



**2020 Education Innovation and Research (EIR) Project Directors
and Evaluators Technical Assistance Meeting**

Virtual Meeting | October 21–23, 2020

Recovering from COVID-19 in 2021: Modifying Evaluation Designs

Cristofer Price
Abt Associates



Cris Price



Role: TA Leader & TA Liaison



Background: Cris is a Principal Associate in Abt Associates' Social and Economic Policy division, with more than 30 years of experience in behavioral and educational research. He is an evaluation TA team leader for the EIR and i3 evaluation technical assistance projects. He has been the technical lead on a wide variety of projects representing a multitude of different types of study designs, data elements, units of analysis, and analytical methods.



Agenda

- Overview of the impact study designs in EIR and how they were affected by the COVID school closures
 - Focus on the spring 2020 school shutdowns
 - Too early to know how COVID will affect the 2020-21 school year and the evaluations
- Potential adaptations to impact study designs resulting from the spring 2020 COVID shutdowns
 - Some issues cannot be resolved
- Questions and comments

Overview of EIR Evaluation Designs and the Effect of COVID

Types of Impact Study Designs in EIR (But No Two Impact Evaluations Are the Same)

Study Design	District	School	Teacher	Student	Total
Randomized control trial (RCT)		23	8	10	41
Quasi-experimental design (QED)	1	14	1	6	22
Comparative short interrupted time series (CSITS)		15			15

How the COVID School Closures Affected EIR Impact Evaluations

1. Lost a baseline and an outcome measure
2. Lost an outcome, but not a baseline measure
3. Lost a baseline, but not an outcome measure
4. Baseline or outcome from 2020 was measurable, but expected to be quite different from prior years
5. Interruption or change in treatment – because spring of 2020 was between the planned baseline and outcome measurements, data collection was not affected
6. No effect of shutdowns on plans – because intervention was not being implemented in school year 2019-20 and no data collection was planned for spring of 2020

Examples of Lost Measures and Those Still Available

Lost Measures

- Achievement tests
- Surveys
- Assessments of students' social-emotional learning

Available Measures

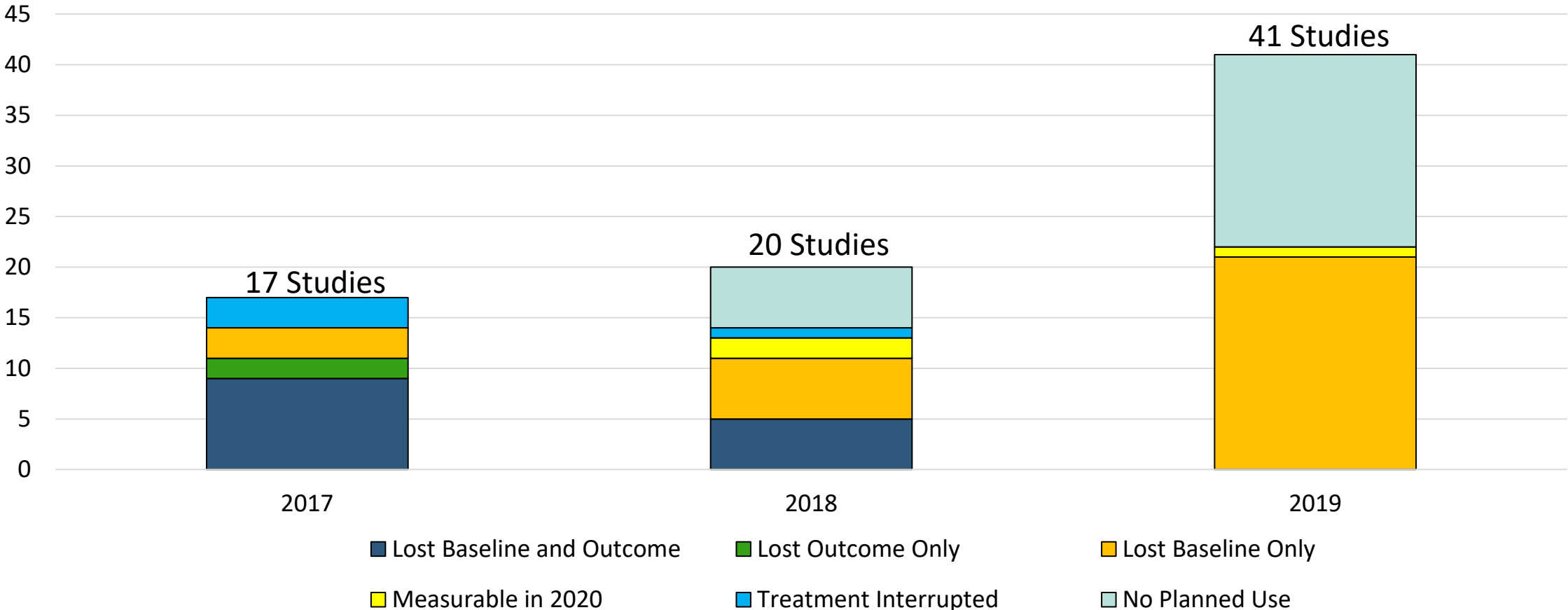
- Attendance
- Office discipline referrals
- Suspensions
- Dropout
- Graduation
- Taking or passing an Advanced Placement exam

Effect of COVID School Shutdowns on Evaluation Quality

- Loss of outcome measure from spring 2020 could result in:
 - Loss of power to detect effects
 - Inability to answer a research question
- Loss of baseline measure from spring 2020 could result in:
 - Loss of power to detect effects
 - Inability to establish baseline equivalence
- Interruption in implementation of the intervention or major change in implementation model could result in:
 - Loss of power to detect effects
 - Difficulty with interpreting results

Summary of the COVID Shutdown Effects by Cohort

How COVID School Shutdowns Affected Grants



Group 1: Loss of Baseline and Outcome Measures

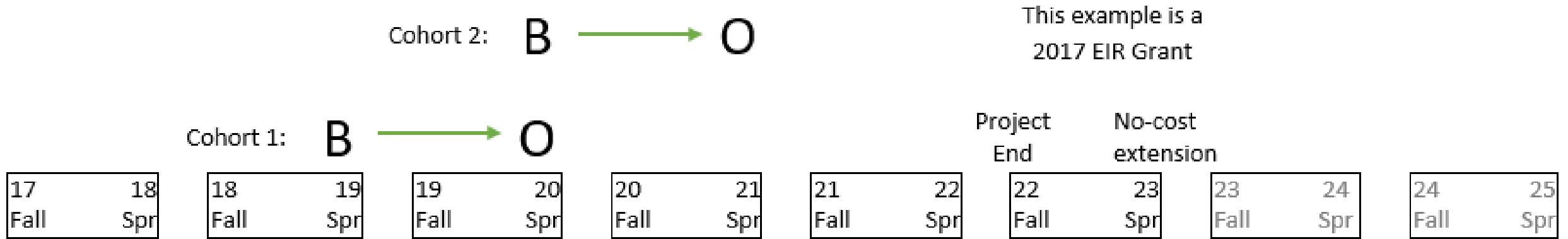
Spring 2020 school shutdowns caused:

