Why Do We Need Multivariate Models To Evaluate Programs?

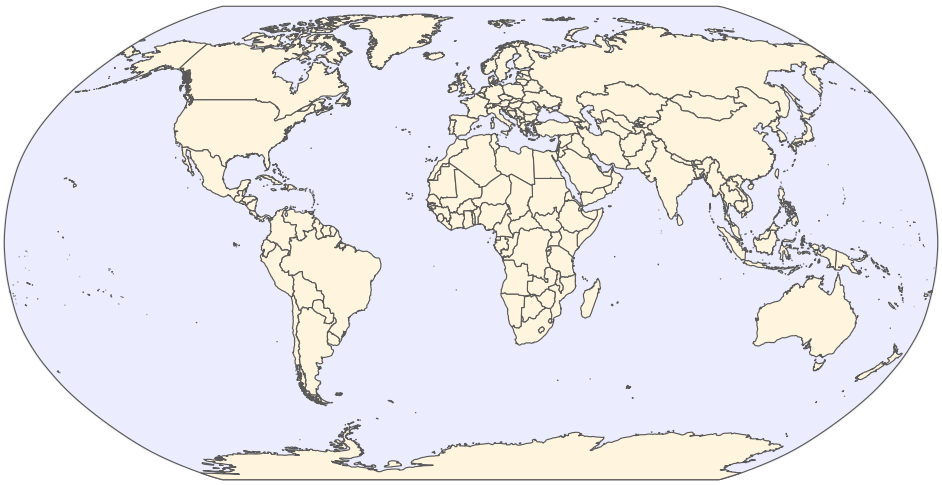
Andy Grogan-Kaylor

2026-02-11

Table of contents

1 Introduction

Across the world, there is a great deal of suffering. Many people have mental health issues or substance use issues. People often suffer the after effects of discrimination, poverty, trauma, violence or conflict.



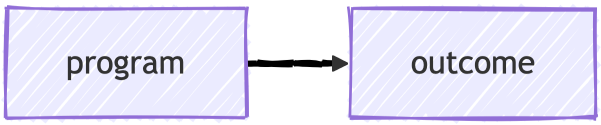
Understandably, many people and organizations try to develop *interventions* or *programs* for those who suffer from such difficulties.

Yet evaluating such *programs* may be more difficult than it appears.

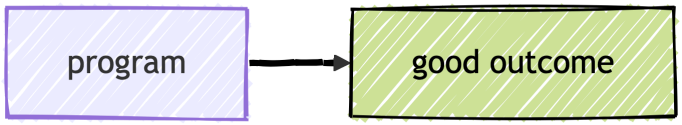
2 A Simple Evaluation

Let's consider a simple evaluation of a *program* designed to improve mental health.

In its simplest form, an evaluation might consist of looking at the outcomes–e.g. mental health outcomes–for those who participate in an *program*.



If the program appears to be associated with good outcomes, we might be tempted to claim that the program is successful.



3 Our Worry

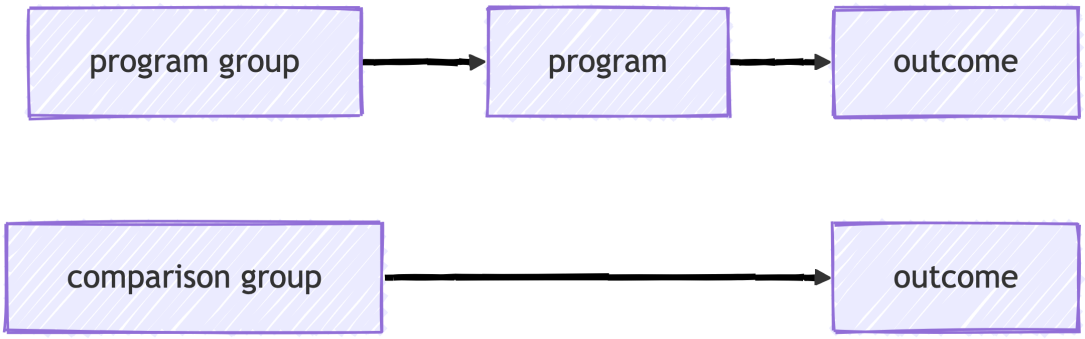
However, we might wonder, or worry, about a number of issues. For example:

- What were the outcomes like for this group of people *before* they participated in the *program*?
- If the outcomes of the *program* were favorable, it might not be that the program is particularly better, but that people improve or get better *naturally* over time.

