Randomized Controlled Trials - Overview

Jeremy Magruder, Ph.D.

Assistant Professor
University of California, Berkeley

April, 2008

Why RCT?

- Randomized Controlled Trials are the gold standard of impact analysis
- Random individuals represent an ideal control group
- This section: overview the idea of a randomized control trial, frequent pitfalls, etc.

A simple example

- Suppose we want to estimate the effect of a new nutrition and vitamin program

A simple example

- Suppose we want to estimate the effect of a new nutrition and vitamin program
- This program, a group of nurses goes to villages and gives information to parents about proper nutrition for their children and also provides some vitamins.

A simple example

- Suppose we want to estimate the effect of a new nutrition and vitamin program
- This program, a group of nurses goes to villages and gives information to parents about proper nutrition for their children and also provides some vitamins.
- For the first year, the program can only be implemented in 1/3 of households because of costs

A simple example

- Suppose we want to estimate the effect of a new nutrition and vitamin program
- This program, a group of nurses goes to villages and gives information to parents about proper nutrition for their children and also provides some vitamins.
- For the first year, the program can only be implemented in 1/3 of households because of costs
- Suppose we want to investigate the effects of this program on child health and nutrition

Without a RCT

- Suppose the first year, the program is given to one province but not to another province

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health
- That is, if $\bar{Y}^P$ is the mean health of program children and $\bar{Y}^{NP}$ is the mean health of non program children, can measure $\bar{Y}^P - \bar{Y}^{NP}$

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health
- That is, if $\bar{Y}^P$ is the mean health of program children and $\bar{Y}^{NP}$ is the mean health of non program children, can measure $\bar{Y}^P - \bar{Y}^{NP}$
- We might find that the children who received the program are healthier than those who did not.

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health
- That is, if $\bar{Y}^P$ is the mean health of program children and $\bar{Y}^{NP}$ is the mean health of non program children, can measure $\bar{Y}^P - \bar{Y}^{NP}$
- We might find that the children who received the program are healthier than those who did not.
- But, we can't say whether this is due to the program – we don't know how people who received it are different from those who did not.

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health
- That is, if $\bar{Y}^P$ is the mean health of program children and $\bar{Y}^{NP}$ is the mean health of non program children, can measure $\bar{Y}^P - \bar{Y}^{NP}$
- We might find that the children who received the program are healthier than those who did not.
- But, we can't say whether this is due to the program – we don't know how people who received it are different from those who did not.
- Maybe the province with the program is a richer province where people have more money to spend on health services

Without a RCT

- Suppose the first year, the program is given to one province but not to another province
- Then, we compare the children that received the program to those that did not in terms of child health
- That is, if $\bar{Y}^P$ is the mean health of program children and $\bar{Y}^{NP}$ is the mean health of non program children, can measure $\bar{Y}^P - \bar{Y}^{NP}$
- We might find that the children who received the program are healthier than those who did not.
- But, we can't say whether this is due to the program – we don't know how people who received it are different from those who did not.
- Maybe the province with the program is a richer province where people have more money to spend on health services
- These children might have been healthier even without the program. If we just compare them and attribute that effect to the program, we might have a bad estimate

Treatment on the Treated

Important distinction: let G_1 be the program group and G_2 be the non-program group. What we want to measure is

$$\bar{Y}^{P}_{G_1+G_2} - \bar{Y}^{NP}_{G_1+G_2}$$

What we actually measure is

$$\bar{Y}^{P}_{G_1} - \bar{Y}^{NP}_{G_2}$$

If $\bar{Y}^{P}_{G_1} \neq \bar{Y}^{P}_{G_2}$ or $\bar{Y}^{NP}_{G_1} \neq \bar{Y}^{NP}_{G_2}$ then, we are measuring the wrong thing. This happens if G_2 is different from G_1 in a way which causes them to be more or less healthy without the program or more or less impacted by the program.

Time differences

- What if we have data on the same program over time?

