# HarvardX Study Supporter Pre-Analysis Plan

Note: this project has involved collecting several rounds of data because of the relatively small sample sizes we have managed to recruit. The below was first written on 10/23/2014 at 10:33 am. After several rounds of editing this, I emailed it to Sam Skowronek with the following email with only “Live” in the body of the email. Though I intended to publish this on [www.socialscienceregistry.org](http://www.socialscienceregistry.org), I somehow forgot to.

We are about to analyze the data associated with our second round of recruitment. I am now posting the exact same pre-analysis plan I emailed at that time.



Also note that when we initially developed this preanalysis plan it was not clear in which other HarvardX courses we would be permitted to conduct this study. At the time of posting this pre-analysis plan (10/22/2015) we are planning to analyze a different course named Science and CookingX. For this and all subsequent HarvardX courses associated with this pre-analysis plan, our primary dependent variables will be whether the student earned a certificate as defined for each course by HarvardX.

Also, when analyzing the binary outcome measure (certificate-earning) we will use logit regression unless the logit analysis rejects specific strata due to homogeneity in the outcome variable within the strata (e.g., perfect prediction). In this scenario we will use OLS regression.

# Basic Design:

Students are recruited to participate in the study and asked to name two friends or family members with whom they communicate regularly—one of whom they do not live with (Non-Cohabitant) and one they do live with (if applicable) (Cohabitant). Students are told that one of their “Supporters” may be contacted throughout the course. The Supporters will be asked to inquire about the course when they communicate with the student and to encourage the student. Students are randomly assigned to one of three conditions:

* Non-Cohabitant Supporter condition
* Cohabitant Supporter condition
* Control condition

# Primary Hypotheses:

Our primary hypotheses depend on us being able to successfully communicate with Supporters. For this reason we will look separately at students based on the quality and extent of the Supporter contact information they provided. Below is the rank order of the information that students provided that we predict to be most treatment responsive. We will examine students who, for at least one Supporter:

* 1. Provided a valid SMS number for the targeted supporter (they may or may not also have included an email address)
	2. Only provided valid email address for the targeted supporter

### H1a: Students in either Supporter condition will be more likely to “pass” the course compared to those in Control condition

We will define “pass” as students who have earned a course certificate. Students earn a course certificate by:

* Tangible Things (USW30x) – completing 70% of the self-assessment assignments. At the end of every unit, there is a multiple choice section where students mark whether they completed the videos, readings, exercises and discussion participation. Everything in the course is honor based.
* Leaders of Learning (GSE2x) – Receiving a score of 60% or higher. Students receive grades on the quizzes that appear in each section (50% of score), as well as the self-reflection (25% of score) and peer assessment questions (25% of score). Certificates do not indicate the student’s score, only that the student passed.

We will use a logit regression controlling for the following when available. If coverage on these variables is poor, we may either exclude it or use some other strategy for imputing the variable. The variables are: student’s gender, age, and whether the student has a friend currently enrolled in the same course, and the Study Supporter’s education level.

### H1b: Students in the Non-Cohabitant Supporter condition will be more likely to “pass” the course compared to those in Control condition

We will define “pass” as students who have earned a course certificate. Students earn a course certificate by:

* Tangible Things (USW30x) – completing 70% of the self-assessment assignments. At the end of every unit, there is a multiple choice section where students mark whether they completed the videos, readings, exercises and discussion participation. Everything in the course is honor based.
* Leaders of Learning (GSE2x) – Receiving a score of 60% or higher. Students receive grades on the quizzes that appear in each section (50% of score), as well as the self-reflection (25% of score) and peer assessment questions (25% of score). Certificates do not indicate the student’s score, only that the student passed.

We will use a logit regression controlling for the following when available. If coverage on these variables is poor, we may either exclude it or use some other strategy for imputing the variable. The variables are: student’s gender, age, and whether the student has a friend currently enrolled in the same course, and the Study Supporter’s education level.

### H2: Those in the Non-Cohabitant Supporter condition will show a larger effect of being in the Supporter condition (compared to the control condition) on “passing” the course than those who were in the Cohabitant Supporter condition.

We will define “pass” as students who have earned a course certificate. Students earn a course certificate by:

* Tangible Things (USW30x) – completing 70% of the self-assessment assignments. At the end of every unit, there is a multiple choice section where students mark whether they completed the videos, readings, exercises and discussion participation. Everything in the course is honor based.
* Leaders of Learning (GSE2x) – Receiving a score of 60% or higher. Students receive grades on the quizzes that appear in each section (50% of score), as well as the self-reflection (25% of score) and peer assessment questions (25% of score). Certificates do not indicate the student’s score, only that the student passed.

We will use a logit regression controlling for the following when available. If coverage on these variables is poor, we may either exclude it or use some other strategy for imputing the variable. The variables are: student’s gender, age, and whether the student has a friend currently enrolled in the same course, and the Study Supporter’s education level.

This will be tested using an interaction term 2 (Live together: yes, no) X 2 (Condition: Supporter, Control).

# Exclusion Criteria:

We will only include in analyses students who meet the following criteria:

* We can match data from their opt-in survey with course performance data
* Entered at least one valid form of contact for their Supporters (valid email address, valid phone number).
* We have at least one valid form of contact for the student (valid email address, valid phone number).