If we fail to account for these possibilities, we are potentially declaring a program successful, when in fact it has no effect. We are potentially advocating that time energy and money be put into this program, when our resources would be better allocated elsewhere. Advocating for programs which have not been successfully evaluated, and which are not backed up by evidence, could be seen as an *ethical* issue.

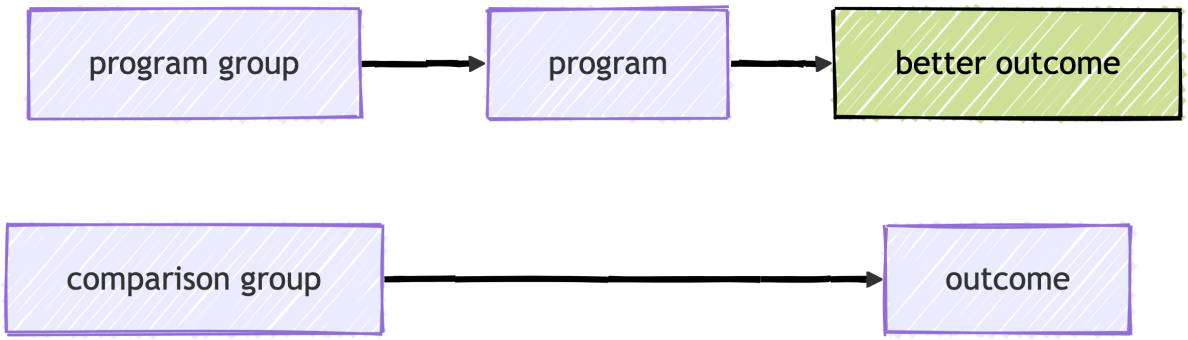
4 A More Sophisticated Evaluation

Therefore a more sophisticated research design would be to have one group of people–an *program group*–participate in the *program*, while another group–a *comparison* group–does not participate. We would then compare outcomes across the two groups.[1]



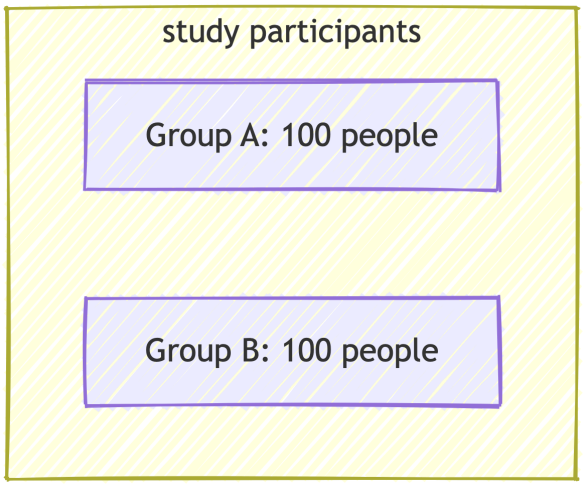
We hope that our results will show that those participants who have participated in the program group have better outcomes than those who were members of the comparison group.

[1]Valid questions could be raised about the *ethics* of such an approach, specifically denying participation in the *program* to one group of people. If a program is of *unknown* benefit, it is *ethical* to evaluate this program with a comparison group approach, because it is *not yet known* whether the program confers benefit on its participants, and whether the program represents a valid use of time, energy and financial resources, or whether the program is a waste of resources, and of participant's time. Indeed, an evaluation might uncover the fact that the program has no beneficial effects, or even that the program is harmful! Once a program has been established as *beneficial*, it would likely be *unethical* to conduct an evaluation where the program is withheld from some participants. However, we could then consider a comparison of the program with an enhanced version of the program that might confer even more benefits.



