performing students benefited more from the program, as all interactions are negative, although all the parameters are quite imprecisely estimated. The results for SPS/South Africa indicate that high-performing students learned to read more words per minute than low- performing students from the program in both Eastern Cape and KwaZulu-Natal (just as in MA/Zambia), as the coefficients on the interactions between treatment and baseline quartiles are positive, increasing in the quartiles in absolute values, and actually statistically significant for the highest quartile.

Table 3. Treatment heterogeneity across baseline reading skills – Oral reading fluency (CWPM)

		READ CO/Ethiopia				SPS/South Africa	
		Amhara		Oromia			
	MA/Zambia	School-based	School + Community	School-based	School + Community	Eastern Cape	KwaZulu-Natal
Treatment	1.78**	6.32	5.25	5.34*	1.29	-0.72	-0.70
	(0.65)	(3.19)	(4.13)	(2.65)	(2.65)	(0.98)	(1.14)
Interactions between treatment status and baseline reading skills quartiles							
Quartile 2	1.08	5.02	0.58	-1.43	0.03	0.48	1.04
	(1.30)	(2.84)	(4.41)	(3.79)	(3.59)	(1.07)	(1.22)
Quartile 3	1.94	-0.63	-4.45	-0.49	2.63	1.50	2.64
	(1.18)	(3.70)	(4.64)	(4.18)	(4.63)	(1.12)	(1.43)
Quartile 4	1.77	-1.25	-3.66	-2.67	-1.00	2.90*	3.80*
	(1.67)	(3.30)	(4.40)	(3.52)	(3.79)	(1.33)	(1.85)
	1961	933	929	681	592	4110	4450

Note: The outcome of interest is the difference between endline and baseline for oral reading fluency. All regressions include dummies for baseline reading skills quartiles. Standard errors are clustered at the community level for SPS/South Africa and at the school level for all other experiments. Asterisks denote significance at 0.1% (***), 1% (**) and 5% (*) level.

The results for MA/Zambia and SPS/South Africa indicate that high-performing students learned to read more words per minute thanks to the program than low performing students. This, combined with the fact that performance at baseline is negatively correlated with attrition, suggests that Lee bounds overstate the impact of the program over the original baseline sample.

5. Conclusions

Survey attrition is inevitable in longitudinal studies. In this paper we present evidence that attrition implications in early reading interventions may be minor, at least according to the IPW and Lee bounds approaches. However, the assumption attached to the former, selection on observables, is strong and may not correctly describe the data generation process. The assumption inherent to Lee bounds, monotonicity, seems weaker but ultimately as untestable as IPW's. Manski bounds impose the least restrictions to the data generation process but fail to provide informative bounds across the analyzed experiments, even when binary outcomes were considered. Manski bounds will probably produce not informative bounds in other applications, unless overall attrition is very small, in which case it is probably inconsequential to begin with. Finally, Manski bounds derivatives, the Kling and Liebman bounds, are restrictive in the sense that they assume that mean outcomes of those that attrit will fall within essentially arbitrary ranges.

Aside from Manski bounds, Lee bounds are probably the most conservative method to address attrition in program evaluations. Therefore, we recommend that researchers consider this approach to bound the treatment effect in the presence of attrition.

However, in addition to assuming monotonicity, Lee bounds only document the impact on those always observed. This is problematic in the context of early reading interventions because the resulting bounds down-weight the impact on low performers, given that, as we showed, reading skills at baseline and attrition are negatively correlated. The extension we propose to Lee bounds documents the impact on the compliers too, but still leaves out of the analysis the impact on the never observed.

Researchers interested in documenting treatment impacts on low-performers should plan for this type of student to be more likely to attrit. To tackle this problem, one strategy is to oversample low performing students at baseline. This may be difficult because in most cases researchers only learn the reading skills of students after baseline data are collected, but information available ex-ante could be used to oversample low performing students, for example if there are test score data on standardized exams, even at the school level, researchers could oversample schools that underperform in these exams. Another strategy is to allocate additional resources to track down and assess at-home low-performers at endline that would otherwise attrite from the sample.[7] In addition, researchers should explore whether there is treatment effect heterogeneity across baseline reading skills (as well as other determinants of attrition). This exercise would inform whether Lee bounds may produce a biased estimate of the impact the program

[7] Molina and Macours (2019) propose methodologies to use intense tracking information to correct for selection bias.

would have on the entire original sample. The results for MA/Zambia and SPS/South Africa suggest that these programs benefited more high-performing students, which combined with the fact that reading skills are negatively correlated with attrition imply that Lee bounds may overestimate the impact the program would have on the original sample.

