Field Review: Empirical Specifications

Sergio Puerto

June 11 , 2025

Agenda

1. Stage 2: Identifying Treatment Effects
2. IDEAL-preferred Specification
3. Author-preferred
4. Rounds of data collection

1 Identifying Treatment Effects

Treatment Effects

- In the context of a randomized evaluation: “causal effect of a treatment”, “impact of an intervention”, “change in outcomes in an experiment” …
- Stage 2: Focus on exhibit-level information (e.g., tables and figures)
- TE = <contrast> x <outcome> x <rounds> x <specification>

Intervention
→ Arms
→ Treatment
→ Control

Randomization
→ Unit of analysis

Rounds of data collection
→ datasources

→ Estimand
→ Empirical model:
Author-preferred
IDEAL-preferred

Treatment Effects in Exhibits

Regression Table

	Outcome 1 (…)	
	Spec 1 (…)	Spec S
Contrast 1 (…)	TE-1-1-1	TE-1-s-r
Covariates		
Other info	round 1 (…)	round r

Number of TEs: Contrast-Outcome pair x s x r

Means and tests

	Contrast 1		
	Group 1	Group 2	Diff
Outcomes (…)	μ_1	μ_2	Spec 1
Other info			

Number of TEs: Outcomes x Specs

Field 1: Number of Treatment Effects

- Coding protocol:

Variable name in SurveyCTO	[tfx_num]
Field name (LABEL on SurveyCTO)	How many IDEAL **eligible** treatment effects in the exhibit are you going to report?
Definition (LABEL on SurveyCTO)	The number of IDEAL eligible treatment effects in the table to report.
Response options (open-text, numeric, date, text-CV)	Numeric
CV (Choices on SurveyCTO)	Integer greater than zero
SurveyCTO repeat level	Repeat for each table identified in Stage 1.

Field 1: Number of Treatment Effects

What estimates to count:

The total number of IDEAL-**eligible** treatment effects depends on the specification and the rounds of data collection:

1. Count outcome-contrast pairs
2. Count estimates for which full sample results are reported.
3. Count Count ITT estimates. Only count LATE / TOT estimates if they are preferred by the author or are the only estimates reported.
4. Count each treatment effect associated with different rounds of data collection.
5. Do not count estimates from interaction terms or linear combinations of estimates.

Field 1: Number of Treatment Effects

What estimates to count:

6. Count estimates with the highest-ranked IDEAL preferred specification:

#	Strata correction	Baseline outcome	Other controls	No other controls
1	✓	✓		
2	✓			✓
3	✓	✓	✓	
4	✓		✓	
5		✓		
6				✓
7		✓	✓	
8			✓	

Field 1: Number of Treatment Effects

Example 1: Riley (2024)

Outcomes: 3

Contrasts: 2

Specifications: 1

Rounds of data collection: ?

Number of TEs: 5

TABLE 1—TREATMENT EFFECTS ON WOMAN'S BUSINESS PROFITS, SAVINGS, AND BUSINESS CAPITAL

	Profit (1)	Savings (2)	Capital (3)
Mobile Account	2.88	0.93	13.48
	(3.61)	(9.54)	(24.18)
	[0.43]	[0.92]	[0.58]
	{1.00}	{1.00}	{1.00}
Mobile Disburse	17.61	8.46	
	(3.54)	(10.23)	
	[0.00]	[0.41]	
	{0.00}	{0.75}	
Observations	2,639	2,639	2,639
R^2	0.44	0.35	0.51
Control mean	109.81	155.33	678.28
Control mean baseline	116.62	134.32	655.90
p-value *MA* = *MD*	0.00	0.43	0.01