Loss of Baseline measure for one cohort

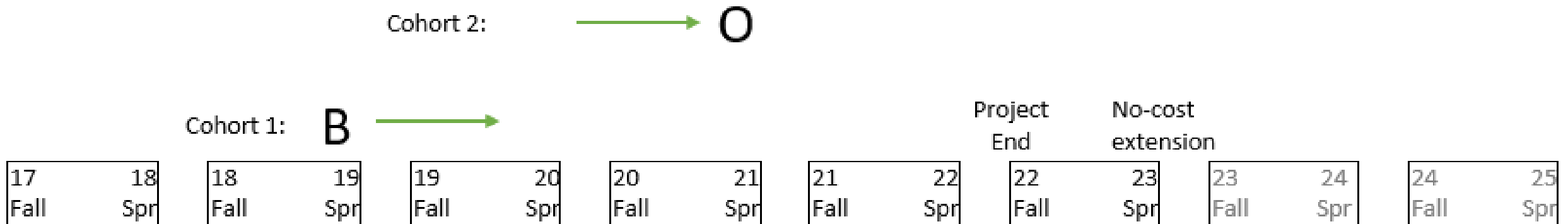
Loss of Outcome measure for another cohort

Group 1: Loss of Baseline and Outcome Due to COVID

Plan prior to COVID

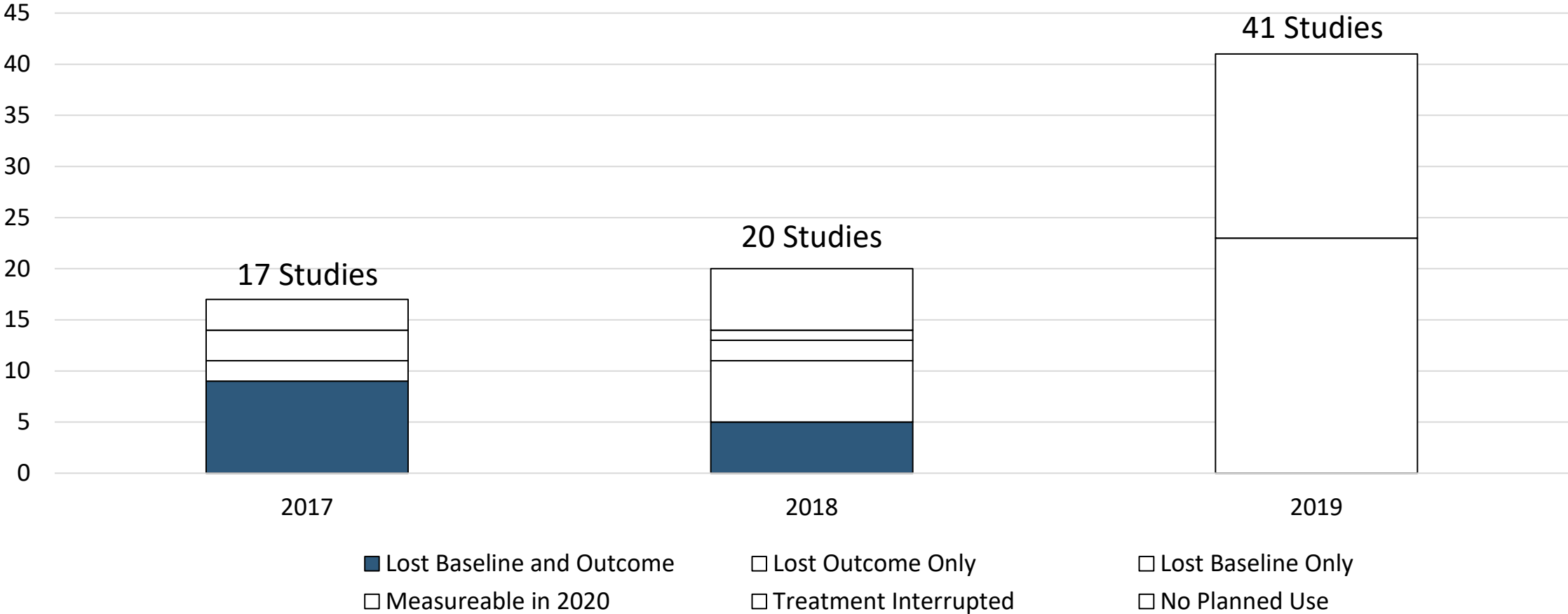


Due to COVID, no assessment data from spring of 2020



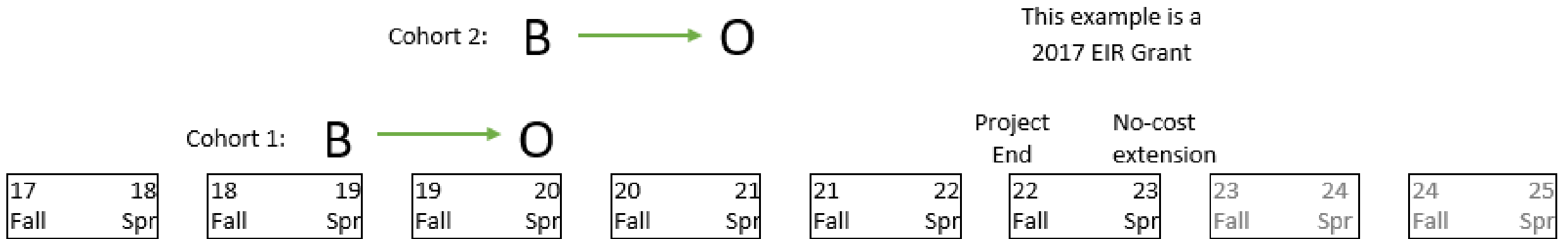
Over Half of 2017 Studies Lost a Baseline and an Outcome