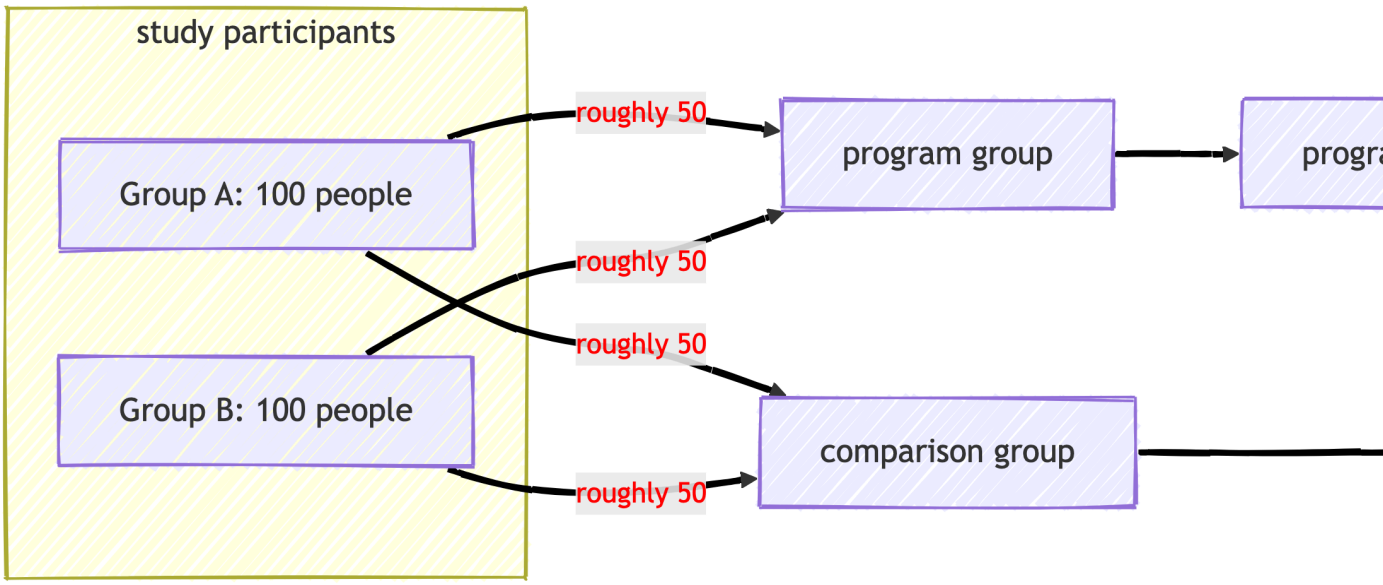
5 Different Subgroups of Individuals

One big question or concern in conducting an evaluation of this type is that our group of participants may be composed of different subgroups of individuals. These groups might be different in that they might repsresent individuals of different *racial*, *ethnic* or *gender* identities, might be people from different communities, or might have quite different sets of past experiences.



