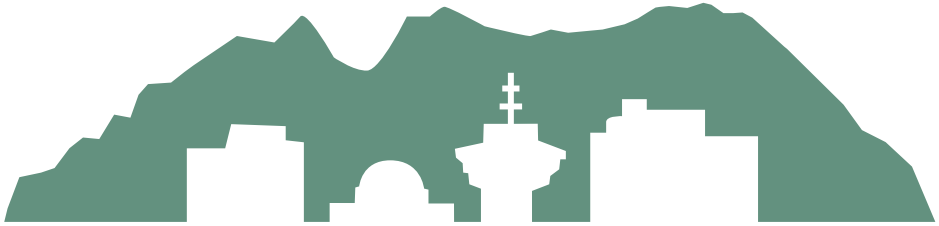


METROPOLIS BRITISH COLUMBIA

Centre of Excellence for Research on Immigration and Diversity

Experimental Methods in Evaluation

Krishna Pendakur and Phil Oreopoulous

Simon Fraser University

University of Toronto and Harvard University

What Are We Trying To Do?

- Evaluate the effect of a *treatment* on the person who is treated
  - *Treatment* could be getting settlement services or doing a language training program
- Try to get more people to accept treatment
  - Say we think the treatment is good: how do we get more people treated?

Evaluate Treatment Effect

- Treatment Effects
  - "treatment" is experimental language. We "treat" a plant with a pesticide and see what happens.
  - The treatment effect we are interested in is causal: the treatment causes the effect.
  - We'd like experimental data: we give treatment to people at random, and measure how they change.
    - But, in real life and in real programs, people choose whether or not to get treated.
    - Is the treatment effect different across people?

Get More People Treated

- If we believe the treatment is ‘good’ and cheap, then you’d want more people to do it.
- Sometimes, you can change little things to ‘nudge’ people in the right direction.
  - E.g., you have ‘opt-out’ RRSP savings rather than ‘opt-in’ RRSP savings.
- So, you use an experiment to try a nudge and figure out if it worked

Treatment Effects

- Suppose you have data on English proficiency for immigrants who've done language training and immigrants who haven't.

	Proficiency	Difference
Got training	Grade 11	
Didn't	Grade 12	-1 grade

- Is -1 grade the treatment effect?
  - That would suck, because we think language training should improve proficiency not worsen it.
  - People who got training might have had worse initial proficiency—proficient people don't need training.

Treatment Effects

- Suppose you have information on the average language proficiency of people before and after they take your English training

Before	After	Difference
Grade 9	Grade 11	2 grades

- Is 2 grades the treatment effect?
  - It's the average effect **on those who were treated**.
  - But, it may vary a lot across those people
  - And may be different for untreated people
  - Maybe those who received training would have improved more anyway without it

Treatment Effects: People Aren't Passive

- Suppose that underlying this effect for treated people is a potential amount of learning for different people

Potential	**Before**	**After**	**Difference**
Would take course	Grade 9	*Grade 11*	**2 grades**
Would not take course	*Grade 12*	Grade 13	1 grade

  - Difference in *italicized* numbers was the first pass.
  - **Bold** number was the second pass.
- These numbers give us everything we'd need to evaluate the program.
  Too bad we seldom get to observe both outcomes

Experimental Ideal: Random Assignment

- An experimental methodology could generate the numbers in that table.
  - It would *randomly assign* treatment (language training) to people (whether they want it or not), and measure the outcome (proficiency) for those treated (treatment group) in comparison to those not treated (control group).
  - Those not treated are like 'before'; those treated are like 'after'.
- So, experiments are awesome.

Random Assignment

- How do you isolate the treatment effect from other external factors?
  - People have lots of stuff going on, not just language training, e.g., differing family and community support; differing workplaces; etc.
- Random assignment solves this problem.
  - The odds of getting treated are the same no matter what external factors the person faces.
  - The measured treatment effect is the average across all that variation.

