

# Analysis Plans in Economics

Benjamin Olken

MIT

December 2012

Where did this come from... for me.



# Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke
  - ▶ What I remember



## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke
  - ▶ What I remember (besides fainting)

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke
  - ▶ What I remember (besides fainting) : the programs had to be completely done, and tested on the actual data, before we knew what was treatment and what was control

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke
  - ▶ What I remember (besides fainting) : the programs had to be completely done, and tested on the actual data, before we knew what was treatment and what was control
  - ▶ Only after all programs were done did we “break the code” and look at which were treatment and which were control

## Where did this come from... for me.

- ▶ High school internship in a medical lab doing research on anti-stroke drugs
  - ▶ Kinematsu et al (1993): “Effects of a novel NMDA antagonist on experimental stroke rapidly and quantitatively assessed by diffusion-weighted MRI”
  - ▶ Double-blind drug study on lab rats
  - ▶ Gave them treatment vs. placebo drug and then tested how they responded to strokes
  - ▶ My job: writing the computer programs to code how much of their brain had a stroke
  - ▶ What I remember (besides fainting) : the programs had to be completely done, and tested on the actual data, before we knew what was treatment and what was control
  - ▶ Only after all programs were done did we “break the code” and look at which were treatment and which were control
- ▶ Completely standard practice in medical trials – and required in fact by the FDA

Where did this come from... for me.

- ▶ Deaton (2009): “Randomization in the tropics”:

## Where did this come from... for me.

### ► Deaton (2009): “Randomization in the tropics”:

“It is possible, for example, to estimate whether the average treatment effect is positive or negative for various subgroups of interest. One immediate charge against such a procedure is data mining. A sufficiently determined examination of any trial will eventually reveal some subgroup for whom the treatment yielded a significant effect of some sort, and there is no general way of adjusting standard errors to protect against the possibility. In drug trials, the FDA rules require that analytical plans be submitted prior to trial, and at least one economic experiment—moving to opportunity—has imposed similar rules on itself, see the protocol by Feins and McInnis (2001). I am not arguing against post-trial subgroup analysis, only that any special epistemic status (as in “gold standard,” “hard,” or “rigorous” evidence) possessed by RCTs does not extend to subgroup analysis if only because there is no general guarantee that a new RCT on post-experimentally defined subgroups will yield the same result.”

## On the other hand...

- ▶ Deaton: “Randomization in the tropics”:

“It is clearly absurd to discard data because we do not know how to analyze it with sufficient purity. Indeed, many important findings have come from post-trial analysis of experimental data, both in medicine and in economics, for example of the negative income tax experiments of the 1960s. None of which resolves the concerns about data-mining. In large-scale, expensive, trials, a zero or very small result is unlikely to be welcome, and there is likely to be overwhelming pressure to search for some subpopulation that gives a more palatable result.”

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:



# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante
  - ▶ Specified a way of structuring the analysis – what program did, impact on main indicators, long run health and education patterns, channels, pitfalls of incentives, etc – with tables

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante
  - ▶ Specified a way of structuring the analysis – what program did, impact on main indicators, long run health and education patterns, channels, pitfalls of incentives, etc – with tables
  - ▶ Specified key dimension of heterogeneity to examine (baseline levels of performance)

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante
  - ▶ Specified a way of structuring the analysis – what program did, impact on main indicators, long run health and education patterns, channels, pitfalls of incentives, etc – with tables
  - ▶ Specified key dimension of heterogeneity to examine (baseline levels of performance)
- ▶ This resulted in a total of 62 tables, each with 10-15 variables.

# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante
  - ▶ Specified a way of structuring the analysis – what program did, impact on main indicators, long run health and education patterns, channels, pitfalls of incentives, etc – with tables
  - ▶ Specified key dimension of heterogeneity to examine (baseline levels of performance)
- ▶ This resulted in a total of 62 tables, each with 10-15 variables.
- ▶ How to solve data mining?



# Some examples from my current experience

A classic program-evaluation: Project Generasi (with Onishi and Wong)

- ▶ Setting:
  - ▶ Classic program evaluation of a CDD program aimed at improving infant and child health and education, on 12 main indicators
  - ▶ Three groups: with incentives, without incentives, control
- ▶ Analysis plan (inspired by the Deaton paper):
  - ▶ Specified all variables to be examined ex-ante
  - ▶ Specified a way of structuring the analysis – what program did, impact on main indicators, long run health and education patterns, channels, pitfalls of incentives, etc – with tables
  - ▶ Specified key dimension of heterogeneity to examine (baseline levels of performance)
- ▶ This resulted in a total of 62 tables, each with 10-15 variables.
- ▶ How to solve data mining? Ex-ante designated one table as “key results” table and one number, the average standardized effects for 12 main indicators, as key metric of success.