Lost a Baseline Measure and an Outcome Measure

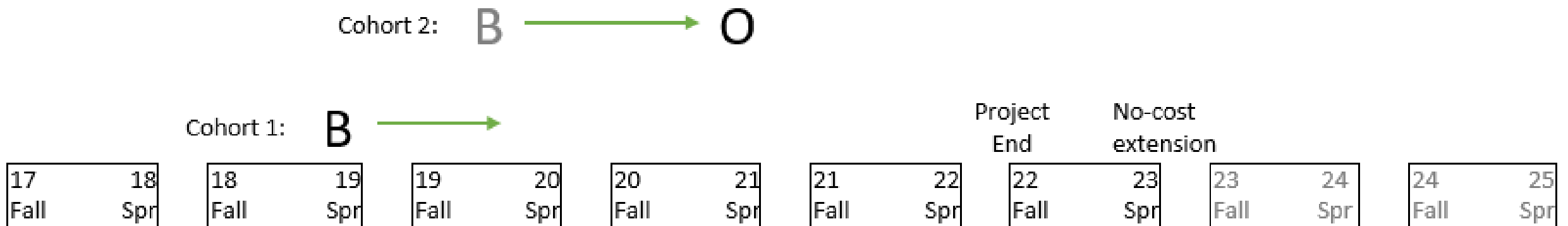


Group 1: Example Adaptation – Replace Lost End-of-Year Measure with a Mid-Year Measure

Plan prior to COVID

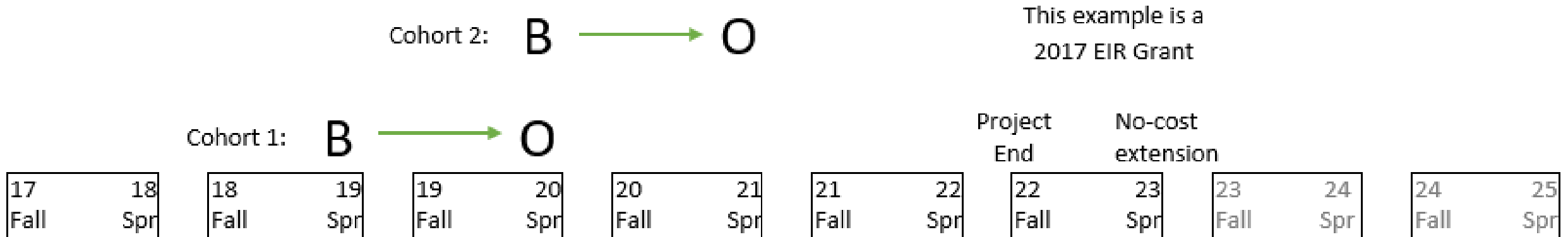


Adaptation: Cohort 2 baseline comes from earlier, mid-year assessment; size of sample increased for cohort 2

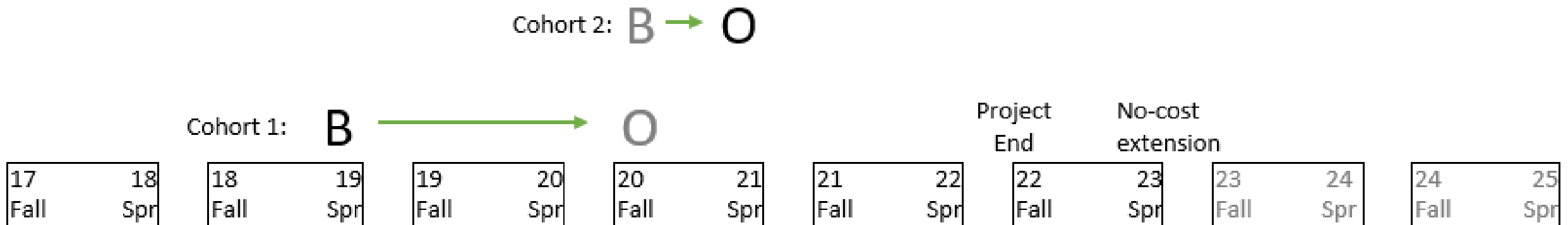


Group 1: Alternative Adaptation – Replace Lost End-of-Year Measure with a Fall Measure

Plan prior to COVID



Adaptation: Evaluator administers assessments in fall of 2020



Group 2: Loss of Outcome Measure

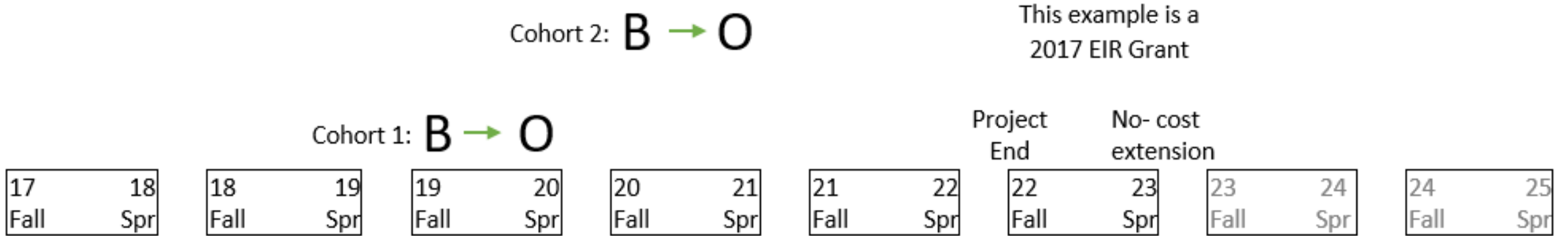
Spring 2020 school shutdowns caused:

Loss of **O**utcome measure for one cohort

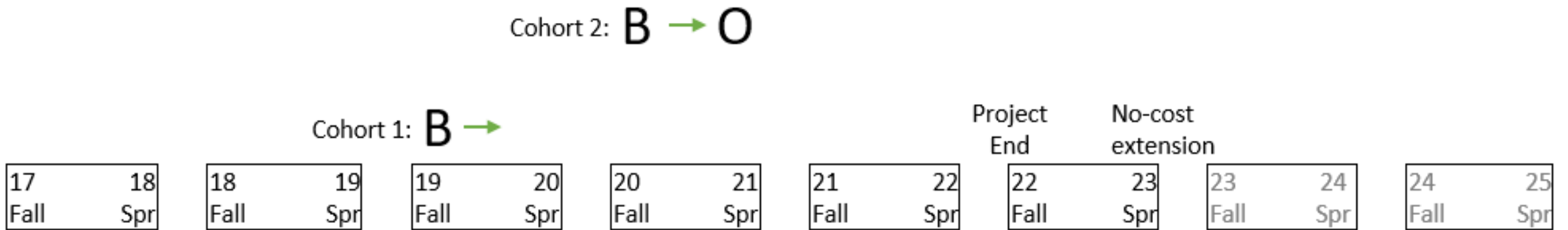
Change in intervention model for subsequent cohorts