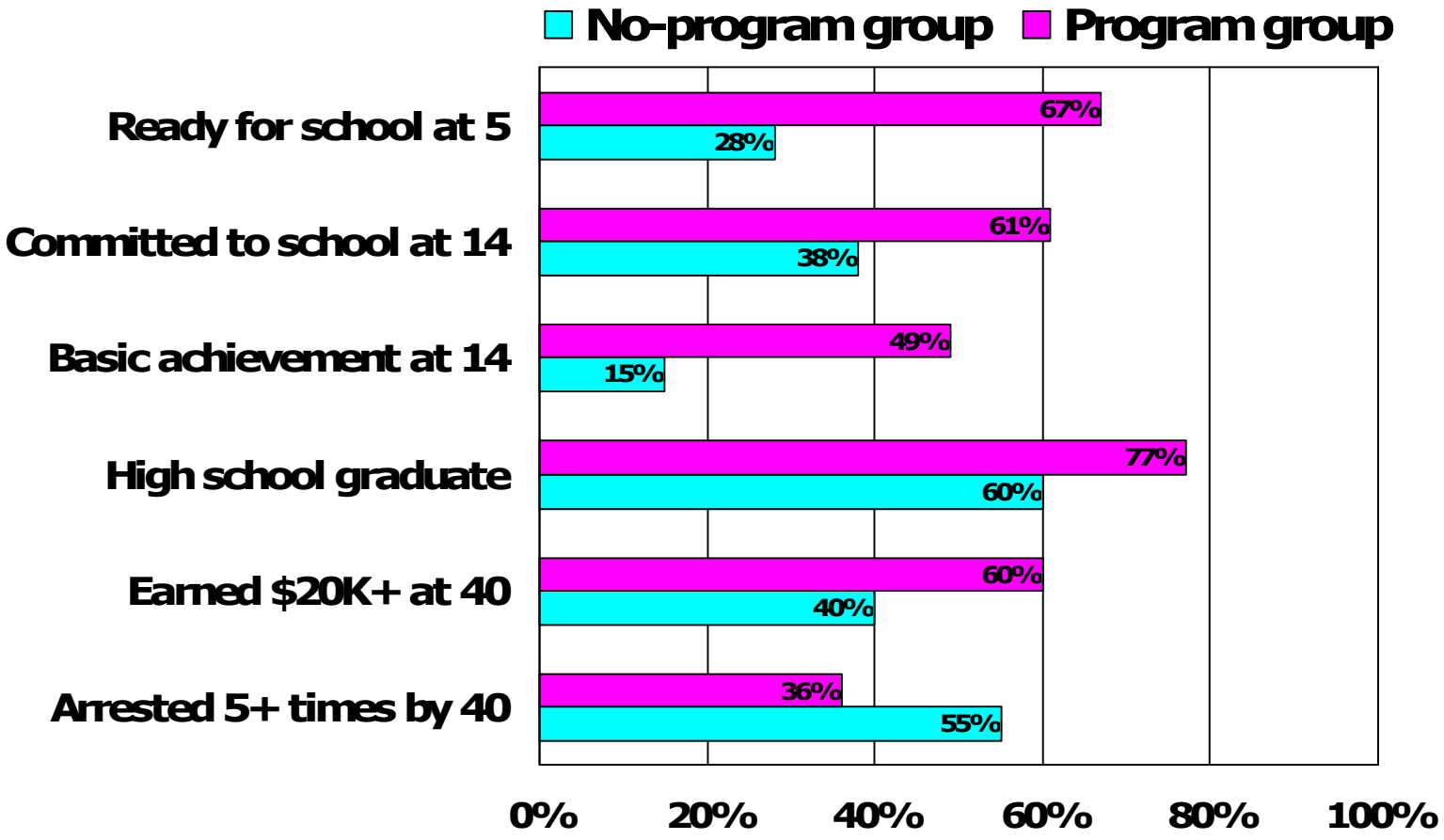
Random Assignment Opportunities

- Pilot Projects
- Line-ups and Waiting Lists
- Administration of services
- Low take-up in programs/services
- Introduction of new programs
- In the Lab

Example: Perry Preschool Project Randomize Program Participation

- From 1962 through 1967 sample of 123 low-income African-American children in Michigan were assessed to be at high risk of school failure
- randomly assigned: 58 of them to a program group that received a high-quality preschool program at ages 3 and 4
- 2.5 hours of classroom time each weekday and 90 minute home visit each week Monthly parent facilitated meeting as well.
- 65 to another group that received no preschool program.