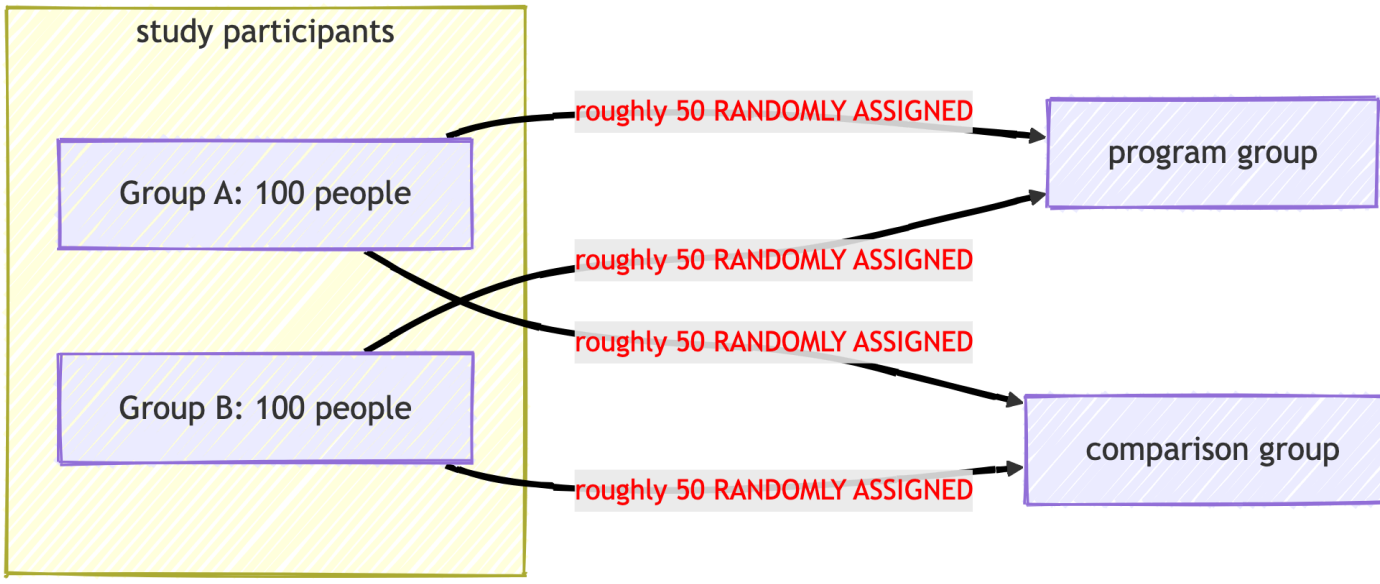
6 Are Those In the Program Group And Comparison Group Similar?

If participants from Group A and Group B are evenly distributed across the program and treatment groups, then we are not worried about the idea that an apparent effect of the program is because of unequal allocation of groups to the program.



7 Random Assignment

One way of accomplishing this *even distribution* would be by *randomly* assigning participants to the program.



8 Random Assignment Is Sometimes Not Possible

However, often the nature of the *program* is such that we want to allow *participants* in the study to *select their own level of participation*, or non-participation, in the program.

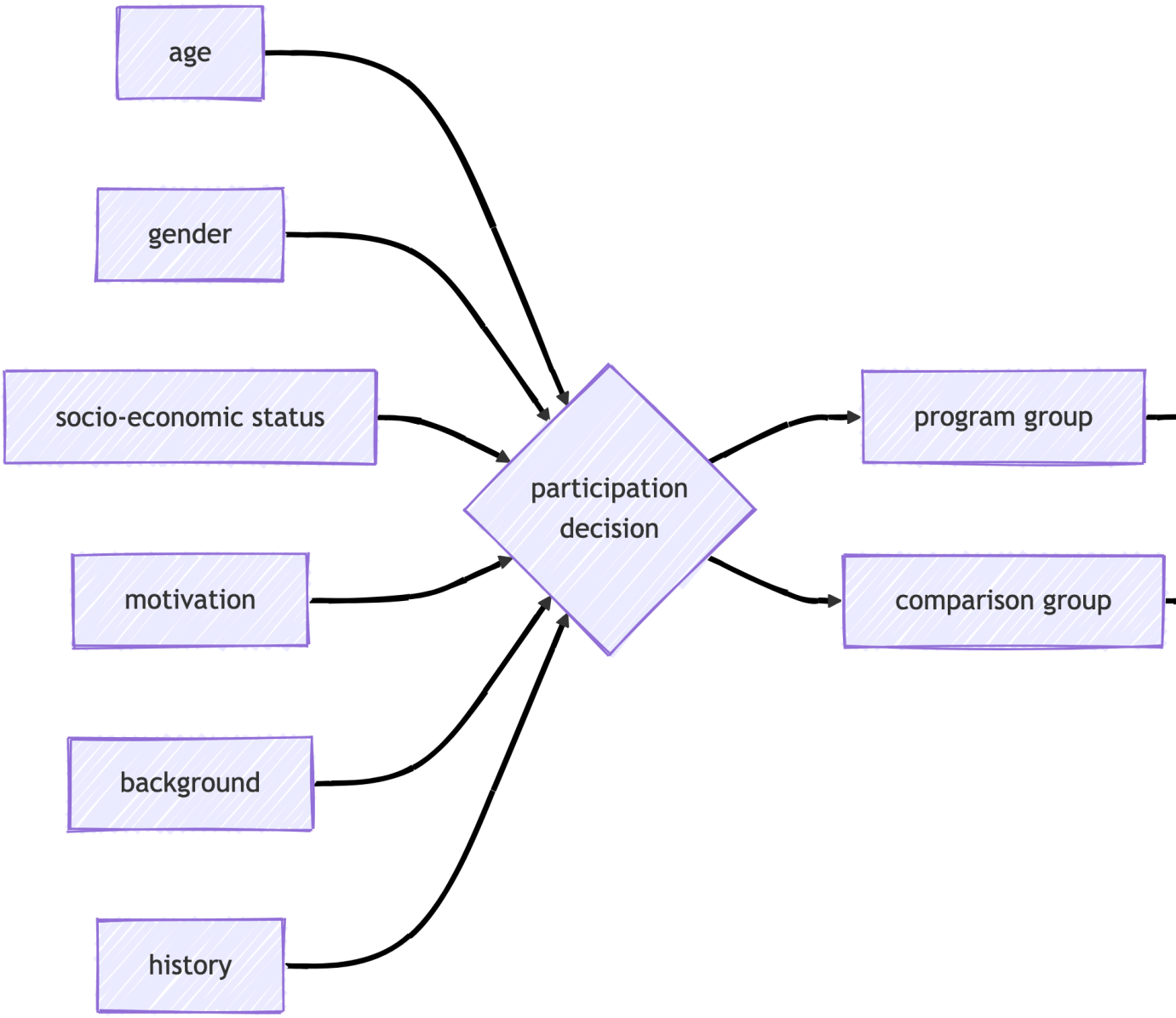
Often funders have objections to random assignment. Or the individuals or communities who are participating may have objections to random assignment.