Notes: Intent-to-treat estimates; monetary outcomes are winsorized at the 99 percent level and in US dollars. All regressions include strata dummies and include the baseline value of the outcome. Mobile Account is the treatment where only a mobile money account was provided and the loan was disbursed as cash. Mobile Disburse is the treatment where a mobile money account was provided and the loan also disbursed onto this account. "Profit" refers to the woman's self-reported monthly business profit. "Savings" is individual savings held by the woman. "Capital" is the value of all assets the woman uses in her business plus the value of inventory held for her business. Control mean endline is the mean value of the outcome in the control group at endline. Control mean baseline is the mean value of the outcome in the control group at baseline. False discovery rate (FDR) adjusted *p*-values, also known as *q*-values, were used to correct for multiple hypothesis testing. They are shown in curly brackets. These were calculated following the method of Benjamini, Krieger, and Yekutieli (2006). Robust *p*-values in square brackets. Robust standard errors in parentheses.

Field 1: Number of Treatment Effects

Example 2: Basinga et al., (2011)

Outcomes: 8

Contrasts: 1

Specifications: 1

Rounds of data collection: 1

(2 different sources)

Number of TEs: 8

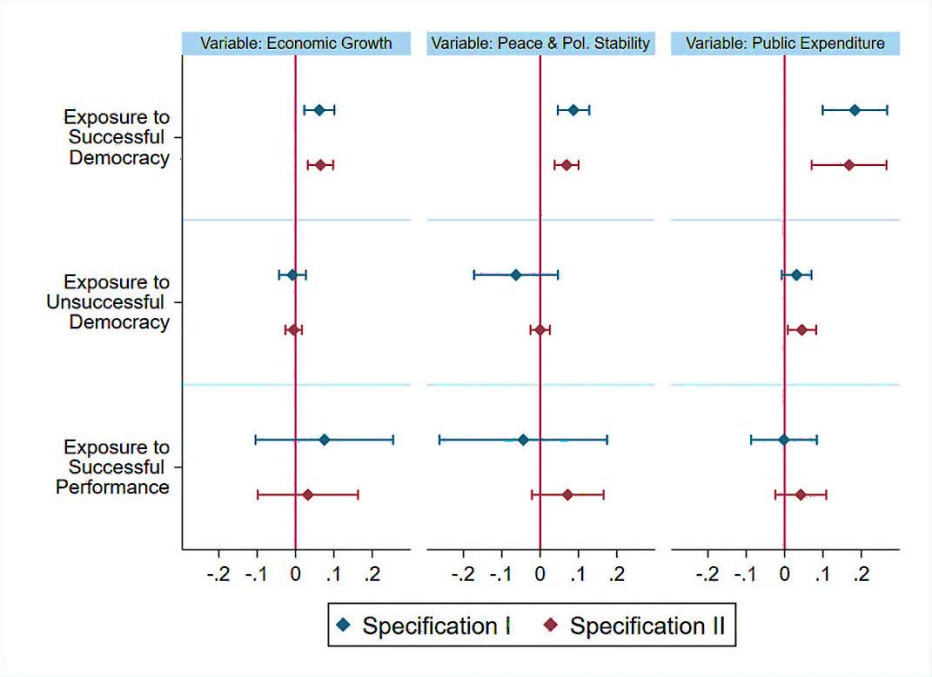
	N	β (95% CI)	p value	% D*
Maternal care use†				
Any prenatal care	2309	0·002 (–0·021 to 0·025)	0·875	0·2%
Four or more prenatal care visits	2223	0·008 (–0·063 to 0·079)	0·825	4·4%
Institutional delivery	2108	0·081¶ (0·015 to 0·146)	0·017	23·2%
Quality of prenatal care‡				
Tetanus vaccine during prenatal visit	2856	0·051|| (–0·002 to 0·103)	0·057	7·2%
Standardised total quality score	3826	0·157¶ (0·026 to 0·289)	0·020	N/A
Child preventive care use§				
Younger than 23 months preventive visit, previous 4 weeks	1971	0·119** (0·041 to 0·198)	0·004	55·9%
24–59 months preventive visit, previous 4 weeks	2902	0·111** (0·059 to 0·162)	0·000	131·6%
12–23 months fully immunised	872	–0·055 (–0·184 to 0·074)	0·390	–8·1%