Major Findings: High/Scope Perry Preschool Study by Age 40



Example: The Self-Sufficiency Project Eligibility Randomization

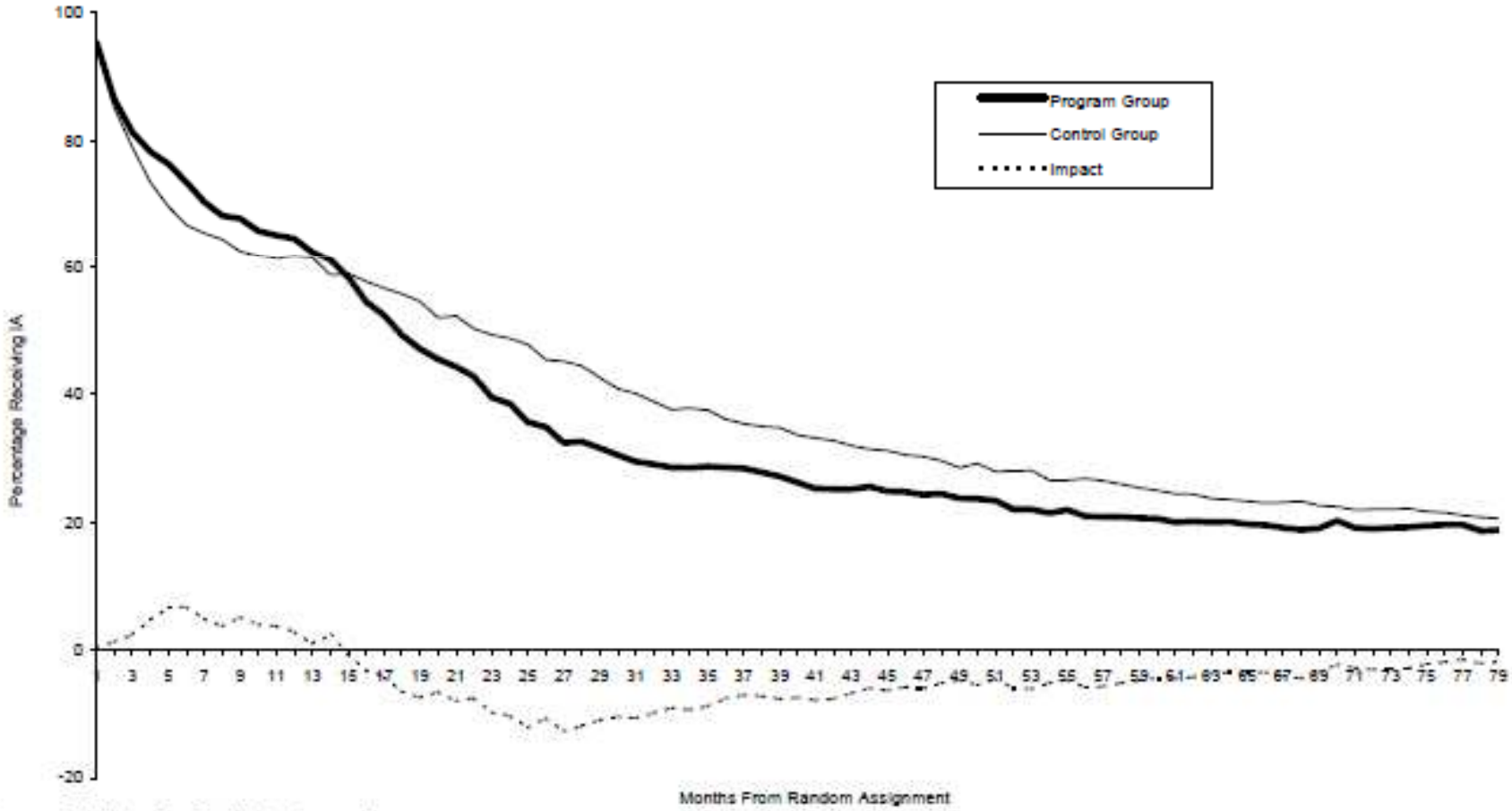
- In New Brunswick and BC, from 1992 to 1999, single parents entering welfare randomly told that if they remained on welfare for a year, they would become eligible for a temporary earnings supplement , paid on top of earnings from employment for up to three continuous years, as long as the person continued to work full time and remained off of IA. Very large payoff from work (50% the difference between $37,000 and actual salary).

- Would have to find fulltime work in first year of eligibility to remain eligible for remaining three years

- Note incentive to stay on welfare

Example 2: The Self-Sufficiency Project

Figure ES.1: Percentage Receiving Income Assistance, by Months From Random Assignment, in the SSP Applicant Study



Source: Calculations from IA administrative records.