Sometimes, instead of relying on random assignment, we may wish to observe the effects of a program more *naturalistically*.

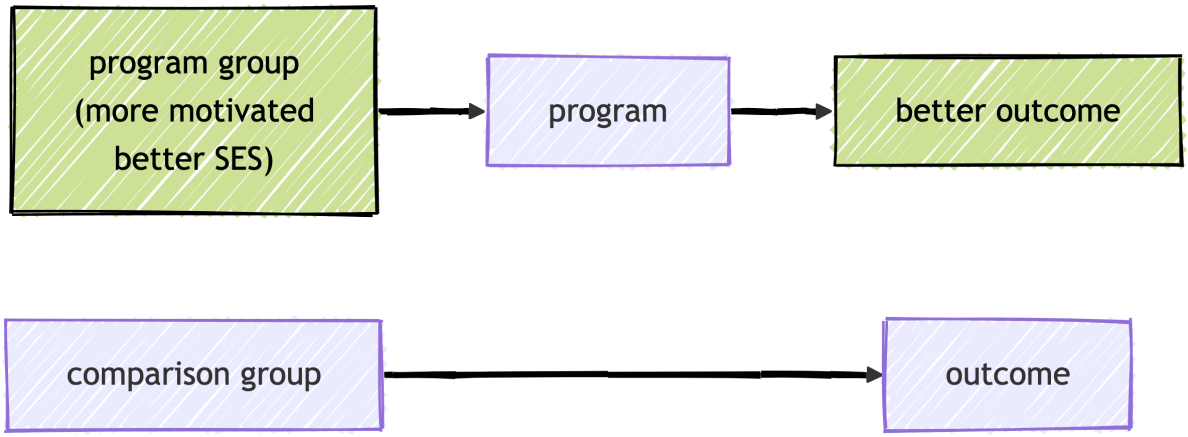
And lastly, the *logistical* demands of random assign may require *smaller samples*, when we wish to observe the outcomes of a program with a *larger*, and possibly more *generalizable*, sample.

9 When Participants Choose Their Own Level Of Particpation

If *participants* select their own participation, then it is likely that many background characteristics of participants–such as their age, gender, socio-economic resources, and motivation–will affect their decision to participate, and their level of participation.