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program
- So, if group G_1 gets treated in time period t, we can observe $\bar{Y}^P_{G_1,t+1} - \bar{Y}^P_{G_1,t-1}$ as the program effect

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program
- So, if group G_1 gets treated in time period t, we can observe $\bar{Y}^P_{G_1,t+1} - \bar{Y}^P_{G_1,t-1}$ as the program effect
- But, other things might have changed too – maybe everyone got healthier in this time period

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program
- So, if group G_1 gets treated in time period t, we can observe $\bar{Y}^P_{G_1,t+1} - \bar{Y}^P_{G_1,t-1}$ as the program effect
- But, other things might have changed too – maybe everyone got healthier in this time period
- We could use group 2 as a control group – then we can compare $\bar{Y}^P_{G_1,t+1} - Y^P_{G_1,t-1} - (\bar{Y}^{NP}_{G_2,t+1} - \bar{Y}^{NP}_{G_2,t-1})$

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program
- So, if group G_1 gets treated in time period t, we can observe $\bar{Y}^P_{G_1,t+1} - \bar{Y}^P_{G_1,t-1}$ as the program effect
- But, other things might have changed too – maybe everyone got healthier in this time period
- We could use group 2 as a control group – then we can compare $\bar{Y}^P_{G_1,t+1} - Y^P_{G_1,t-1} - (\bar{Y}^{NP}_{G_2,t+1} - \bar{Y}^{NP}_{G_2,t-1})$
- That is, use the changes in group 2 as a control group for changes in group 1

Time differences

- What if we have data on the same program over time?
- We can observe how people's health changes in response to the program
- So, if group G_1 gets treated in time period t, we can observe $\bar{Y}^P_{G_1,t+1} - \bar{Y}^P_{G_1,t-1}$ as the program effect
- But, other things might have changed too – maybe everyone got healthier in this time period
- We could use group 2 as a control group – then we can compare $\bar{Y}^P_{G_1,t+1} - Y^P_{G_1,t-1} - (\bar{Y}^{NP}_{G_2,t+1} - \bar{Y}^{NP}_{G_2,t-1})$
- That is, use the changes in group 2 as a control group for changes in group 1
- but, if people get to choose to participate in the program, might expect the biggest changes to be in group 1 – that's why they want to be in the program

With an RCT

- Suppose we randomly treat 1/2 of people and not others

With an RCT

- Suppose we randomly treat $1/2$ of people and not others
- Then, these people should be the same on average except for the program – this means $\bar{Y}^{P}_{G_1} = \bar{Y}^{P}_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$

With an RCT

- Suppose we randomly treat 1/2 of people and not others
- Then, these people should be the same on average except for the program – this means $\bar{Y}^{P}_{G_1} = \bar{Y}^{P}_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$
- We can then compare how healthy children are in program villages to non-program to understand the effect of the program

With an RCT

- Suppose we randomly treat 1/2 of people and not others
- Then, these people should be the same on average except for the program – this means $\bar{Y}^P_{G_1} = \bar{Y}^P_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$
- We can then compare how healthy children are in program villages to non-program to understand the effect of the program
- That is, $\bar{Y}^P_{G_1} = \bar{Y}^P_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$ so we $\bar{Y}^P_{G_1} - \bar{Y}^{NP}_{G_2}$ should give us a consistent estimate of the program.

With an RCT

- Suppose we randomly treat 1/2 of people and not others
- Then, these people should be the same on average except for the program – this means $\bar{Y}^P_{G_1} = \bar{Y}^P_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$
- We can then compare how healthy children are in program villages to non-program to understand the effect of the program
- That is, $\bar{Y}^P_{G_1} = \bar{Y}^P_{G_2}$ and $\bar{Y}^{NP}_{G_1} = \bar{Y}^{NP}_{G_2}$ so we $\bar{Y}^P_{G_1} - \bar{Y}^{NP}_{G_2}$ should give us a consistent estimate of the program.
- When/how to do this?

How do we define groups?

- We discussed randomizing over groups G_1 and G_2

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?
  - First: randomization will tell us the average of the effect of the program on everybody in the area.

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?
  - First: randomization will tell us the average of the effect of the program on everybody in the area.
  - Second: randomization will tell us the effect of the program on people who would adopt in the area. Both outcomes might be of interest.