N/A=not applicable. *The % D = (β / baseline mean) × 100, where the baseline mean of the dependent variable is the 2006 mean of the treatment group from tables 2 and 3. †The β is the estimated intervention effect controlling for a year dummy, facility-fixed effects, individual-level characteristics (age, education, partner lives in household, and number of pregnancies) and household characteristics (health insurance, number of household members, distance from the facility, land ownership, and assets value quartile). SEs were adjusted for clustering at the district-year level. ‡The β is the estimated treatment effect controlling for a year dummy, facility fixed effects, patient-level characteristics (age, education, partner lives in household, and insurance enrolment), and the source of the information (patient exit interview or household survey). SEs were adjusted for clustering at the district-year level. The number of observations in the tetanus model is less than in the quality score model because tetanus is only given to women with five or fewer previous pregnancies. SEs were adjusted for clustering at the district-year level. §The β is the estimated treatment effect, controlling for a year dummy, facility-fixed effects, individual-level characteristics (age, sex), parental-level characteristics (height, age and education of mother, father lives in household) and household characteristics (health insurance, number of household members, number of household members younger than 6 years, land ownership, assets value quartile). SEs were adjusted for clustering at the district-year level. ¶p<0·05. ||p<0·1. **p<0·01.

***Table 6*: Estimated effect of P4P schemes on maternal and child health-care services**

Field 1: Number of Treatment Effects

Example 3: Acemoglu et al. (2021)



Outcomes: 3

Contrasts: 3

Specifications: 2

Rounds of data collection: ?

Number of TEs: 0 or 9 or 18

-> Requires additional information

Field 1: Number of Treatment Effects

- The purpose of this field is to prompt you to think about the basic elements of treatment effects.
- Other fields in Stage 2 describe the factors that identify treatment effects.
- Not every field in Stage 2 will be asked: Single specification or period.
- Your response to this field should match the total number of treatment effects generated by the rest of Stage 2.

Key

- The survey form uses your response to Field 1 to later check your responses.
- If your answer does not match this number, an **Error** message will appear.

Field 2: Outcome-Contrast Pair

- Asked for all exhibits.

Variable name in SurveyCTO	[contrast_col]
Field name (LABEL on SurveyCTO)	Outcome-contrast pairs reported for treatment effects.
Definition (LABEL on SurveyCTO)	The contrasts for which treatment effects are reported in the table for each outcome.
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select multiple Mark as checked if the pre-specified pair of Outcomes and Contrasts is present in the exhibit.
SurveyCTO repeat level	Display all outcomes and contrasts identified for the specified table in Stage 1 for mapping.

Field: Outcome-Contrast Pair

Example 1: Riley (2024)

- **Outcome-Contrast pairs**: 5

TABLE 1—TREATMENT EFFECTS ON WOMAN'S BUSINESS PROFITS, SAVINGS, AND BUSINESS CAPITAL

	Profit (1)	Savings (2)	Capital (3)
Mobile Account	☑	☑	☑
Mobile Disburse	☑	☑	

2 IDEAL-preferred Specification

Field 2: Estimand

- Coding protocol:

Variable name in SurveyCTO	[estimand_col]
Field name (LABEL on SurveyCTO)	Estimand
Definition (LABEL on SurveyCTO)	Estimand for the outcome-contrast pairs reported for treatment effects in the table.
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select one -ITT -TOT or LATE
SurveyCTO instructions	Repeats for each outcome-contrast pair; If multiple estimands in exhibit, display CV alternatives; If a single estimand, display label specified in Stage 1.

Field 2: Estimand

Example 1: Riley (2024)

- Outcome-Contrast pairs: 5
- **Estimand**: ITT

Notes:

-The survey form collects either: ITT or TOT/LATE per exhibit, but not both.

-These terms are primarily used in Economics. Other fields may reference similar estimates differently or may not mention them at all.