If, for example, *more motivated individuals* and those with *more socio-economic resources* are the ones who choose to participate in the program, then any better outcomes might **not** be the result of the *program*, but a result of the *characteristics of individuals who chose to participate.*



10 What To Do?

When participation in a program is not randomly assigned, we need to think about a way to account for the various background and demographic factors that may be different between participants and non-participants.

10.1 A Visual Exploration

10.1.a A Simple Model of the Data

Let's start with a simple visual model of the data.

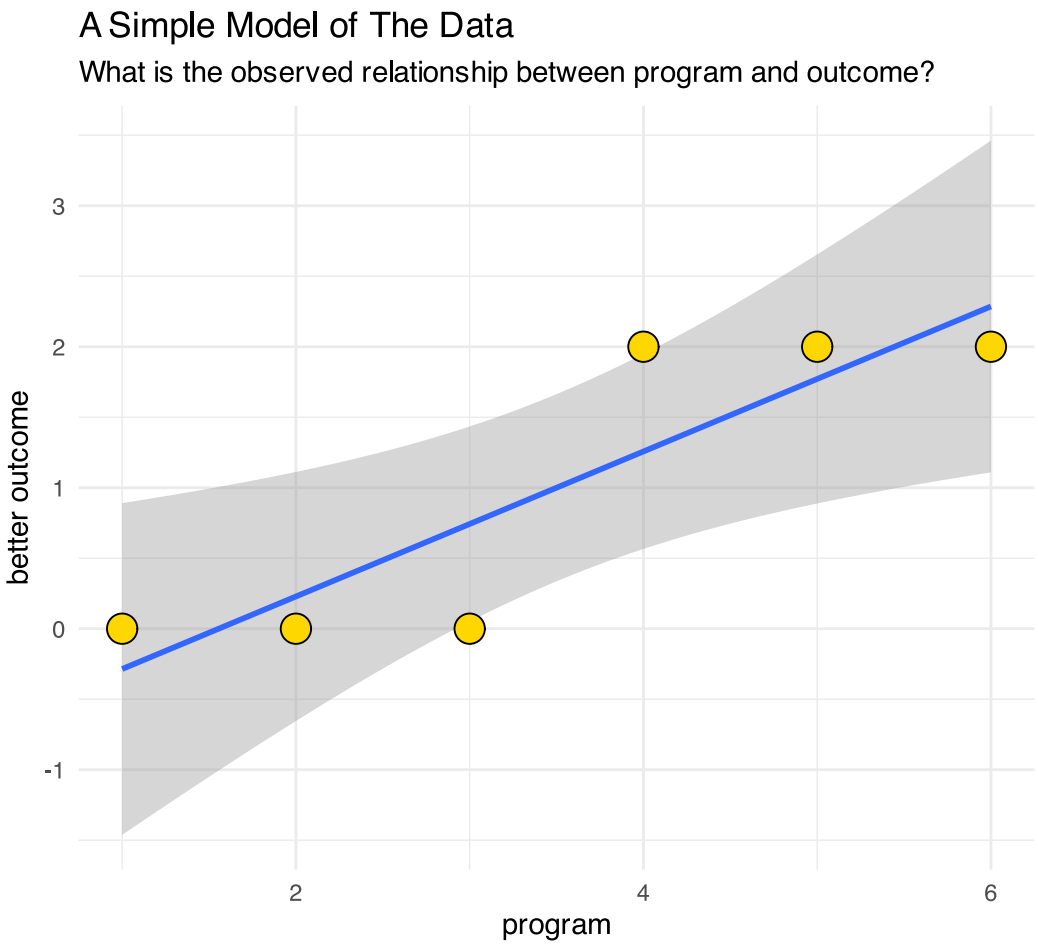


Figure 1

> **Key Takeaway**
>
> This simple model of the data (Figure 1) appears to indicate that the more participation in the program is associated with more of the good outcome.

Here we are considering two aspects, or *dimensions* of the data: participation in the program; and level of the good outcome.

10.1.b A More Complicated Model of the Data

However, our understanding becomes more complicated if we consider more dimensions of our study (Figure 2). Let's consider participation in the *program*, the level of the *outcome*, but also the fact that there are different *groups* of people in the study. For example, there might be a *more highly motivated* group, and a *less motivated* group.