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?
  - First: randomization will tell us the average of the effect of the program on everybody in the area.
  - Second: randomization will tell us the effect of the program on people who would adopt in the area. Both outcomes might be of interest.
    - Potential problem: often don't know who would have adopted in Control areas. So $\bar{Y}_{G_2}^{NP} = \bar{Y}_{G_2^{all}}^{NP}$.

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?
  - First: randomization will tell us the average of the effect of the program on everybody in the area.
  - Second: randomization will tell us the effect of the program on people who would adopt in the area. Both outcomes might be of interest.
    - Potential problem: often don't know who would have adopted in Control areas. So $\bar{Y}_{G_2}^{NP} = \bar{Y}_{G_2^{all}}^{NP}$.
    - If we don't know who would have adopted, and compare adopters in treatment to everyone in control group, not a perfect control group.

How do we define groups?

- We discussed randomizing over groups G_1 and G_2
- How are these defined?
  - Could make the program available to people in one (random) area but not another
    - Group could be everyone in the randomly assigned area (call it G_1^{all})
    - Or, Group could be the people who would adopt treatment in that area (call it G_1^{Treat})
- What do we measure with each of these groups?
  - First: randomization will tell us the average of the effect of the program on everybody in the area.
  - Second: randomization will tell us the effect of the program on people who would adopt in the area. Both outcomes might be of interest.
    - Potential problem: often don't know who would have adopted in Control areas. So $\bar{Y}_{G_2}^{NP} = \bar{Y}_{G_2^{all}}^{NP}$.
    - If we don't know who would have adopted, and compare adopters in treatment to everyone in control group, not a perfect control group.
    - So, either need to use G_1^{all} or would need to know who would and wouldn't adopt in each group.

Could we do an RCT?

- What level could we randomize at?

Could we do an RCT?

- What level could we randomize at?
- How long should it take to observe the results?

Could we do an RCT?

- What level could we randomize at?
- How long should it take to observe the results?
- Are there spillovers to the control group?

Could we do an RCT?

- What level could we randomize at?
- How long should it take to observe the results?
- Are there spillovers to the control group?
- What variables can we measure to see how effective it is?

Could we do an RCT?

- What level could we randomize at?
- How long should it take to observe the results?
- Are there spillovers to the control group?
- What variables can we measure to see how effective it is?
- Is it ethical to randomize at this level?

Could we do an RCT?

- What level could we randomize at?
- How long should it take to observe the results?
- Are there spillovers to the control group?
- What variables can we measure to see how effective it is?
- Is it ethical to randomize at this level?
- Can we generalize the result to learn more about other programs that might work?

Level of RCT

- For some programs, you can randomize at a household level, for others, a village level, for others maybe bigger.

Level of RCT

- For some programs, you can randomize at a household level, for others, a village level, for others maybe bigger.
- How to randomize will depend on the level – may be able to just shuffle some envelopes with different treatments, or run a computer program to randomly assign

Level of RCT

- For some programs, you can randomize at a household level, for others, a village level, for others maybe bigger.
- How to randomize will depend on the level – may be able to just shuffle some envelopes with different treatments, or run a computer program to randomly assign
- Household level advantages - households within the same village may be more comparable, more households to choose from

Level of RCT

- For some programs, you can randomize at a household level, for others, a village level, for others maybe bigger.
- How to randomize will depend on the level – may be able to just shuffle some envelopes with different treatments, or run a computer program to randomly assign
- Household level advantages - households within the same village may be more comparable, more households to choose from
- Household level disadvantage - spillovers – other households may learn from the ones who get treated

Level of RCT

- For some programs, you can randomize at a household level, for others, a village level, for others maybe bigger.
- How to randomize will depend on the level – may be able to just shuffle some envelopes with different treatments, or run a computer program to randomly assign
- Household level advantages - households within the same village may be more comparable, more households to choose from
- Household level disadvantage - spillovers – other households may learn from the ones who get treated
- Household level disadvantage - ethics – other households may demand treatment

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages
- But... spillovers? Will these households talk to other households in the village? Seems likely for information

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages
- But... spillovers? Will these households talk to other households in the village? Seems likely for information
- Then what if we compare people with the program to those without?