Eligibility Randomization allows one to estimate mean effect from offering program and mean effect on those that participate

- Eligibility randomization allows one to directly estimate the mean effect of eligibility for the program on the population included in the experiment:

- the effect of eligibility for the program on outcomes is known as the **Intent to Treat (ITT) effect**.

- With some additional assumptions (the effect for those eligible who do not participate , one can also estimate is zero) the TOT from an eligibility randomization experiment with partial compliance: TOT = ITT/P(Participate |Eligible)

Example: Impact Study of Big Brothers Randomized Waiting List

- Big Brothers Big Sisters of America 100 year old mentoring program that pairs unrelated adult volunteers with youth from single-parent households: meet 2-4 times per month for at least one year

- Despite long existence, effects of mentoring program were not credibly documented

- 10-16 year olds who applied to participate in program offered to participate in study: 50% chance of being selected right away; 50% chance of being selected at least 18 months later (typical waiting period longer than 12 months). Only 2.7% rejected offer to participate.

Table 16 How Youth Benefit from Big Brothers Big Sisters Relative to Similar Non-Program Youth 18 Months After Applying

Outcome	Change
Antisocial Activities	
Initiating Drug Use	-45.8%
Initiating Alcohol Use	-27.4
Number of Times Hit Someone	-31.7
Academic Outcomes	
Grades	3.0%
Scholastic Competence	4.3
Skipped Class	-36.7
Skipped Day of School	-52.2
Family Relationships	
Summary Measure of Quality of the Parental Relationship	2.1%
Trust in the Parent	2.7
Lying to the Parent	-36.6
Peer Relationships	
Emotional Support	2.3%
Number of Youth	959

Example: Random nudges

- Benefit programs are often constructed as though mere existence ensures take-up

- Research in Psychology and economics demonstrates effectiveness of nudges (seemingly small-differences in signup procedures and marketing that lead to large differences in participation)

- Nudges allow one to evaluate direct impact and subsequent impacts

- Changing defaults (Beshears et al. 2006a)
- Simplifying options (Beshears et al. 2006b)
- Information (Hastings and Weinstein, 2008)
- Marketing (Bertrand et al. 2006)
- Incentives or disincentives (Vollp et al)

FAFSA Experiment
Random ‘Nudges’ to increase take-up

- Two treatments to parents and individuals going to H&R Bock after consenting to participate in experiment about college:

- 1. provide information about estimated costs and aid eligibility to H&R Block clients

- 2. provide information and assistance with filling out FAFSA

Table3. Summary of Results

	(1)	(2)	(3)
	Outcome during first year following experiment		
	Filed FAFSA (Based on DOE data)	**Attended College** (Based on NSC and OBR data)	**Attended College and Received Pell Grant** (Based on DOE data)
Dependent Participants (N=868)			
Control Group Mean	0.399	0.342	0.296
FAFSA Treatment Effect	0.157	0.081	0.106
	[0.035]***	[0.035]**	[0.034]***
Info Treatment Effect	-0.012	-0.004	0.004
	[0.060]	[0.058]	[0.056]
Independent Participants, No Prior College (N=9232)			
Control Group Mean	0.161	0.095	0.111
FAFSA Treatment Effect	0.267	0.015	0.029
	[0.009]***	[0.007]**	[0.007]***
Info Treatment Effect	-0.019	0.003	-0.016
	[0.014]	[0.012]	[0.012]

Example: Audit Studies

5354 Russell View Road, Mississauga, ON, L5M 5V8, (647) 477-2873, martin.john15@gmail.com

John Martin

Professional Summary

- Experienced in various business aspects; accounting, customer relations, computer training, sales, marketing, negotiations, presentations, and office operations.
- Highly motivated.
- Experienced multi-tasker.
- Analytical and detail-oriented, problem solver.
- Efficient with time management skills.
- Excellent skills in Visual Basic, HTML and Microsoft Office.

Experience

KPMG Corp. 2006 to date Toronto, Ontario

Financial Analyst

- Assisted the Corporate Finance Director with budgeting and forecasting exercises. Identified, explained and communicated variances for operating plans and latest forecast. Examined the feasibility of business projects and prepare a plan of action based on financial analysis. Reconciled monthly bank statements entries via AS400.