Group 2: Loss of Outcome (Only) Due to COVID

Plan prior to COVID



Due to COVID, no assessment data from spring of 2020



Group 2: None of the 2018 or 2019 Studies Lost Only an Outcome

Lost an Outcome Measure



Group 2: Example Adaptation – Add Another Cohort

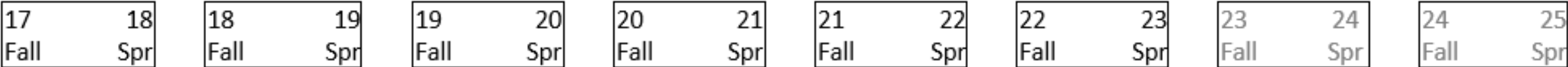
Plan prior to COVID

Cohort 2: **B** → **O**

This example is a 2017 EIR Grant

Cohort 1: **B** → **O**

Project End No-cost extension



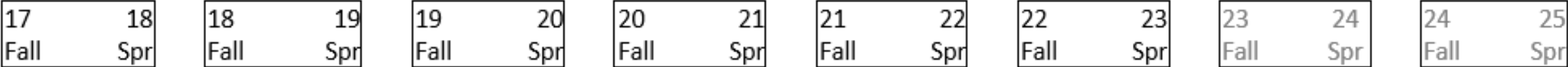
Adaptation: Add another cohort

Cohort 2: **B** → **O**

Cohort 1: **B** →

Cohort 3: **B** → **O**

Project End No-cost extension



Group 3: Loss of Baseline Measure

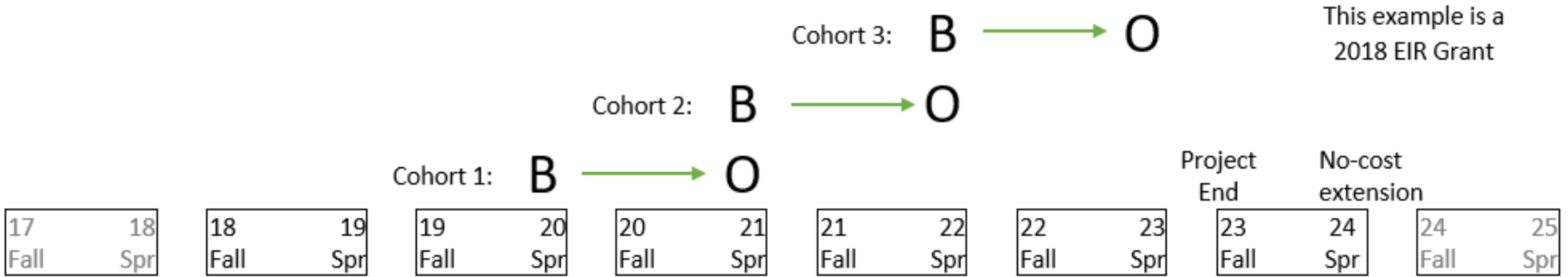
Spring 2020 school shutdowns caused:

Loss of **B**aseline measure for one cohort

Change in intervention model for subsequent cohorts

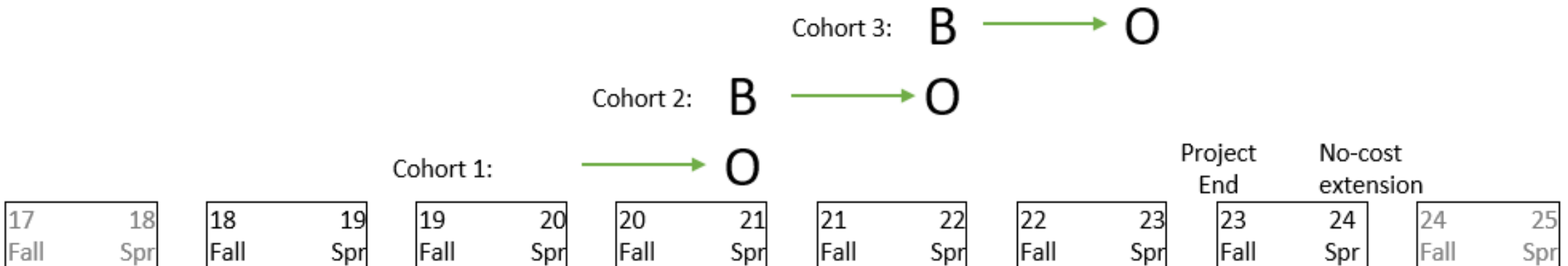
Group 3: Loss of Baseline (Only) Due to COVID