Finally, Lee bounds will probably fail to be informative when attrition rates are very different between treatment and control, which was not the case in most of the experiments examined in this paper. When treatment status is expected to affect school attendance (e.g., school feeding programs, conditional cash transfers), it would be recommendable that student surveys are conducted not only at school but at students' homes when needed, so differences in attrition rates by treatment status would be lower than if surveys were conducted at the schools.

References

1. Ahrens, Achim, Christian Hansen, and Mark Edwin Schaffer. (2018). LASSOPACK: Stata Module for LASSO, Square-root LASSO, Elastic Net, Ridge, Adaptive LASSO Estimation and Cross-validation.
2. Alderman, Harold and Donald Bundy. (2012). “School Feeding Programs and Development: Are We Framing the Question Correctly?” *The World Bank Research Observer* 27, no. 2: 204–21, https://doi-org.proxy.uchicago.edu/10.1093/wbro/lkr005
3. Aurino, Elisabetta, Aulo Gelli, Clement Adamba, Isaac Osei-Akoto, and Harold Alderman. (2020). “Food for Thought? Experimental Evidence on the Learning Impacts of a Large-Scale School Feeding Program.” *J. Human Resources* 1019-10515R1; published ahead of print December 14, 2020, doi:10.3368/jhr.58.3.1019-10515R1
4. Bando, Rosangela, Francisco Gallego, Paul Gertler, and David R. Fonseca. (2017). “Books or Laptops? The Effect of Shifting from Printed to Digital Delivery of Educational Content on Learning.” *Economics of Education Review* 61: 162-73.
5. Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2005): “Long-term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico.” IAI Discussion Papers, no. 122, Georg-August-Universität Göttingen, Ibero-America Institute for Economic Research (IAI), Göttingen.
6. Bundervoet, Tom. (2018). “Internal Migration in Ethiopia: Evidence from a Quantitative and Qualitative Research Study.” World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/32097 License: CC BY 3.0 IGO.
7. Cilliers, Jacobus, Brahm Fleisch, Cas Prinsloo, and Stephen Taylor. (2018). “How to Improve Teaching Practice? Experimental Comparison of Centralized Training and In-classroom Coaching.” RISE-WP-18/024
8. Drake, Lesley, Meena Fernandes, Elisabetta Aurino, Josephine Kiamba, Boitshepo Giyose, Carmen Burbano, Harold Alderman, Lu Mai, Arlene Mitchell, and Aulo Gelli. (2017). “School Feeding Programs in Middle Childhood and Adolescence.” In *Disease Control Priorities* 3, ed. D. Bundy, N. De Silva, S. Horton, D. Jamison, and G.C. Patton. Washington, D.C.: The World Bank. http://dcp-3.org/sites/default/files/chapters/DCP3CAHD_Ch 12.pdf
9. Doyle, Orla, Colm Harmon, James J. Heckman, Caitriona Logue, and Seong Hyeok Moo. (2016). “Early Skill Formation and the Efficiency of Parental Investment: A Randomized Controlled Trial of Home Visiting.” *Labour Economics*. http://dx.doi.org/10.1016/j.labeco.2016.11.002.
10. Fiszbein Ariel and Norbert R. Schady. (2009). “Conditional Cash Transfers: Reducing Present and Future Poverty.” World Bank Policy Research Report. Washington, DC: World Bank.
11. Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. (1998). “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics.” *The Journal of Human Resources* 33, no. 2: 251–99.
12. Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra. (2020). Testing Attrition Bias in Field Experiments. Working Papers 202010, University of California at Riverside, Department of Economics.
13. Glewwe, Paul and Karthik Muralidharan. (2016). “Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications.” In *Handbook of the Economics of Education* 5, 653–743. Amsterdam: Elsevier.
14. Grant, Monica J. and Jere R. Behrman. (2010). “Gender Gaps in Educational Attainment in Less Developed Countries.” *Population and Development Review* 36, no. 1, 71-89.
15. He, Fan, Leigh Linden, and Margaret MacLeod. (2009). “A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial.” Unpublished manuscript.