ZAC Marketing Inc 2004 - 2006 Toronto, Ontario

Actuarial Analyst

- Performed actuarial and statistics analysis of risk to provide the underwriting department with keys contract valuation metrics. Developed actuarial models used for pricing and/or risk management. Performed segmentation analysis on the behalf of insurance companies to determine best and worst performing products/classes and recommend strategies for growing/correcting those areas as appropriate. Examined expert risk reports on larges individual corporate risks. Improved the decision making process significantly and the quality of internal statistical and technical reporting documents by creating an Access based program that offered a wide range analyses of the company's portfolio of reinsurance contracts. Improved the average technical account reconciliation time by more than fifty percent.

FGF Brands Inc. 2002 - 2004 Toronto, Ontario

Investment Analyst

- Independently performed fundamental research on assigned securities (distribution sectors). Participated in the decision making process with respect to portfolio management by making buy, sell and hold recommendations. Analyzed and tracked key data and statistics related to individual stocks and portfolios. Created and maintained financial models for stock and portfolio analysis. Produced performance reports that include analyses of returns, risk, added value and portfolio characteristics. Created an Excel-based application that calculates more than 20 financial and operational ratios. Elected employee of the month four times.

Education

University of Waterloo 1998 - 2002 Waterloo, Ontario

- **Bachelor of Commerce, Accounting**

Additional Interests and Activities

World Traveller: Travelled to thirty-one countries on five continents.
Big Brothers: Mentor for disadvantaged youth.
Other Activities: Competitive squash player, classical piano player, recreational photographer.

5354 Russell View Road, Mississauga, ON, L5M 5V8, (647) 477-2873, zhang.long11@gmail.com

Zhang Long

Professional Summary

- Experienced in various business aspects; accounting, customer relations, computer training, sales, marketing, negotiations, presentations, and office operations.
- Highly motivated.
- Experienced multi-tasker.
- Analytical and detail-oriented, problem solver.
- Efficient with time management skills.
- Excellent skills in Visual Basic, HTML and Microsoft Office.

Experience

KPMG Corp. 2006 to date Toronto, Ontario

Financial Analyst

- Assisted the Corporate Finance Director with budgeting and forecasting exercises. Identified, explained and communicated variances for operating plans and latest forecast. Examined the feasibility of business projects and prepare a plan of action based on financial analysis. Reconciled monthly bank statements entries via AS400.

ZAC Marketing Inc 2004 - 2006 Toronto, Ontario

Actuarial Analyst

- Performed actuarial and statistics analysis of risk to provide the underwriting department with keys contract valuation metrics. Developed actuarial models used for pricing and/or risk management. Performed segmentation analysis on the behalf of insurance companies to determine best and worst performing products/classes and recommend strategies for growing/correcting those areas as appropriate. Examined expert risk reports on larges individual corporate risks. Improved the decision making process significantly and the quality of internal statistical and technical reporting documents by creating an Access based program that offered a wide range analyses of the company's portfolio of reinsurance contracts. Improved the average technical account reconciliation time by more than fifty percent.

FGF Brands Inc. 2002 - 2004 Toronto, Ontario

Investment Analyst

- Independently performed fundamental research on assigned securities (distribution sectors). Participated in the decision making process with respect to portfolio management by making buy, sell and hold recommendations. Analyzed and tracked key data and statistics related to individual stocks and portfolios. Created and maintained financial models for stock and portfolio analysis. Produced performance reports that include analyses of returns, risk, added value and portfolio characteristics. Created an Excel-based application that calculates more than 20 financial and operational ratios. Elected employee of the month four times.

Education

University of Waterloo 1998 - 2002 Waterloo, Ontario

- **Bachelor of Commerce, Accounting**

Additional Interests and Activities

World Traveller: Travelled to thirty-one countries on five continents.
Big Brothers: Mentor for disadvantaged youth.
Other Activities: Competitive squash player, classical piano player, recreational photographer.