Plan prior to COVID



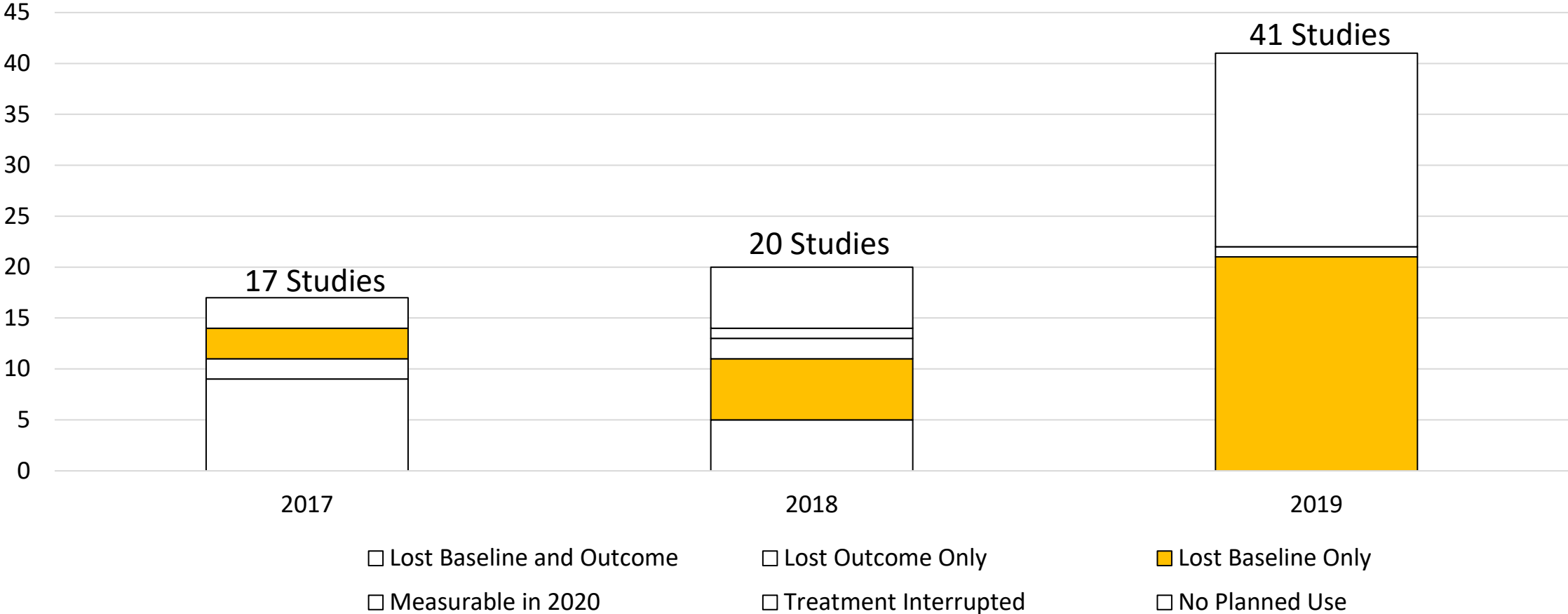
This example is a 2018 EIR Grant

Due to COVID, no baseline assessment data from spring of 2020



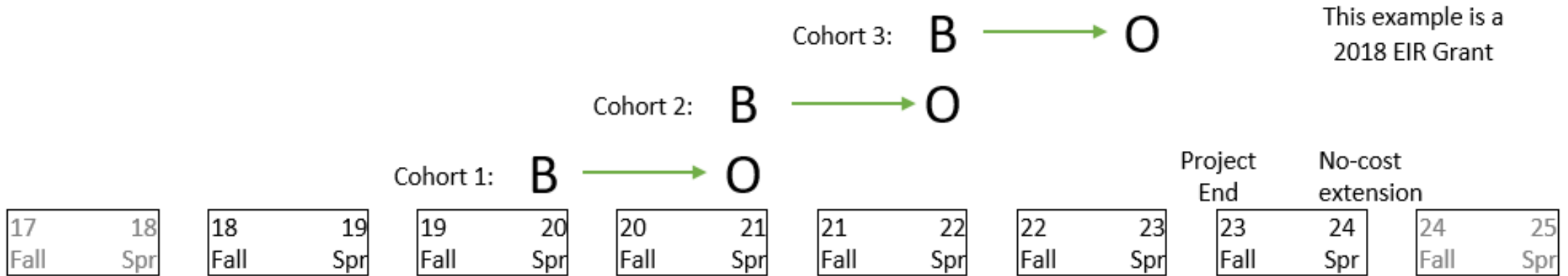
Group 3: Half of the 2019 Studies Lost Only a Baseline

Lost a Baseline Measure

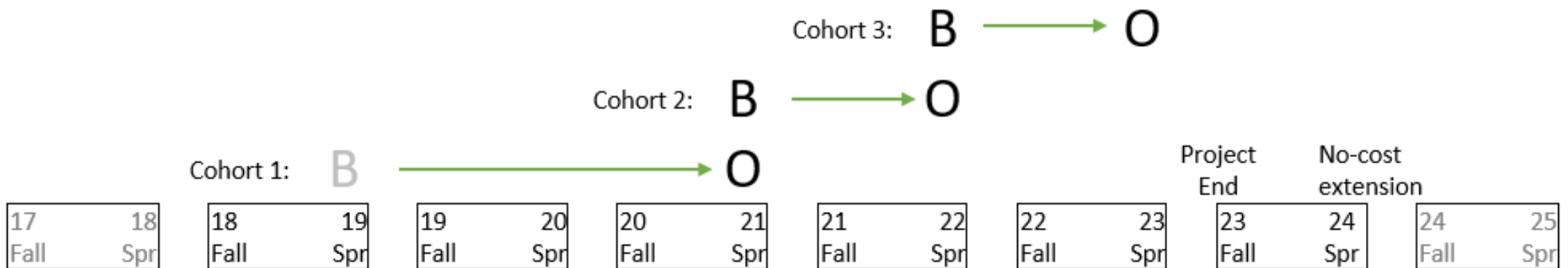


Group 3: Example Adaptation – Get Baseline from One Year Prior to Lost Year

Plan prior to COVID



Adaptation: Cohort 1 baseline comes from a year prior to the start of treatment



Group 3: Issues with Using an Older Baseline Measure (From spring 2019 Instead of Spring 2020)

- The older baseline will likely have a lower correlation with the outcome resulting in lower power
- If original plan was to use a 3rd grade baseline, and adaptation means going back one year to get 2nd grade baseline, most states don't have English language arts (ELA) and math assessments in 2nd grade.
- If baseline comes from a whole year prior to the start of the intervention, and if the plan was to show that treatment and comparison...
 - ... **students** were equivalent at baseline, the study can still meet WWC standards
 - ... **schools** were equivalent at baseline, it is not clear that the study can still meet WWC standards
 - (More on this to follow)

What the WWC Says About Baseline Equivalence^(a)

- “Any of the following three samples can be used to satisfy the baseline equivalence requirement for the analytic sample of clusters...

(a) The analytic sample of individuals from any pre-intervention time period.

(b) Individuals from the same cohort as the individuals in the analytic sample, within the same clusters at the time that clusters were assigned to conditions or up to one year prior to when clusters were assigned to conditions.

(c) Individuals from the previous (adjacent) cohort and in the same grade as individuals in the analytic sample, within the same clusters.”

(a) This language is from the WWC Review of Individual Studies Protocol

Can the Studies in Group 3 Meet WWC Standards Using a 2019 Baseline?

Grant Cohort	RCT	Traditional QED Baseline Equivalence of Students	Traditional QED Baseline Equivalence of Schools/Districts	CSITS Baseline Equivalence of Schools
2017		1		2
2018	2	2		2
2019	9	5	1	6
Can pass WWC standards using a spring of 2019 baseline?	Yes	Yes	Maybe (But this study is delaying 1 year such that it will have a 2021 baseline)	Probably not (But at least 3 of the 6 studies for 2019 are delaying 1 year such that they will have a 2021 baseline)

Group 3: Example of Typical CSITS Design in EIR

Research Question: Does the intervention have an effect on school mean 5th grade ELA scores?

