Survey methodology

I’ve put together a write-up of what I’ll be doing for the survey and why. The Research Question, Sample Selection, and Analysis sections explain my reasoning for design choices I propose we make. The Survey Design section provides documentation for how we will do the survey, to cut down on confusion once we are administering it.

# Research Question

### Hypotheses

Developing the hypotheses to test before we start tabulating results helps prevent spurious findings. When presented with data, people instinctively – and often subconsciously – organize those data into a narrative that fits their existing theories. Formulating beforehand the assumptions that we have forces us form questions without that temptation; additionally, it helps us avoid reading too much into any “noise” in our results. An excellent, and comic, explanation can be found [here](https://xkcd.com/882/).

I’ve come up with the following hypotheses:

* More contact with clients is correlated with a better relationship.
* More contact with clients is correlated with clients understanding what is happening in their case.
* More contact with clients is correlated with clients feeling like their attorney is fighting for them.
* Clients feeling like their attorney is fighting for them is correlated with a better relationship.
* Clients feeling listened to is correlated with a better relationship.
* Clients feeling listened to is more strongly correlated with clients feeling like their attorney fought for them than is the outcome of their case.
* Clients feeling listened to is more strongly correlated with clients feeling like their attorney fought for them than is the number of times a lawyer contacted them.
* Clients getting referred to ancillary services is correlated with clients saying they received ancillary services.
* Clients being contacted by employees of NYCDS other than their attorney is correlated with a better relationship.
* Clients suffer from major collateral consequences, even when they are not found guilty, plea to a violation, plea to a minor misdemeanor, or are found guilty of a minor misdemeanor.

All this said, it is likely that we will not have enough of a response rate to be able to answer most of these questions. Please see the Response Rate section of this methodology for what we will do if we get a low response rate.

# Sample selection

### Intended variables

Our clients are a diverse group of people. Because of that, to get a representative sample we want to make sure that our sample is as diverse as our client population. These are intended variables: the aspects of clients’ demographics and cases that we want to make sure are represented in our sample. Potential intended variables include:

* Age
* Charge type
* Number of previous cases
* Incarceration status
* Outcome of case
* Use of services

### Sampling Population

The final sample will be about 600 people. We will sample among clients who have fully closed cases, which were disposed of in the last 3 years. Fully closed means that the client will have no expectation of interacting with their lawyer again; any repleader, conditional discharge, or ACD would have to be finished.

We will **not** sample anyone:

* without an address
* under 18
* ever 7/30’d
* currently on open case, as identified by OCA
* who does not speak English, Spanish, or French (potentially)
* whose case had more than one assigned attorney (potentially)

### Sampling method

I will perform a stratified random sample, which basically means that I will break the eligible population into groups (*strata*), and then will randomly sample from each group. Potential strata include:

* Trial
* Felony/misdemeanor
* Female
* Incarcerated – Rikers/upstate/non

### Oversampling

Oversampling means that we include a larger number of a particular group of people in our final sample. Oversampling is done for two main reasons: for small groups, it allows for large enough numbers to make statistical analysis possible. For example, if 5% of our clients are women (a liberal estimate), then a representative sample of 600 people would include only 30 women. With such a small group, even small differences of opinion look much larger. I plan to oversample female clients and clients whose cases went to trial.

Oversampling also allows us to correct for expected differences in response rates, so that our responses reflect the actual client population, outlined above. For example, if we expect 90% of incarcerated clients to respond, but only 70% of non-incarcerated clients to respond, we’d want to sample more non-incarcerated clients. I plan to oversample non-incarcerated clients for this reason.

# Survey Design

The following section is mainly administrative detail, and mostly to think through the logistics of the survey.

### Preparing for the survey

The final sample list will need to include:

* Client name
* Client contact information
* Attorney name
* Top charge
* Number of appearances
* Final disposition
* Incarceration status
* Gender
* Language
* Survey ID

Client name and contact information will be used only for administering the survey and generating letters, envelopes and/or surveys. Attorney name will be used only for generating the letter.

### Pre-notification

Previous studies have shown that pre-notification letters with information about the survey improve survey response rates. Our pre-notification letter would need to include a reminder of what the person was on trial for, that they were represented by NYCDS, and when their case was. I propose to send a pre-notification letter to participants 2 weeks before we mail the survey. This will also let us see how many addresses are no longer correct.

### Survey mailing

The survey packet will need to include a cover letter, which will be basically the same as the pre-notification letter; the survey itself; and a stamped return envelope.

### Processing responses

I propose that responses be processed in the following way:

* The preparer will redact any names or identifying information about the client or attorney
* The preparer will scan the survey. Any additional documents (a page of notes, etc) should be scanned to a separate file.
* The preparer will save the scanned image save it to a folder on the U Drive. The image name should be the ID of the survey.
* The preparer will check the quality of the image, and make sure that it is machine-readable.
* The preparer will check off the survey as completed on the list of surveys.
* Finally, the preparer will shred the paper copy of the survey.

# Analysis

### Response rate

Surveys take advantage of the fact that a randomly selected group of items will look roughly like the population they come from, especially as the number of selected items goes up. This allows us to make inferences about the population. However, if the sample becomes biased or otherwise non-random, then we can no longer reliably make inferences about the whole. Because of this fact, researchers have debated at length how high the response rate needs to be in order for the survey results to be valid.

In general, researchers agree that non-response – that is, how many people who got the survey but did not respond– does not affect the results **as long as** the non-response is random. But non-response is usually not random; in fact, respondents are often the people who feel the most strongly about an issue. Because low response rates can indicate that phenomenon, social scientists generally agree that a response rate of 60% is problematic, 70% is acceptable, and 80% is good. We will not do an analysis on results if the response rate is 60% or less. If the response rate is between 60% and 80%, we will analyze the results but do a thorough non-response bias check, described below.

### Post-survey reliability testing

A non-response bias check is simply a doing sort of mini-survey of the people who did not respond (but got the survey), using a different survey method as the original survey. That mini-survey would include a small subset of questions to see if the non-respondents differ in significant and systematic ways from the respondents. For our survey, a non-response bias check could be a telephone call to a random sample of non-respondents, asking the first few questions.

### Adjusting weights

To get population-wide estimates, we will need to adjust the weight given to survey responses, to account for any oversampling we did in the sample selection and any non-response bias that occurred.