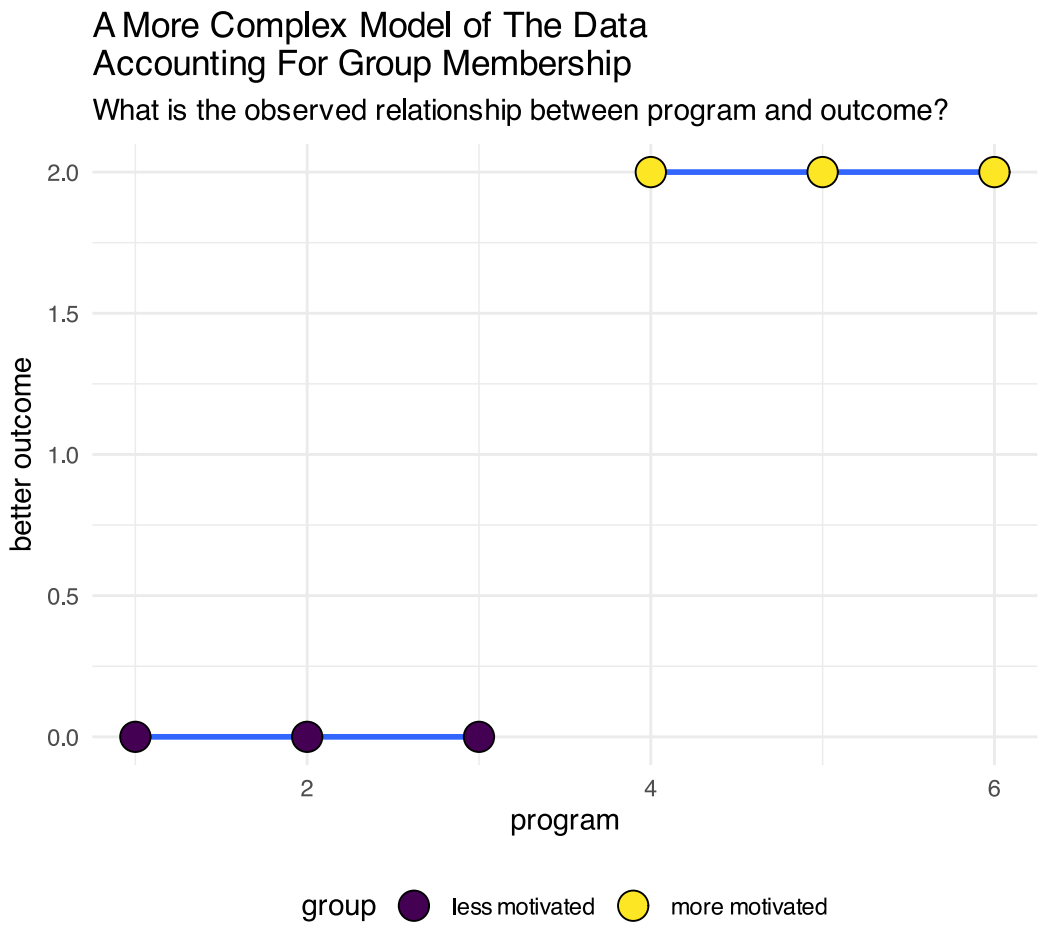


Figure 2

We see that when we consider a third dimension of the data, *motivation*, that our analysis changes. *More motivated* people are more likely to participate in the *program*, but among both *more motivated people* and *less motivated people*, participation in the *program* is associated with *no change* in the good outcome.

> **Key Takeaway**
>
> In essence, after considering people's pre-existing levels of *motivation*, the program has *no effect*!

Since we are considering three *dimensions* of the data, it may be illustrative to provide a 3 dimensional graph. This three dimensional graph illustrates the same idea as Figure 2: more participation in the program initially appears to be associated with better outcomes; after consideration of the third factor of motivation, the program appears to have no effect!

10.2 A Statistical Exploration

Accounting for, *ruling out*, or *controlling for* multiple background factors is one of the major reasons we use statistics.