TABLE 1—TREATMENT EFFECTS ON WOMAN'S BUSINESS PROFITS, SAVINGS, AND BUSINESS CAPITAL

	Profit (1)	Savings (2)	Capital (3)
Mobile Account	☑	☑	☑
Mobile Disburse	☑	☑	

Notes: Intent-to-treat estimates; monetary outcomes are winsorized at the 99 percent level and in US dollars. All regressions include strata dummies and include the baseline value of the outcome. Mobile Account is the treatment where only a mobile money account was provided and the loan was disbursed as cash. Mobile Disburse is the treatment where a mobile money account was provided and the loan also disbursed onto this account. "Profit" refers to the woman's self-reported monthly business profit. "Savings" is individual savings held by the woman. "Capital" is the value of all assets the woman uses in her business plus the value of inventory held for her business. Control mean endline is the mean value of the outcome in the control group at endline. Control mean baseline is the mean value of the outcome in the control group at baseline. False discovery rate (FDR) adjusted *p*-values, also known as *q*-values, were used to correct for multiple hypothesis testing. They are shown in curly brackets. These were calculated following the method of Benjamini, Krieger, and Yekutieli (2006). Robust *p*-values in square brackets. Robust standard errors in parentheses.

Field 3: IDEAL-preferred specification

- Coding protocol:

Variable name in SurveyCTO	[specification_col]
Field name (LABEL on SurveyCTO)	Outcome-Contrast Pairs
Definition (LABEL on SurveyCTO)	IDEAL preferred empirical specifications for the outcome-contrast pairs reported for treatment effects in the table.
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select multiple -No controls or any other variables -Strata fixed effects. -Baseline value of outcome. -Controls that were either observed before treatment assignment or considered static. -Controls that were either observed after treatment assignment or are not considered static.

Field 3: IDEAL-preferred specification

Ranking:

#	Strata correction	Baseline outcome	Other controls	No other controls
1	✓	✓		
2	✓			✓
3	✓	✓	✓	
4	✓		✓	
5		✓		
6				✓
7		✓	✓	
8			✓	

Field 3: IDEAL-preferred specification

Example 1: Riley (2024)

- Outcome-Contrast pairs: 6
- Estimand: ITT

- **Specification for all O-C pairs**: Strata Fixed Effects + Baseline value of outcome

Notes:

-At least one CV option must be selected.

-If other options are chosen, "No controls or any other variables" cannot be selected.

TABLE 1—TREATMENT EFFECTS ON WOMAN'S BUSINESS PROFITS, SAVINGS, AND BUSINESS CAPITAL

	Profit (1)	Savings (2)	Capital (3)
Mobile Account	☑	☑	☑
Mobile Disburse	☑	☑	

Notes: Intent-to-treat estimates; monetary outcomes are winsorized at the 99 percent level and in US dollars. All regressions include strata dummies and include the baseline value of the outcome. Mobile Account is the treatment where only a mobile money account was provided and the loan was disbursed as cash. Mobile Disburse is the treatment where a mobile money account was provided and the loan also disbursed onto this account. "Profit" refers to the woman's self-reported monthly business profit. "Savings" is individual savings held by the woman. "Capital" is the value of all assets the woman uses in her business plus the value of inventory held for her business. Control mean endline is the mean value of the outcome in the control group at endline. Control mean baseline is the mean value of the outcome in the control group at baseline. False discovery rate (FDR) adjusted p-values, also known as q-values, were used to correct for multiple hypothesis testing. They are shown in curly brackets. These were calculated following the method of Benjamini, Krieger, and Yekutieli (2006). Robust p-values in square brackets. Robust standard errors in parentheses.