Plan Prior to COVID:

Use school mean 5th grade ELA scores for 4 years prior to start of intervention and 3 years after

Assess baseline equivalence of schools using scores from spring of 2020; that is, using:

“(c) Individuals from the previous cohort and in the same grade as individuals in the analytic sample, within the same clusters.”

Assignment	Spring 2017	Spring 2018	Spring 2019	Spring 2020	Spring 2021	Spring 2022	Spring 2023	Number of Schools
Treatment	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	15
Comparison	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	5 th Gr	45

Unshaded cells are pre-treatment years

Shaded cells are treatment years

Group 3: Most EIR CSITS Studies Will Not Meet WWC Standard Using a 2019 Baseline Measure

Due to COVID, no assessment data from spring of 2020

Using spring of 2020 data would NOT be using:

“(c) Individuals from the previous cohort and in the same grade as individuals in the analytic sample, within the same clusters.”

Assignment	Spring 2017	Spring 2018	Spring 2019	Spring 2020	Spring 2021	Spring 2022	Spring 2023	Number of Schools
Treatment	5 th Gr	5 th Gr	5 th Gr		5 th Gr	5 th Gr	5 th Gr	15
Comparison	5 th Gr	5 th Gr	5 th Gr		5 th Gr	5 th Gr	5 th Gr	45

What about using...

“(b) Individuals from the same cohort as the individuals in the analytic sample, within the same clusters at the time that clusters were assigned to conditions or up to one year prior to when clusters were assigned to conditions.”

For 5th grade outcomes in spring of 2021, could potentially meet standards by showing baseline equivalence of that cohort back in 2019 when the students in that cohort were in 3rd grade.

But for 5th grade outcomes in spring of 2022, would need to show baseline equivalence of that cohort back in 2019 when the students in that cohort were in 2nd grade (but typically no state ELA assessments for 2nd graders!)

But for 5th grade outcomes in spring of 2023, would need to show baseline equivalence of that cohort back in 2019 when the students in that cohort were in 1st grade (but typically no state ELA assessments for 1st graders!)

Group 3: An Alternative Adaptation that Uses a Mid-Year Assessment to Establish Baseline Equivalence

Due to COVID, no assessment data from spring of 2020

Adaptation – Use District Assessment from January 2020 to show Baseline Equivalence, and Use that Measure as a Covariate, but Do Not Use that Measure in Time Series

Assignment	Spring 2017	Spring 2018	Spring 2019	Spring 2020	Spring 2021	Spring 2022	Spring 2023	Number of Schools
Treatment	5 th Gr	5 th Gr	5 th Gr		5 th Gr	5 th Gr	5 th Gr	15
Comparison	5 th Gr	5 th Gr	5 th Gr		5 th Gr	5 th Gr	5 th Gr	45

Let $5th\ Gr_{ij}$ represent an end-of-year state test score from the i^{th} year in the j^{th} school

Let $5th\ Gr_{January20}$ represent a district administered test score from January of 2020

Show baseline equivalence using $5th\ Gr_{January20}$

Use a CSITS impact model that has $5th\ Gr_{January20}$ as a right-hand-side covariate:

$$\begin{aligned}
 5th\ Gr_{ij} = & (\beta_0 + \alpha_{0j}^{Schs}) + (\beta_1 + \alpha_{1j}^{Schs})(Time_{ij}) + \beta_2(TSch_i) + \beta_3(TSch_i * Time_{ij}) \\
 & \beta_5(TYr1_{ij}) + \beta_6(TYr2_{ij}) + \beta_7(TYr3_{ij}) \\
 & \beta_9(TSch * TYr1_{ij}) + \beta_{10}(TSch * TYr2_{ij}) + \beta_{11}(TSch * TYr3_{ij}) \\
 & + \beta_4(5th\ GR_{January20}) + \epsilon_{iGj}^{Years}
 \end{aligned}$$

Group 4: Baseline or Outcome Measurable, but Different From Prior Years

Group 4: Baseline or Outcome From 2020 Was Measurable, but Expected to Be Quite Different From Prior Years

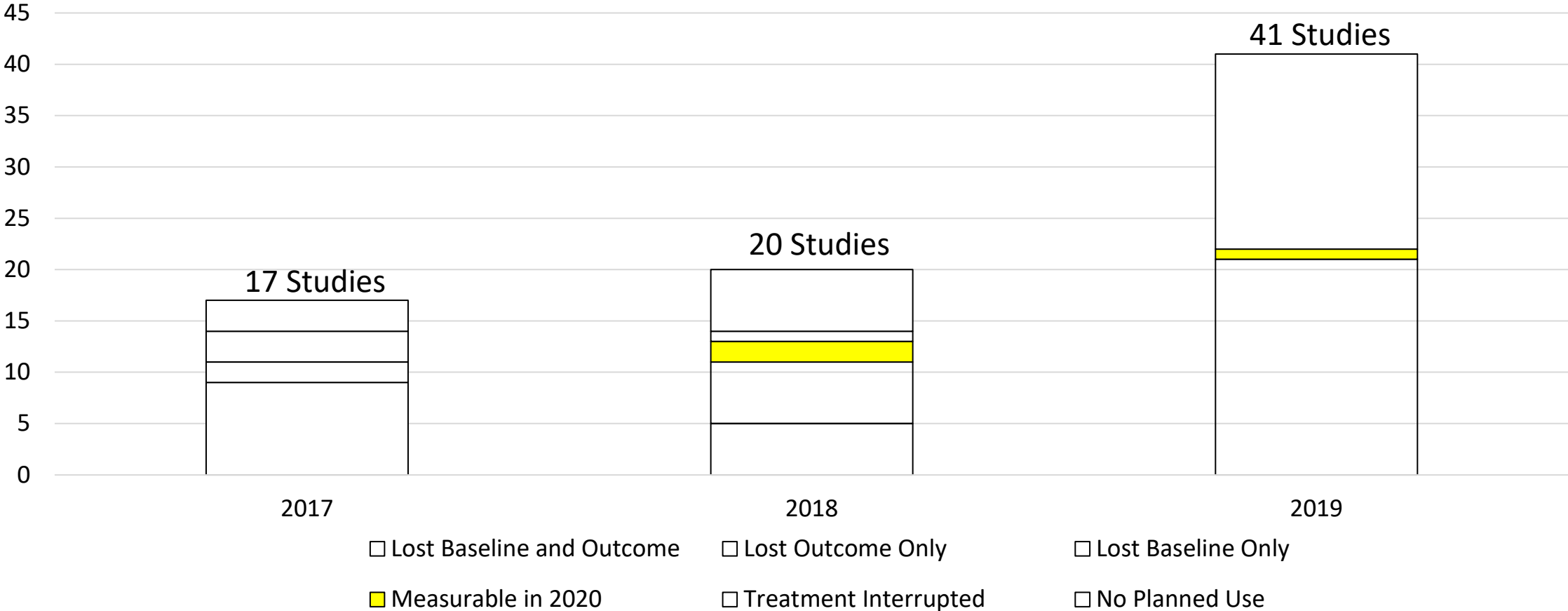
- These are studies that planned to measure a baseline or an outcome in spring of 2020, but all confirmatory outcomes were measurable
- Examples include measures of:
 - Attendance, office discipline referrals, suspensions
 - Dropout, graduation, matriculation to post-secondary
 - Taking or passing an Advanced Placement exam