# The good and the bad

- ▶ The Good.

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.*

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
    - ▶ Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
    - ▶ Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
    - ▶ I absolutely recommend this in the case of when you have an interested party!

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
    - ▶ Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
    - ▶ I absolutely recommend this in the case of when you have an interested party!
  - ▶ *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.

# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
    - ▶ Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
    - ▶ I absolutely recommend this in the case of when you have an interested party!
  - ▶ *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.
  - ▶ *Specification.* Lags, first difference, which controls, blah blah. None of this will matter much, but since we all are obsessed with p-values at critical cutoffs (0.049 vs. 0.051; 0.099 vs. 0.101), it matters a bit.



# The good and the bad

- ▶ The Good.
  - ▶ *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
    - ▶ Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
    - ▶ I absolutely recommend this in the case of when you have in interested party!
  - ▶ *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.
  - ▶ *Specification.* Lags, first difference, which controls, blah blah. None of this will matter much, but since we all are obsessed with p-values at critical cutoffs (0.049 vs. 0.051; 0.099 vs. 0.101), it matters a bit. This is the perfect thing to pre-specify.

# The good and the bad

## ► The Good.

- *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
  - Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
  - I absolutely recommend this in the case of when you have an interested party!
- *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.
- *Specification.* Lags, first difference, which controls, blah blah. None of this will matter much, but since we all are obsessed with p-values at critical cutoffs (0.049 vs. 0.051; 0.099 vs. 0.101), it matters a bit. This is the perfect thing to pre-specify.
- *Randomization balance checks.*

# The good and the bad

## ► The Good.

- *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
  - Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
  - I absolutely recommend this in the case of when you have in interested party!
- *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.
- *Specification.* Lags, first difference, which controls, blah blah. None of this will matter much, but since we all are obsessed with p-values at critical cutoffs (0.049 vs. 0.051; 0.099 vs. 0.101), it matters a bit. This is the perfect thing to pre-specify.
- *Randomization balance checks.* Which variables go in the balance check?

# The good and the bad

## ► The Good.

- *Dealing with negative results.* We had a truly unexpected result where our health and education program made enrollments worse.
  - Could have killed it by looking at alternative variable definitions, but had to present it as is to the government!
  - I absolutely recommend this in the case of when you have in interested party!
- *Interactions.* We pre-specified one key dimension of heterogeneity: baseline level. Turned out that was a good idea – makes a big difference. Much more credible since we specified it ex-ante.
- *Specification.* Lags, first difference, which controls, blah blah. None of this will matter much, but since we all are obsessed with p-values at critical cutoffs (0.049 vs. 0.051; 0.099 vs. 0.101), it matters a bit. This is the perfect thing to pre-specify.
- *Randomization balance checks.* Which variables go in the balance check? (Aside I find this irritating).

# The good and the bad

- ▶ The Bad.

# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.

# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.
  - ▶ Why?

# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.
  - ▶ Why? Because you need to specify the tables before you know the results. Makes it very hard to tell a story, and puts you in to a very standardized methodology.



# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.
  - ▶ Why? Because you need to specify the tables before you know the results. Makes it very hard to tell a story, and puts you in to a very standardized methodology.
- ▶ Some thoughts:

# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.
  - ▶ Why? Because you need to specify the tables before you know the results. Makes it very hard to tell a story, and puts you in to a very standardized methodology.
- ▶ Some thoughts:
  - ▶ *Endline vs. midline.* We did a new endline analysis plan again after having looked at midline data, but before looking at endline data. This seems totally kosher to me: this is fresh data.

# The good and the bad

- ▶ The Bad.
  - ▶ This can make for a really boring paper.
  - ▶ Why? Because you need to specify the tables before you know the results. Makes it very hard to tell a story, and puts you in to a very standardized methodology.
- ▶ Some thoughts:
  - ▶ *Endline vs. midline.* We did a new endline analysis plan again after having looked at midline data, but before looking at endline data. This seems totally kosher to me: this is fresh data.
  - ▶ *One-sided tests.* My view is we should be allowed to specify one-sided p-values in our analysis plans. At least we would get something in return for pre-specifying, since in many cases – like my education example – that's how we react anyway!

# Some examples from my current experience

## Targeting the Poor (with Alatas Banerjee, Hanna, and Tobias)

- ▶ Setting:
  - ▶ Study in Indonesia on how to identify poor households for anti-poverty programs
  - ▶ Three treatments: PMT, Community Targeting, and Hybrid (to prevent elite capture of community process)

## The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.

# The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.
  - ▶ Is this the right use of analytical effort?

# The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.
  - ▶ Is this the right use of analytical effort?
  - ▶ Can we even do this properly? Very hard to think hard contingent on actual results; even harder to think hard about any possible results you could possibly have.

