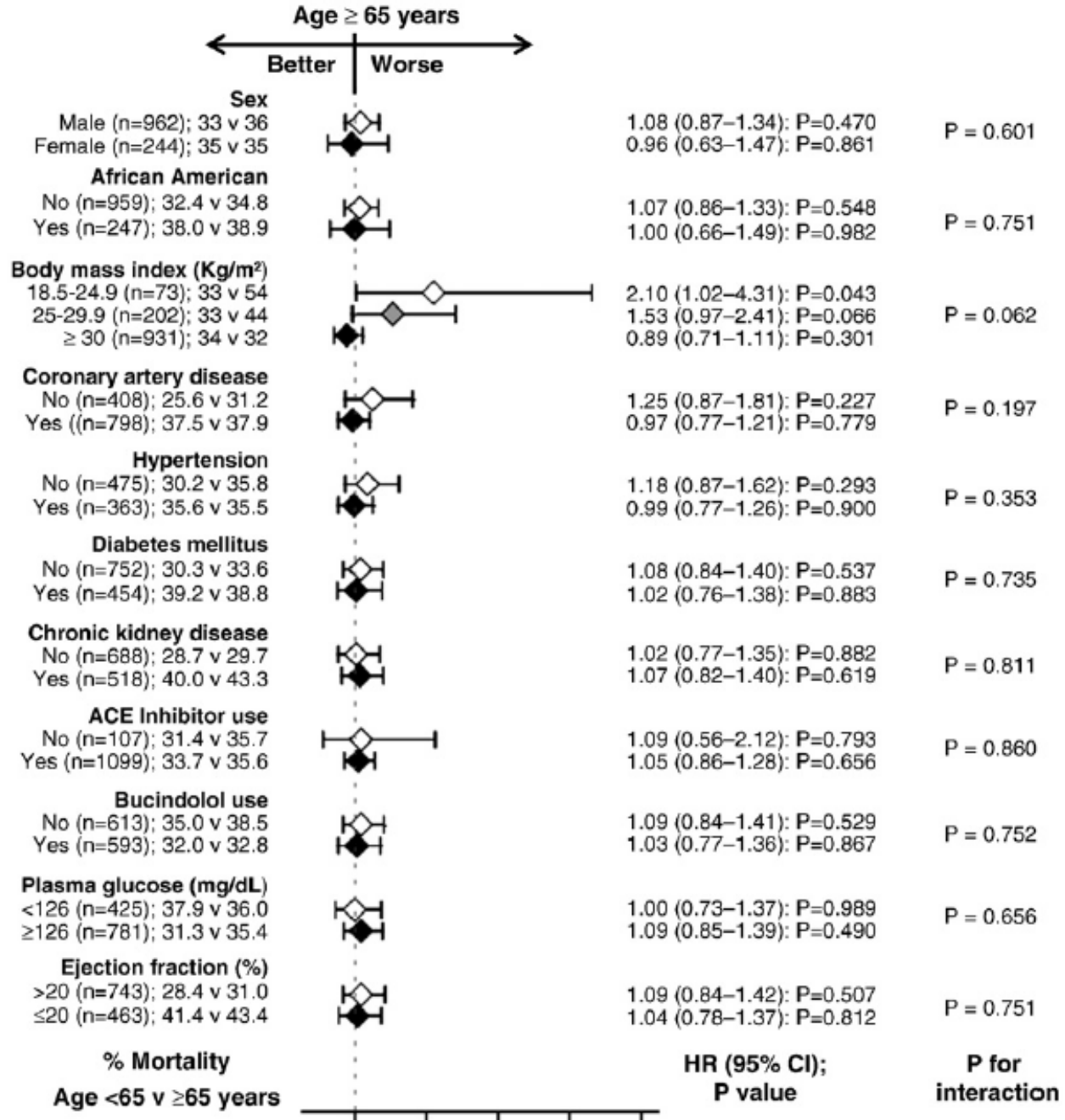
Table 2
Age ≥ 65 years and outcomes in the matched cohort.

Outcomes	Events (%)		Absolute risk difference[a] (%)	Hazard ratio (95% confidence interval)	*P* value
	<65 years ($n = 603$)	≥65 years ($n = 603$)			
All-cause mortality	202 (33%)	215 (36%)	+2	1.05 (0.87–1.27)	0.614
Cardiovascular mortality	182 (30%)	168 (28%)	−2	0.91 (0.74–1.12)	0.379
Heart failure mortality	52 (9%)	80 (13%)	+5	1.51 (1.07–2.14)	0.020
All-cause hospitalization	375 (62%)	409 (68%)	+6	1.07 (0.93–1.23)	0.372
Heart failure hospitalization	229 (38%)	240 (40%)	+2	1.01 (0.84–1.21)	0.951

[a] Absolute risk differences were calculated by subtracting percent events in the <65 year group from the percent events in the ≥65 year group (before rounding).

- Significant association of older age with increased risk of HF mortality.

Figure 3 (Subgroup Analyses)



Conclusions

In conclusion, in patients with advanced chronic systolic HF, older age is an important marker of increased mortality and hospitalization, but has no intrinsic effect on outcome. Therapeutic decisions in older adults with advanced HF should not be biased on the basis of age alone.

Funding and Acknowledgements

Dr. Ahmed is supported by the National Institutes of Health through grants (R01-HL085561 and R01-HL097047) from the National Heart, Lung, and Blood Institute and a generous gift from Ms. Jean B. Morris of Birmingham, Alabama.

"The Beta-Blocker Evaluation of Survival Trial (BEST) study was conducted and supported by the NHLBI in collaboration with the BEST Investigators. This manuscript was prepared using a limited access dataset obtained by the NHLBI and does not necessarily reflect the opinions or views of the BEST Study or the NHLBI." The authors of this manuscript have certified that they comply with the Principles of Ethical Publishing in the *International Journal of Cardiology*.

Rosenbaum Chapter 5

1. What was the most important thing you learned from reading Chapter 5?
2. What was the muddiest, least clear thing that arose in your reading?

Next Time

- Designing Observational Studies (Rubin 2001 and Rubin 2004)
- Discussion of Project Proposals