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages
- But... spillovers? Will these households talk to other households in the village? Seems likely for information
- Then what if we compare people with the program to those without?
- Will the doctors/others enumerating the survey be able to avoid giving out extra vitamins to others in the community?

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages
- But... spillovers? Will these households talk to other households in the village? Seems likely for information
- Then what if we compare people with the program to those without?
- Will the doctors/others enumerating the survey be able to avoid giving out extra vitamins to others in the community?
- What if other villages learn about this program and want it?

Info and Vitamins program level

- Advantages of randomizing at the household level: more households to choose from, households in the same village are probably more comparable than households in different villages
- But... spillovers? Will these households talk to other households in the village? Seems likely for information
- Then what if we compare people with the program to those without?
- Will the doctors/others enumerating the survey be able to avoid giving out extra vitamins to others in the community?
- What if other villages learn about this program and want it?
- What are the ethics of giving the program to some and not others?

Spillovers

- Suppose we give several households at random information about nutrition as well as some vitamin pills and not others

Spillovers

- Suppose we give several households at random information about nutrition as well as some vitamin pills and not others
- Then, we compare the health outcomes of the households with information to those without

Spillovers

- Suppose we give several households at random information about nutrition as well as some vitamin pills and not others
- Then, we compare the health outcomes of the households with information to those without
- What if information is the only important thing, and these households just share this information?

Spillovers

- Suppose we give several households at random information about nutrition as well as some vitamin pills and not others
- Then, we compare the health outcomes of the households with information to those without
- What if information is the only important thing, and these households just share this information?
- Then, the households without the intervention will look just as well off as those who received it

Spillovers

- Suppose we give several households at random information about nutrition as well as some vitamin pills and not others
- Then, we compare the health outcomes of the households with information to those without
- What if information is the only important thing, and these households just share this information?
- Then, the households without the intervention will look just as well off as those who received it
- In other words, with spillovers, the control group may become invalid. May be a case for village-level randomization.

Successful Randomization

- How can we be sure it is random?

Successful Randomization

- How can we be sure it is random?
- Can compare statistics of treated individuals to those of untreated

Successful Randomization

- How can we be sure it is random?
- Can compare statistics of treated individuals to those of untreated
- So, if it's village level, would want to check if education, income, ethnicities, etc. are similar in treatment and control villages

Successful Randomization

- How can we be sure it is random?
- Can compare statistics of treated individuals to those of untreated
- So, if it's village level, would want to check if education, income, ethnicities, etc. are similar in treatment and control villages
- Why do it randomly? Why not just pick villages which are similar in terms of education, income, etc.?

Successful Randomization

- How can we be sure it is random?
- Can compare statistics of treated individuals to those of untreated
- So, if it's village level, would want to check if education, income, ethnicities, etc. are similar in treatment and control villages
- Why do it randomly? Why not just pick villages which are similar in terms of education, income, etc.?
  - Don't observe everything about these villages; villages that look similar but behave differently may be the most different from each other

Timing of results

- If everyone is going to get the program, can't randomize totally

Timing of results

- If everyone is going to get the program, can't randomize totally
- but may be able to randomize timing

Timing of results

- If everyone is going to get the program, can't randomize totally
- but may be able to randomize timing
- For example, if it takes time to expand access, we may only be able to give it to 1/3 of people this year anyway

Timing of results

- If everyone is going to get the program, can't randomize totally
- but may be able to randomize timing
- For example, if it takes time to expand access, we may only be able to give it to 1/3 of people this year anyway
- In that case, why not make that a random 1/3, then next year can look at that 1/3 and compare to others

Timing of results

- If everyone is going to get the program, can't randomize totally
- but may be able to randomize timing
- For example, if it takes time to expand access, we may only be able to give it to 1/3 of people this year anyway
- In that case, why not make that a random 1/3, then next year can look at that 1/3 and compare to others
- Importance: In this case, can only pick up effects that happen within 1 year

Nutrition and Info within 1 year

- What might change?

Nutrition and Info within 1 year

- What might change?
- height?

Nutrition and Info within 1 year

- What might change?
- height?
- weight?