3 Author-preferred Specification

Field 4: Author-preferred specification

- Coding protocol:

Variable name in SurveyCTO	[preferred_col]
Field name (LABEL on SurveyCTO)	Is the IDEAL-preferred specification also preferred by the author(s)?
Definition (LABEL on SurveyCTO)	Indicates whether authors preferred the same empirical specifications as IDEAL for the outcome-contrast pairs.
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select one *-Check for Yes: IDEAL-preferred is the same as Author-preferred.* *-Leave unchecked for No: Both specifications are different.*
SurveyCTO instructions	Repeat for each outcome-contrast pair (conditional)

Field 4: Author-preferred specification

- A **directly preferred** estimate would be explicitly flagged by authors as a "preferred" estimates, “main” specification or model, or similar.

  -> If authors only report one estimate.

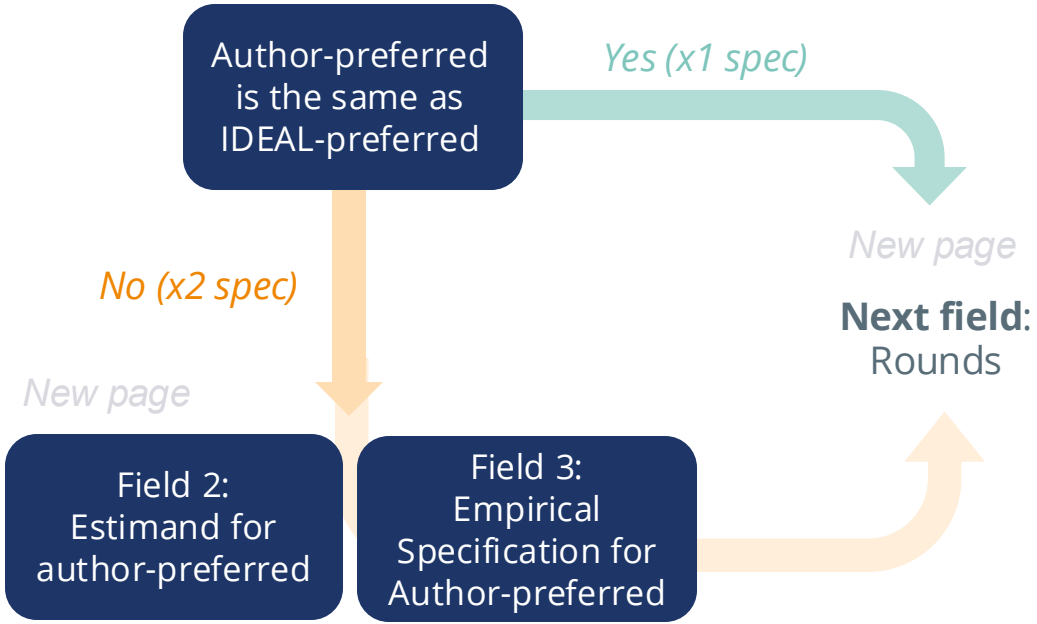
- Estimates are **indirectly preferred** when multiple estimates from different specifications reported in the paper, but the authors choose one to report in the abstract, introduction, or conclusion.

- If authors consistently report a range of estimates from different specifications without indicating a preference, they **do not have a preferred estimate.**

  -> We collect the IDEAL-preferred specification based on the ranking.

Field 4: Author-preferred specification

- Skip logic:



Field 4: Author-preferred specification

- Example: Ozler et al., (2018)

Table 3
Impacts on child assessments - 18-month follow-up.