Group 4: Example of How the 2020 Measure Was Different

- Advanced Placement (AP) exams were administered in the spring of 2020, but:
 - They were shorter than usual
 - Were taken from home
 - There were widely reported IT issues
- We would expect these issues to affect treatment and comparison groups in similar ways, but if spring of 2020 was a pre-treatment year and if...
 - ... it was to be used as a baseline in a traditional QED or RCT design, the correlation between baseline and outcome will likely be lower, which will affect power to detect effects
 - ... the measure was part of a time series (as in CSITS design), there will be an unusual score from that year, which will affect the slope of the baseline trend

Group 4: Studies Where the Baseline or Outcome was Measurable in Spring of 2020

Baseline or Outcome Was Measurable in Spring of 2020



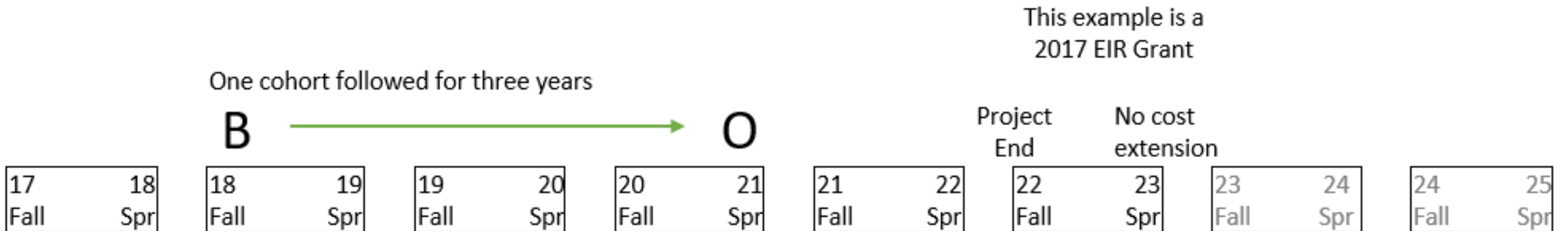
Group 4: Strategies and Adaptations – When 2020 Baseline Measure is Atypical and CSITS Design is Used

- 1) No adaptation – let the baseline trend be influenced by the unusual measure from 2020 under the assumption that the trends should be influenced similarly in treatment and comparison schools.
- 2) Measure the outcome (e.g., office disciplinary referral) only through March of every year, so that the 2020 measure won't be atypical.
- 3) Adapt the impact analysis model in a similar manner to the adaptation described previously, when the 2020 baseline was a mid-year district assessment and all other years were end-of-year state tests

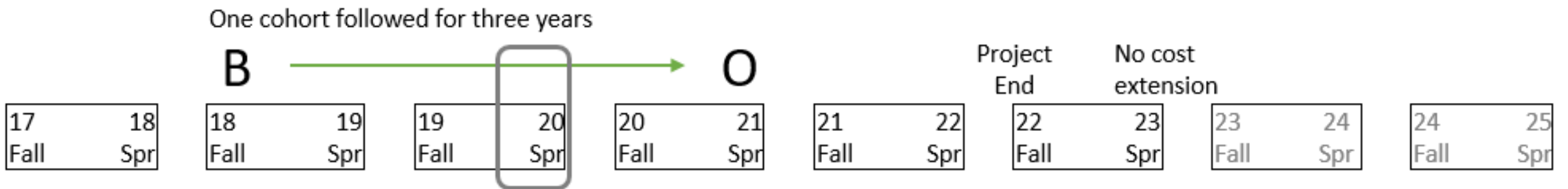
Group 5: Interruption or Change in Treatment Between Planned Baseline and Outcome Measurements

Group 5: Interruption or Change in Treatment Between Baseline and Outcome Due to COVID