Callback Rates by Resume Type
(Difference Compared to Type 0)
[Standard Error of Difference, * indicates sign. Diff. compared to prev. type]
{Callback Ratio: Type 0 / Type}

		Ethnic Origin					
		English-Canada	India	China	Pakistan	Britain	India/China/Pakistan
Type 0	**English Name** **Cdn Educ/Exp**	**0.158**					
Type 1	**Foreign Name** **Cdn Educ** **Cdn Exp**		**0.121** (-0.037) [0.019]* {1.31}	**0.108** (-0.050) [0.018]*** {1.46}	**0.11** (-0.048) [0.016]*** {1.44}	**NA**	**0.113** (-0.045) [0.011]*** {1.40}
Type 2	**Foreign Name** **Foreign Educ** **Cdn Exp**		**0.122** (-0.036) [0.022] {1.30}	**0.094** (-0.064) [0.020] {1.68}	**0.14** (-0.018) [0.027] {1.13}	**0.129** (-0.029) [0.019] {1.22}	**0.114** (-0.044) [0.014] {1.39}
Type 3	**Foreign Name** **Foreign Educ** **Mixed Exp**		**0.075** (-0.083) [0.019]*** {2.11}	**0.103** (-0.055) [0.021] {1.53}	**0.078** (-0.080) [0.020]*** {2.03}	**0.157** (-0.001) [0.023] {1.01}	**0.088** (-0.070) [0.013]*** {1.80}
Type 4	**Foreign Name** **Foreign Educ** **Foreign Exp**		**0.051** (-0.107) [0.017]** {3.10}	**0.053** (-0.105) [0.018]*** {2.98}	**0.052** (-0.106) [0.015]** {3.04}	**0.141** (-0.017) [0.021] {1.12}	**0.052** (-0.106) [0.011]*** {3.04}

Possible directions for you to consider

- Ways to improve program take-up (e.g. incentives, marketing, timing)
- Perhaps strong enough take-up impacts to test effectiveness (treatment on the treated)
- Randomize type of training

Some notes on random assignment

- Note that researcher is often restricted on things that can be randomized
- E.g., Not looking here at hiring decision, wages, longer term outcomes
- Randomized experiments often expensive, do not mimic real world situations, or use representative samples of total population

Suggestions for Practicing More Evidence Based Policy

- Acknowledge Preference for Evidence Based Policy
  - Favour outside research that incorporates experimental or quasi-experimental designs
  - Discount research that does not
  - Document best-practices for evidence based policy
  - (similar to US Department of Education setup)
- Support proposals for experiments that could help
  - Support proposals for experimental designs
  - Offer logistical support
- Initiate own experiments
  - Seek outside advice
  - Arms length evaluation and peer-review

Random Assignment Concerns

- **Concern 1: Set-up and analysis take too long**
  - ✓ Short-term outcomes on their own may be enough to motivate
  - ✓ Set-up not long in many circumstances;
  - ✓ Alternative options worse: they don't involve evidence based policy
- **Concern 2: Too many policy options to consider**
  - ✓ Same true in absence of RA
  - ✓ Lab experiments may help
  - ✓ Alternative options worse: implement policies that we don't know work
- **Concern 3: RA is unfair: some don't receive treatment**
  - ✓ Already programs exist that deny ineligible services
  - ✓ RA does not making people worse off than before
  - ✓ Can offer benefit to all that participate
- **Concern4: Sometimes impractical to setup RA**
  - ✓ True, but many times we can: RA is the gold standard
  - ✓ Pilot Projects with Matched Comparison Group or staggered roll-out good alternatives
  - ✓ Alternative options worse: they don't involve evidence based policy
- **Concern 5: RA too expensive**
  - ✓ Not All of them are
  - ✓ Alternative options worse: RA studies should get more weight that other studies
- **Concern 6: RA programs set up with belief 'program will work'**

A Math Slide

- Let *Y* by the outcome (proficiency), *T* be the treatment (training), let *X* be observed external factors or characteristics (all the other stuff that differs across people), and let *u* be unobserved external factors.
- $Y=aT+bX+u$
- *a* is the treatment effect we want to know.
- Comparing *Y* for *T=0* and *T=1* gives *a* **if either**
  - *b=0*, or
  - knowing *X* is not informative of *T*

Another Math Slide

- $Y=aT+bX+u$
- Does $b=0$?
  - Worries about external factors that affect *Y* are very natural. There's lots of stuff going on in people's lives other than settlement services and language training. So, *b* isn't *0*.
- Does knowing *X* tell me about *T*?
- **Not if treatment (*T*) is randomly assigned.**

The Last Math Slide

- $Y=aT+bX+u$
- Comparing *Y* for *T=0* and *T=1* gives *a* if you have random assignment.
- But, even without random assignment, if *a* is the same for people with different *X*, then comparing *Y* for *T=0* and *T=1* **for a given X** gives *a*.
- This is as good as random assignment.