	Dependent Variable:									
	Attending CBCC: 2012-13		Enrolled in Primary: 2012-13		Malawi Developmental Assessment Tool Score					
					Total		Language Skills		Fine Motor/ Perception Skills	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
T2 (teacher training)	0.084**	0.090***	−0.105***	−0.109***	−0.092	−0.063	−0.041	−0.031	−0.134*	−0.107*
	(0.034)	(0.030)	(0.034)	(0.029)	(0.080)	(0.060)	(0.079)	(0.063)	(0.074)	(0.059)
T3 (T2 + teacher incentives)	0.085**	0.097***	−0.086**	−0.097***	0.013	−0.003	0.088	0.085	−0.081	−0.115*
	(0.035)	(0.031)	(0.034)	(0.031)	(0.081)	(0.067)	(0.084)	(0.072)	(0.075)	(0.065)
T4 (T2 + parenting training)	0.065**	0.078***	−0.064**	−0.075***	0.115	0.126*	0.183**	0.185**	0.008	0.012
	(0.031)	(0.028)	(0.031)	(0.027)	(0.086)	(0.067)	(0.087)	(0.071)	(0.075)	(0.061)
Lagged Dependent Variable (Baseline)		−0.092***		0.090***		0.510***		0.426***		0.444***
		(0.015)		(0.015)		(0.034)		(0.031)		(0.034)
Any Treatment (T2, T3, or T4) - **separate regression**	0.077***	0.088***	−0.083***	−0.092***	0.019	0.027	0.084	0.087	−0.064	−0.065
	(0.027)	(0.024)	(0.027)	(0.023)	(0.071)	(0.055)	(0.071)	(0.058)	(0.065)	(0.051)
Mean and Standard Deviation of dependent variable in the control group	0.695		0.297		0.000		0.000		0.000	
	(0.461)		(0.458)		(1.000)		(1.000)		(1.000)	
F-test for Equality of Parameters (p-value) T2 = T3	0.974	0.843	0.586	0.701	0.191	0.367	0.131	0.106	0.454	0.889
T2 = T4	0.567	0.687	0.234	0.276	0.007***	0.002***	0.005***	0.001***	0.033**	0.049**
T3 = T4	0.546	0.543	0.503	0.478	0.195	0.058*	0.266	0.168	0.199	0.045**
District-bin Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dependent Variable and Age Dummies?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of observations	1925	1925	1925	1925	1936	1936	1936	1936	1936	1936

Notes: 0.01 - ***; 0.05 - **; 0.1 - *; OLS regressions at the child level using standardized test scores at the 18-month follow-up and baseline covariates with standard errors (SEs) in parentheses. Child assessments at the 18-month follow-up were conducted at the end of the 2012-13 school year. The covariate adjustment referred to as the 'lagged dependent variable' is the baseline value of the dependent variable, except in columns (1)–(4), where it is the baseline value of the Malawi Developmental Assessment Tool: Total Score. SEs are clustered at the CBCC level and observations are weighted using sampling weights and tracking weights (for 42 observations randomly assigned to tracking). 'Any treatment (T2, T3, or T4)' refers to a separate OLS regression, presenting the average impact of being in one of the three treatment arms in comparsion with the control group.

4 Rounds of data collection

Field 5: Rounds of data collection

- Coding protocol:

Variable name in SurveyCTO	[period_col]
Field name (LABEL on SurveyCTO)	Rounds of data collection
Definition (LABEL on SurveyCTO)	Rounds of data collection used to estimate the specific treatment effect
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select multiple -Select from pre-populated list of rounds of data collection (from Stage 1)
SurveyCTO instructions	Repeats for each outcome-contrast pair; Only appears if there is more the one round; Survey allows for alternative specification for each round.

Field 5: Rounds of data collection

- Example: Yoshikawa et al. (2015)

Table 1
Impact of the UBC Intervention on Classroom Quality Outcomes at the End of Prekindergarten and Posttest

Classroom quality outcome	End of prekindergarten impact						Posttest impact					
	N	Adj. mean full UBC group	Adj. mean comparison group	Diff.	Sig.	Effect size	*N*	Adj. mean full UBC group	Adj. mean comparison group	Diff.	Sig.	Effect size
Emotional support	76	4.810	4.420	0.390	***	0.814	72	4.831	4.566	0.265	*	0.376
Instructional support	76	1.804	1.605	0.199	**	0.438	72	1.842	1.813	0.029		0.056
Classroom organization	76	4.435	4.131	0.304	*	0.459	72	4.340	3.990	0.341	*	0.432

Note. Adj = adjusted mean for covariates; Diff = difference between adjusted mean in full UBC group and adjusted mean in comparison group; this is equivalent to parameter estimate (beta).