Plan prior to COVID

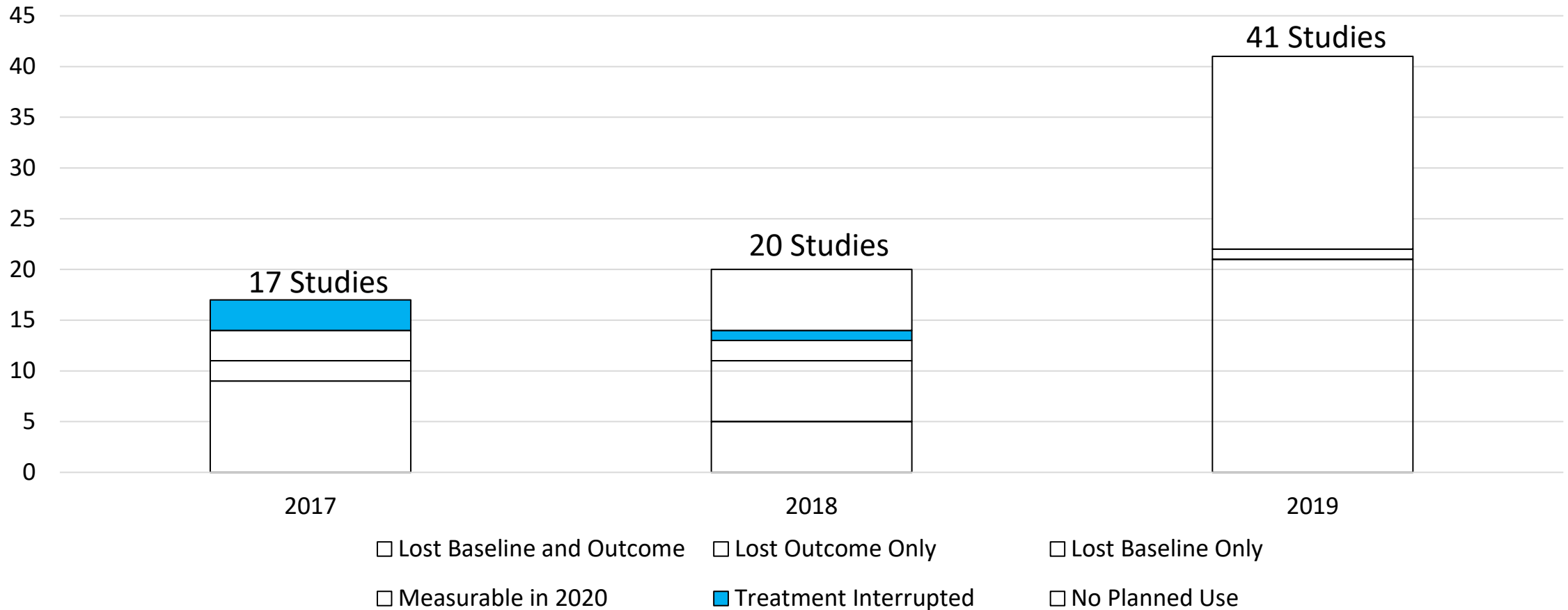


Due to COVID, no treatment for about 1/3rd of school year 2019-20



Group 5: Some Studies Had Treatment Interrupted, but No Planned Data Collection in Spring 2020

Treatment Interrupted, But No Planned Data Collection in Spring of 2020



Group 5: Strategies and Adaptations – When Treatment was Interrupted

- 1) No change in impact study design – Some of the studies in this category will not be changing the impact study design, but may change the measurement and interpretation of what it means to implement the intervention with fidelity after the spring 2020 shutdown and continuing disruptions. The interruption in treatment may also affect power to detect effects if the treatment is less effective due to the shortened 2019-20 school year.
- 2) Add a new cohort of students who will be measured only in the post-shutdown years
 - One grant is trying to raise funds to implement this strategy, but they cannot possibly implement this and finish the study within the current five year grant period.

Group 6: No Effect of School Shutdowns on Plans

Group 6: No Effects of School Shutdowns on Plans

- These are grants where the main, confirmatory impacts studies; had not yet begun in the 2019-20 school year
 - Some had pilot studies underway in 2019-20
 - Others never intended to measure baselines or outcomes from spring of 2020
 - And some that are in this group originally planned to collect baselines from spring of 2020, but decided to delay the start of their studies by one year because of the shutdowns and the uncertainty of how schools would operate in the 2020-21 school year.

Group 6: Almost Half to 2019 Studies Were Not Planning to Use Measures from Spring of 2020

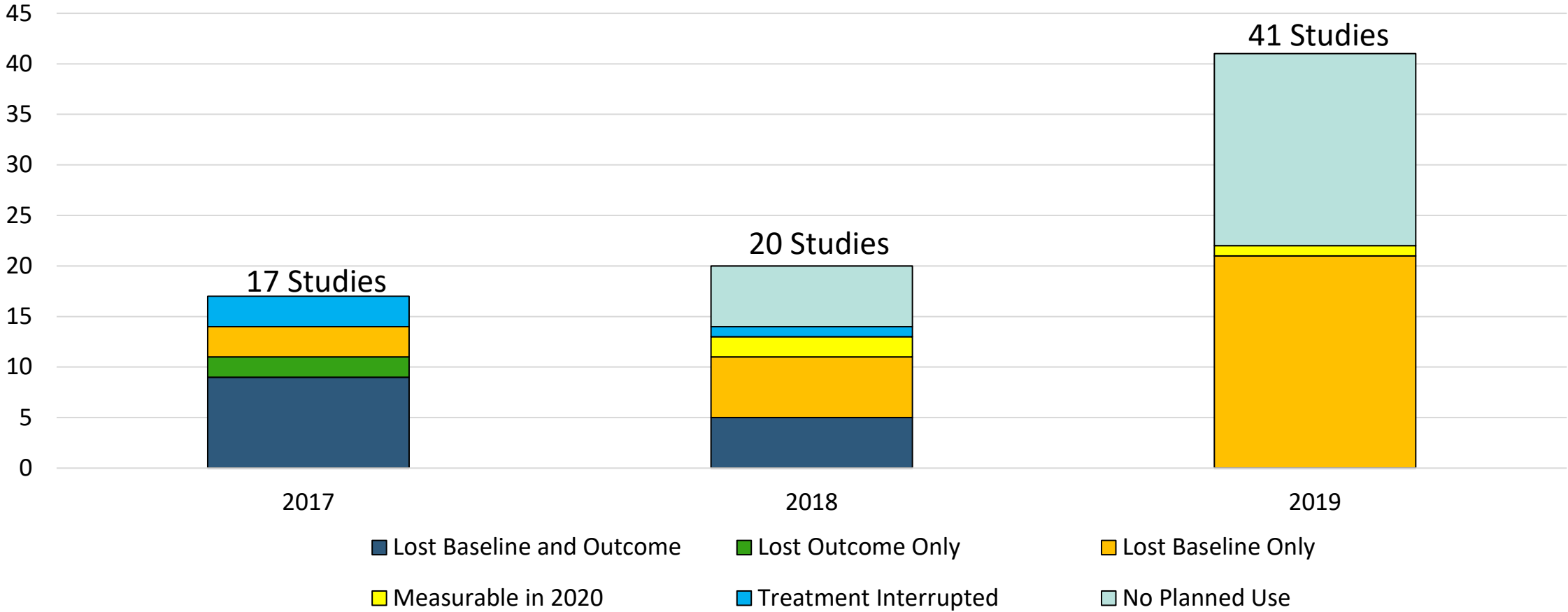
Not Implementing and Not Collecting Baseline or Outcome Data in Spring of 2020



Summary

Summary of the COVID Shutdown Effects by Cohort

How Spring 2020 School Shutdowns Affected Grants



Looking Ahead

- The spring 2020 school shutdown is just the beginning of the story of the effects of COVID on the EIR evaluations
 - All of the interventions that are implementing in the current 2020-21 school year will need to adapt to the mix of in-person, remote, and hybrid school climates this year
 - Intervention logic models may need to be adapted
 - What it means to implement with fidelity may need to change
 - The expected impacts of the altered interventions in the atypical school climates may be different than previous expectations
 - In some states, there may be no state assessments in spring of 2021

Contact

EIR Evaluation TA Team

EIREvalTA@AbtAssoc.com



BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT

www.abtassociates.com

Questions

