Analyze as you randomize

- For simple and complete randomization
  - A simple difference in means is an unbiased estimator for the average treatment effect
- For block randomization
  - A regression with block fixed effects is an estimator for the average treatment effect
  - For designs with equal units in each arm in each block
  - Else block size weighted
- For cluster randomization
  - A regression with clustered standard errors is an estimator for the average treatment effect

Tables

- Should be self contained
  - Elaborate notes
  - Complete
  - Easy to interpret (Marginal Effects)
- Start here, them read interpretation in text
- What to look for
  - Check against design (correct test of H?)
  - Check against sample (attrition?)
  - Check for weird values (descriptives, coefficients)
  - Do results look credible? (do you believe them?)
  - What is not reported (all the data? all the models?)

Recall CDD example

- Complete randomization: treatment vs control
- Many hypothesis and outcomes
- Results:
  - large effects on development hardware (FamA)
  - No difference in participation, democratization, inclusion (FamB)
- Lets have a look:

Balance

TABLE I

BASELINE (2005) COMPARISON BETWEEN TREATMENT AND CONTROL COMMUNITIES

	(1) Baseline mean for controls	(2) T-C difference at baseline	(3) *N*
Panel A: Community characteristics			
Total households per community	46.76	0.30	236
		(3.67)	
Distance to nearest motorable road in miles	2.99	–0.32	236
		(0.36)	
Index of war exposure (range 0 to 1)	0.68	–0.01	236
		(0.02)	
Historical extent of domestic slavery (range 0 to 1)	0.36	0.03	236
		(0.06)	
Average respondent years of education	1.65	0.11	235
		(0.13)	
Panel B: Selected variables from "hardware" family A			
Proportion of communities with a village development committee (VDC)	0.55	0.06	232
		(0.06)	
Proportion visited by ward development committee (WDC) member in past year	0.15	–0.01	228
		(0.05)	
Proportion of communities with a functional grain-drying floor	0.23	0.05	231
		(0.05)	
Proportion of communities with a functional primary school	0.41	0.08	230
		(0.06)	
Average household asset score	–0.06	0.11	235
		(0.08)	
Proportion of communities with any petty traders	0.54	–0.01	226
		(0.06)	

(continued)

ATE

TABLE II

GoBifo Treatment Effects by Research Hypothesis

Hypotheses by family	(1) GoBifo mean treatment effect endex	(2) Naive p-value	(3) FWER-adjusted p-value for all 12 hypos	(4) FWER-adjusted p-value for 11 hypos in 2009 PAP
Family A: Development infrastructure or "hardware" effects				
Mean effect for family A (Hypotheses 1–3; 39 unique outcomes)	**0.298**** **(0.031)**	0.000		
H1: GoBifo project implementation (7 outcomes)	0.703** (0.055)	0.000	0.000	
H2: Participation in GoBifo improves the quality of local public services infrastructure (18 outcomes)	0.204** (0.039)	0.000	0.000	0.000
H3: Participation in GoBifo improves general economic wel fare (15 outcomes)	0.376** (0.047)	0.000	0.000	0.000
Family B: Institutional and social change or "software" effects				
Mean effect for family B (Hypotheses 4–12; 155 unique outcomes)	0.028 **(0.020)**	0.155		
H4: Participation in GoBifo increases collective action and contributions to local public goods (15 outcomes)	0.012 (0.037)	0.738	0.980	0.981
H5: GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent, and accountable (47 outcomes)	0.002 (0.032)	0.944	0.980	0.981
H6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government (25 outcomes)	0.056 (0.037)	0.134	0.664	0.667

TABLE II

(CONTINUED)

Hypotheses by family	(1) GoBifo mean treatment effect endex	(2) Naive p-value	(3) FWER-adjusted p-value for all 12 hypos	(4) FWER-adjusted p-value for 11 hypos in 2009 PAP
H7: Participation in GoBifo increases trust (12 outcomes)	0.042			
	(0.046)	0.360	0.913	0.914
H8: Participation in GoBifo builds and strengthens community groups and networks (15 outcomes)	0.028			
	(0.037)	0.450	0.913	0.914
H9: Participation in GoBifo increases access to information about local governance (17 outcomes)	0.038			
	(0.037)	0.301	0.913	0.913
H10: GoBifo increases public participation in local governance (18 outcomes)	0.090*			
	(0.045)	0.045	0.315	0.322
H11: By increasing trust, GoBifo reduces crime and conflict in the community (8 outcomes)	0.010			
	(0.043)	0.816	0.980	0.981
H12: GoBifo changes political and social attitudes, making individuals more liberal toward women, more accepting of other ethnic groups and "strangers," and less tolerant of corruption and violence (9 outcomes)	0.041			
	(0.043)	0.348	0.913	0.914