-> 3 Outcome-Contrast pairs
-> 2 rounds of data collection used for estimation
-> 1 specification for both

Number of treatment effects: 6

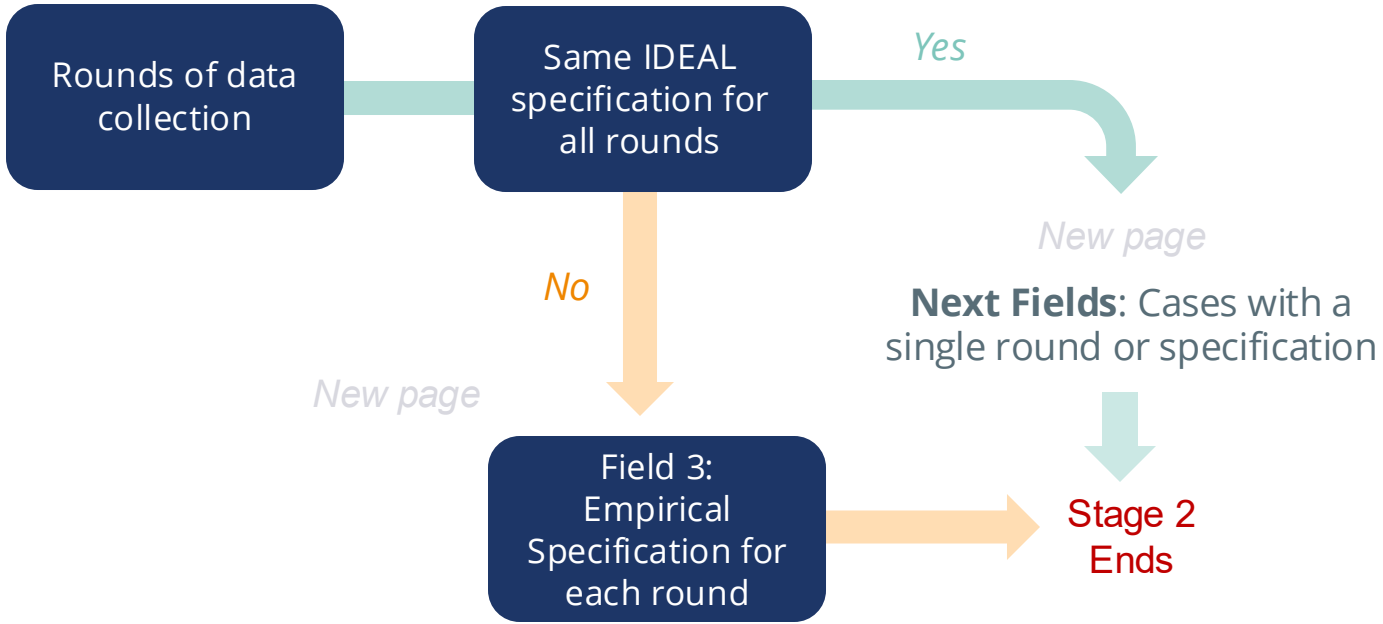
Field 6: Same specification for all rounds

- Coding protocol:

Variable name in SurveyCTO	[same_period_col]
Field name (LABEL on SurveyCTO)	Does the IDEAL-preferred specification apply to all data collection rounds selected above?
Definition (LABEL on SurveyCTO)	Specifies whether the IDEAL-preferred specification is used to estimate the treatment effect in each of the selected rounds of data collection.
Response options (open-text, numeric, date, text-CV)	Text-CV
CV (Choices on SurveyCTO)	Select one - Yes (check): IDEAL-preferred applies to all rounds. - No (unchecked): At least one of the rounds of data collection is used in a different empirical specification.
SurveyCTO instructions	Repeat for each outcome-contrast pair (conditional); Only appears if there is more than one specification.

Field 6: Same specification for all rounds

- Skip logic:



Thank you for listening

Sergio Puerto
sergiopuerto@berkeley.edu