Nutrition and Info within 1 year

- What might change?
- height?
- weight?
- eating habits?

Nutrition and Info within 1 year

- What might change?
- height?
- weight?
- eating habits?
- school attendance (why should this be impacted)?

Nutrition and Info within 1 year

- What might change?
- height?
- weight?
- eating habits?
- school attendance (why should this be impacted)?
- Nutritional knowledge quiz

Nutrition and Info within 1 year

- What might change?
- height?
- weight?
- eating habits?
- school attendance (why should this be impacted)?
- Nutritional knowledge quiz
- vitamin/other nutrient frequencies in blood?

Reporting

- Do we trust these kind of effects?

Reporting

- Do we trust these kind of effects?
- For objective measures, like height/weight/blood frequencies: pretty unassailable

Reporting

- Do we trust these kind of effects?
- For objective measures, like height/weight/blood frequencies: pretty unassailable
- But, expensive to collect

Reporting

- Do we trust these kind of effects?
- For objective measures, like height/weight/blood frequencies: pretty unassailable
- But, expensive to collect
- Survey data, like eating habits, school attendance, easier to collect

Reporting

- Do we trust these kind of effects?
- For objective measures, like height/weight/blood frequencies: pretty unassailable
- But, expensive to collect
- Survey data, like eating habits, school attendance, easier to collect
- But, worry about honesty – people may try to give the "correct" answer to official surveyors

Randomized Review

- Suppose we implement the program in 1/3 of villages (to limit spillovers) at random for the first year

Randomized Review

- Suppose we implement the program in 1/3 of villages (to limit spillovers) at random for the first year
- We compare weight, reported eating habits and schooling attendance between program villages and non-program villages

Randomized Review

- Suppose we implement the program in 1/3 of villages (to limit spillovers) at random for the first year
- We compare weight, reported eating habits and schooling attendance between program villages and non-program villages
- and, ultimately, we find that the program had an impact - children in program villages weigh more, report better eating habits, and go to school more

Randomized Review

- Suppose we implement the program in 1/3 of villages (to limit spillovers) at random for the first year
- We compare weight, reported eating habits and schooling attendance between program villages and non-program villages
- and, ultimately, we find that the program had an impact - children in program villages weigh more, report better eating habits, and go to school more
- dealt with ethics: could only provide to 1/3 of villages, anyway, and will provide to all next year

Randomized Review

- Suppose we implement the program in 1/3 of villages (to limit spillovers) at random for the first year
- We compare weight, reported eating habits and schooling attendance between program villages and non-program villages
- and, ultimately, we find that the program had an impact - children in program villages weigh more, report better eating habits, and go to school more
- dealt with ethics: could only provide to 1/3 of villages, anyway, and will provide to all next year
- But... what can we learn?

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.
- Because program did both vitamins and information, we don't know which is more important

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.
- Because program did both vitamins and information, we don't know which is more important
- For scale-up, it might be useful to know how effective each was in case one is much cheaper than the other

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.
- Because program did both vitamins and information, we don't know which is more important
- For scale-up, it might be useful to know how effective each was in case one is much cheaper than the other
- How would we solve this? Could randomize separately for the two treatments, maybe could collect data on taking vitamins

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.
- Because program did both vitamins and information, we don't know which is more important
- For scale-up, it might be useful to know how effective each was in case one is much cheaper than the other
- How would we solve this? Could randomize separately for the two treatments, maybe could collect data on taking vitamins
- Also, we only know that the program was effective for the group who we studied. Ideally, we can design programs which would be effective for more people so that we know we can scale up.

Generalizability

- We learned that the program worked, that providing both information and vitamins improved child health in this context.
- Because program did both vitamins and information, we don't know which is more important
- For scale-up, it might be useful to know how effective each was in case one is much cheaper than the other
- How would we solve this? Could randomize separately for the two treatments, maybe could collect data on taking vitamins
- Also, we only know that the program was effective for the group who we studied. Ideally, we can design programs which would be effective for more people so that we know we can scale up.
- Over the rest of the afternoon, we'll discuss more case-studies which work on these issues and how best to do RCTs.