# The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.
  - ▶ Is this the right use of analytical effort?
  - ▶ Can we even do this properly? Very hard to think hard contingent on actual results; even harder to think hard about any possible results you could possibly have.
- ▶ Papers change contingent on results.



# The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.
  - ▶ Is this the right use of analytical effort?
  - ▶ Can we even do this properly? Very hard to think hard contingent on actual results; even harder to think hard about any possible results you could possibly have.
- ▶ Papers change contingent on results.
  - ▶ Paper ended up being about something very different than we thought. (No elite capture; about different local welfare metrics)

# The good and the bad

- ▶ It is hard to scope out all possible theories and all possible trees in advance. And we spend a huge amount of time and effort plumbing out parts of the tree that never materialized.
  - ▶ Is this the right use of analytical effort?
  - ▶ Can we even do this properly? Very hard to think hard contingent on actual results; even harder to think hard about any possible results you could possibly have.
- ▶ Papers change contingent on results.
  - ▶ Paper ended up being about something very different than we thought. (No elite capture; about different local welfare metrics)
  - ▶ Luckily in this case our analysis plan was sufficiently general that it was OK. But usually that won't be true. How to retain this flexibility? What if we had been required to report tons of null results on elite capture for parts of the tree we didn't explore?

# Some examples from current experience

## Taxes in Pakistan (with Khwaja and Khan)

- ▶ Setting:
  - ▶ Multi-year RCT in Pakistan on tax collectors
  - ▶ Treatment is pay-for-performance incentives for tax collectors (3 versions thereof), vs. control

# The good and the bad

- ▶ We didn't do an analysis plan.

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?
- ▶ Similar issue with in a new project on ID cards in Indonesia.



# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?
- ▶ Similar issue with in a new project on ID cards in Indonesia.
  - ▶ Asked to report to the Indonesian VP 3 months after start.

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?
- ▶ Similar issue with in a new project on ID cards in Indonesia.
  - ▶ Asked to report to the Indonesian VP 3 months after start.
  - ▶ Focus on getting the project started and organizing the survey.

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?
- ▶ Similar issue with in a new project on ID cards in Indonesia.
  - ▶ Asked to report to the Indonesian VP 3 months after start.
  - ▶ Focus on getting the project started and organizing the survey.
  - ▶ Doing analysis plan rapidly, but hard to think out every detail.

# The good and the bad

- ▶ We didn't do an analysis plan. Why not?
  - ▶ Needed results instantly for policy purposes. Under constant pressure to launch and make payments, government wanted initial results three months after we started. Just didn't have time. Should we reject anything that doesn't have an analysis plan?
  - ▶ Our understanding is changing over two years as we do more qualitative field work and non-experimental data analysis. Final paper will look very different - and analysis we'll want to do - than if committed ex-ante to an analysis plan. How to deal with this if you also need instant results?
- ▶ Similar issue with in a new project on ID cards in Indonesia.
  - ▶ Asked to report to the Indonesian VP 3 months after start.
  - ▶ Focus on getting the project started and organizing the survey.
  - ▶ Doing analysis plan rapidly, but hard to think out every detail.
  - ▶ Is it better to do a not-perfect analysis plan than one at all? Will we be hamstrung in our analysis in the future?

# Some thoughts

- ▶ When pre-analysis plans seem most helpful.
  - ▶ When you have a party with strong vested interest. Get them to sign off up front!
  - ▶ When you suspect subgroup analysis is going to be a big deal. See Deaton.
  - ▶ Specifying your specifications. Usually a good idea.
  - ▶ Specifying control variables for balance checks.
  - ▶ When you have a simple outcome variable or can define one; or, alternatively, when you have a zillion outcome variables and are worried about data mining.

# Some thoughts

- ▶ When pre-analysis plans seem most helpful.
  - ▶ When you have a party with strong vested interest. Get them to sign off up front!
  - ▶ When you suspect subgroup analysis is going to be a big deal. See Deaton.
  - ▶ Specifying your specifications. Usually a good idea.
  - ▶ Specifying control variables for balance checks.
  - ▶ When you have a simple outcome variable or can define one; or, alternatively, when you have a zillion outcome variables and are worried about data mining.
- ▶ When it can be harder... and why we should be careful not to necessarily penalize people for not having them.
  - ▶ When things happen in real time (aside: how does medicine deal with this? optimal stopping rules? I don't know)
  - ▶ When you are testing theories
  - ▶ When your understanding of what you're testing evolves as you move through the project