16. Heckman, James. J. (1979). “Sample Selection Bias as a Specification Error.” *Econometrica* 47, no. 1, 153–61.
17. Horowitz, Joel L. and Charles F. Manski. (2000). “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data.” *Journal of the American Statistical Association* 95: 77-84.
18. Huber, Martin and Giovanni Mellace. (2015). “Sharp Bounds on Causal Effects under Sample Selection.” *Oxford Bulletin of Economics and Statistics*, 77, 0305–9049. doi: 10.1111/obes.12056
19. Kling, Jeffrey R and Jeffrey B Liebman. (2004). “Experimental Analysis of Neighborhood Effects on Youth.” Unpublished manuscript.
20. Kotze, Janeli, Brahm Fleisch, and Stephen Taylor. (2019). “Alternative Forms of Early Grade Instructional Coaching: Emerging Evidence from Field Experiments in South Africa.” *International Journal of Educational Development* 66, 203-13.
21. Lee, David S. (2002). “Trimming for Bounds on Treatment Effects with Missing Outcomes.” NBER Working Paper, 0277.
22. Lee, David S. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *Review of Economic Studies* 76, 1071–1102.
23. Molina Millan, Teresa and Karen Macours. (2019). “Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias.” Unpublished manuscript.
24. Murnane R. J. and Alejandro J. Ganimian. (2014). “Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations.” NBER Working Paper, no. 20284, Inter-American Development Bank, Washington, DC.
25. NORC (2015). USAID/Ethiopia Reading for Ethiopia’s Achievement Developed Community Outreach (READ CO) Program, Evaluation Design Report.
26. NORC (2018). USAID/Ethiopia Impact Evaluation of Reading for Ethiopia’s Achievement Developed Community Outreach (READ CO) Program. Endline Evaluation Report.
27. Ome, Alejandro and Alicia Menendez. (2020). “Using SMS and Parental Outreach to Improve Early Grade Reading Skills in Zambia.” Unpublished Manuscript.
28. Parker, Susan, Luis Rubalcava, and Graciela Teruel. (2007). “Evaluating Conditional Schooling and Health Programs.” In *Handbook of Development Economics*, Volume 4, Chapter 62, 3963-4035. Elsevier.
29. RTI International. (2015). Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development.
30. Tibshirani, Robert. (1996). “Regression Shrinkage and Selection via the LASSO.” *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1: 267-88.
31. UNICEF 2019 https://data.unicef.org/topic/child-protection/child-labour/ accessed online on 12/03/20
32. Wooldridge, Jeffrey. M. (2002). *Econometric Analysis of Cross Section and Panel Data*.. Cambridge, MA, and London, The MIT Press.

Annex I. Non-linearities between attrition and baseline characteristics

	MA/Zambia	READ CO/Ethiopia				SPS/South Africa	
		Amhara		Oromia			
	Eastern province	School-based	School + Community	School-based	School + Community	Eastern Cape	KwaZulu-Natal
Treatment	-0.01	-0.02	-0.00	0.01	0.11*	0.03	0.01
	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)
Female	0.01	-0.05	-0.08**	-0.07**	-0.02	-0.01	0.01
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)
Reading score at baseline							
Quartile 2	-0.08**	-0.22***	-0.16***	-0.14**	-0.04	-0.04**	-0.04**
	(0.02)	(0.04)	(0.05)	(0.04)	(0.04)	(0.01)	(0.01)
Quartile 3	-0.02	-0.25***	-0.20***	-0.16**	-0.04	-0.02	-0.05**
	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)	(0.01)	(0.01)
Quartile 4 (highest score)	-0.07***	-0.32***	-0.27***	-0.34***	-0.23***	-0.05**	-0.06**
	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)
Asset index							
Quartile 2	-0.01	-0.05	-0.04	-0.00	-0.03	-0.02	-0.01
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)
Quartile 3	-0.00	-0.03	-0.07	-0.09	-0.04	-0.00	-0.00
	(0.02)	(0.04)	(0.04)	(0.05)	(0.04)	(0.02)	(0.01)
Quartile 4 (most assets)	0.04	-0.04	-0.06	-0.02	-0.03	0.01	-0.01
	(0.02)	(0.04)	(0.04)	(0.06)	(0.05)	(0.02)	(0.01)
Age groups[(a)]							
Group 1	-0.01	-0.05	-0.03	0.05	0.11**	-0.01	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.01)	(0.01)
Group 2	-0.01	-0.03	0.02	0.14**	0.14***	0.02	-0.02
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)
Group 3 (oldest)	0.02	0.08	0.12**	0.26***	0.31***	0.04*	0.01
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.02)	(0.02)
N	2253	1452	1473	1278	1247	4784	5331

[(a)]Age groups are different between experiments. For MA/Zambia the excluded category is 6-year-olds and the shown groups are 7-8-year-olds, 9-10 year olds, and 11-year-olds and older; for READ CO/Ethiopia the excluded category is 8- year-olds and younger, and the shown groups are 9-year-olds, 10-year-olds, and 11-year-olds and older; for SPS/South Africa the excluded category is 7-year-olds and younger, and the shown groups are 8-year-olds, 9-year-olds, and 0-year- olds and older.