Notes: Significance levels (naive p-value) indicated by $+p < .10$, $*p < .05$, $**p < .01$. Robust standard errors in parentheses. Includes fixed effects for the district council wards (the unit of stratification) and the two balancing variables used in the randomization process—total households per community and distance to nearest motorable road. These mean effect estimates are limited to endline data only and the full sample set of outcomes that excludes all conditional outcomes (i.e., those that depend on the state of another variable, e.g., quality of infrastructure depends on the existence of the infrastructure). Construction of the mean effects index in column (1) gives equal weight to each component (following Kling, Liebman, and Katz 2007) as specified in the PAP. Family-wise error rate (FWER) adjusted p-values limit the probability of making any Type I errors when considering the hypotheses as a group, where the group is defined as the final set of 12 hypotheses or the original 11 hypotheses in the presanalysis plan (Westfall and Young 1993 free step-down resampling method as detailed in Anderson 2008). For the complete list of all variables under each hypothesis—including the exact wording of survey questions and treatment effect estimates—see Online Appendix J.

Conditional Average Treatment Effect (CATE)

- Why might you be interested in looking at **heterogeneous results**?
  - Important question: For whom are there big effects? For whom are there small effects? For whom does treatment generate beneficial or adverse effects?
  - What if you have ATE=0?

- You cannot measure causal effects for individual units (the fundamental problem), but you can do so for certain groups

- **CATE** = the average treatment effect specific to a subgroup of subjects

Conditional Average Treatment Effect (CATE)

- Why is pre-specifying CATE important?
  - no “fishing”
  - Multiple comparisons (pre-register, adjust the p-value (Bonferroni, etc.))
  - Statistical power. Ensure sufficient sample size
- If you are really interested in it: can you do a factorial?
- Can you explore heterogeneous effects conditional on variables that were changed by the intervention?

Heterogeneous effects with endogenous subgroups

- Say you are interested to know if CDD improved institutions (FamB) in in places that have more public goods (where H2 is true)
- The measure of public goods is in the endline
- This is a problem: as these outcomes where affected by the treatment itself
- Also a problem for any outcome under H3-H12

Heterogeneous Effects

- Only condition on pre-treatment outcomes
  - reason to do a baseline
- Or time invariant outcomes (but watch out with those)

- Say we are interested in learning if CDD has an effect in a District

- We add an interaction term to the model, treatment * covariate

Heterogeneity – dummy

$$Y_c = \beta_0 + \beta_1 T_c + \beta_2 X_c + \beta_3 T * X_c + D'_c + \varepsilon_c$$

	(1)	(2)
	FamA	FamB
CDD	0.457**	0.014
	(0.058)	(0.034)
Bombali	0.130	-0.096
	(0.128)	(0.068)
CDD*Bombali	-0.227**	0.023
	(0.074)	(0.042)
Constant	-0.149	0.103*
	(0.105)	(0.046)
N	236	236
R2	0.513	0.331

Heterogeneity - continuous

$$Y_c = \beta_0 + \beta_1 T_c + \beta_2 X_c + \beta_3 T * X_c + D'_c + \varepsilon_c$$

	(1)	(2)
	FamA	FamB
CDD	0.370**	0.076*
	(0.068)	(0.035)
Village size (hhs)	0.006**	0.001
	(0.001)	(0.000)
CDD*village size	-0.001	-0.001
	(0.001)	(0.001)
Constant	0.033	0.082
	(0.114)	(0.073)
N	236	236
R2	0.619	0.337

Heterogeneity: experimental subgroups

- Grants or
- Facilitation?

	Control	CDD Facilitation
Control	A	B
CDD Grant	C	D

Randomization

- Strict definition: a known probability (not equal)
- Put your data where you expect more variance
- Decide to reweight analysis
  - with (inverse) probabilities