Note: Marginal effects evaluated at the mean of the independent variables, after running logistic regressions. Regressions for MA/Zambia and SPS/South Africa include grade fixed effects (READ CO/Ethiopia is just one grade). Standard errors are clustered at the community level for SPS/South Africa and at the school level for all other experiments. Asterisks denote significance at 0.1% (***), 1% (**) and 5% (*) level.

Annex II. Model selection for IPW

Stepwise selection

The stepwise selection procedure follows Doyle et al. (2016). This approach is implemented in three phases: i) Estimate bivariate regressions between attrition and all potential predictors, and retain the ones that are statistically significant (p-value<=0.05); ii). An OLS regression is run using all retained predictors and the corresponding R^2 is recorded, then, to reduce further the number of covariates, multiple regressions are run to eliminate covariates iteratively using the adjusted R^2 as information criterion; iii). A logit model is estimated using the final set of covariates.

Below we list, for each data set, the variables retained in the final model, and the total number of variables considered.

MA/Zambia: Out of 27 variables considered, seven were retained: Two EGRA scores, an indicator for the child being first born, whether the household had planes to move at baseline, variables for household ownership of land plots and small livestock, and a district dummy.

READ CO/Ethiopia: 19 variables were considered across all for experiments. For school-based experiment in Amhara 12 variables were retained: Five EGRA scores and the first principal component of the six EGRA scores, child's age, gender, and interest in reading (self-reported), school attendance the previous week, if the child attended preschool, and household size. For the school + community intervention in Amhara, the same covariates were retained except that instead of school attendance a variable for whether Amharic is the language most spoken at home was retained. For school-based experiment in Oromia, seven variables were retained: One EGRA score and the first principal component of the six EGRA scores, child's age, school attendance the previous week, shift (morning, afternoon, all day), a dummy for whether Afaan Oromo is the language most spoken at home and a dummy variable for whether the child attended preschool. For the school + community intervention in Oromia the variables retained were: One EGRA score and the first principal component of the six EGRA scores, child's age, the treatment dummy and school attendance the previous week.

SPS/South Africa: In the Eastern Cape, eight variables out of 25 were retained - two EGRA scores and the first principal component of the four EGRA scores, a treatment indicator, child's height for age z-score, an indicator that the child co-resides with their mother and two district dummies. In KwaZulu-Natal, five variables out of 23 were retained – baseline oral reading fluency, child's age, child's height for age z-score, an indicator that the child co-resides with their mother and one district dummy.

LASSO

We use LASSO methods (Tibshirani 1996) to select the variables that we include in the attrition models for the IPW estimations. We follow Molina and Macours (2019) and model attrition for treatment and control groups separately, using the bias-corrected Akaike Information Criteria to select penalty levels. Then we estimate an attrition model pooling treatment and control groups, using all the covariates selected in either the treatment or control models, as well as their interactions with the treatment dummy. The estimated probabilities of this pooled model are used to construct the IPW for each experiment. Table B2 shows the IPW results using this method for the attrition model.

Table II1. IPW results using LASSO methods to select predictors

	MA/Zambia	READ CO/Ethiopia				SPS/South Africa	
		Amhara		Oromia			
	Eastern province	School-based	School + Community	School-based	School + Community	Eastern Cape	KwaZulu-Natal
Treatment	3.31***	7.29***	2.54	4.22	1.36	0.61	1.14
	(0.69)	(1.457)	(1.573)	(3.138)	(3.161)	(0.54)	(0.78)
N	1954	933	931	682	594	4109	4450

Note: The outcome of interest is the difference between endline and baseline for oral reading fluency. For predictor selection we use the Stata package LASSOPACK (Ahrens, Hansen, and Schaffer 2018). Standard errors are clustered at the community level for SPS/South Africa and at the school level for all other experiments. Asterisks denote significance at 0.1% (***), 1% (**) and 5% (*) level.

Annex III. Manski bounds for binary variables

Table III1. Manski bounds on the change between baseline and endline in the likelihood that students can read at least one word

		READ CO/Ethiopia				SPS/South Africa	
		Amhara		Oromia			
	MA/Zambia	School-based	School + Community	School-based	School + Community	Eastern Cape	KwaZulu-Natal
lower	-0.07*	-0.28***	-0.31***	-0.39***	-0.55***	-0.17***	-0.17***
	(0.03)	(0.05)	(0.06)	(0.04)	(0.05)	(0.02)	(0.02)
upper	0.19***	0.41***	0.40***	0.51***	0.45***	0.12***	0.16***
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.02)	(0.02)
N	2261	1452	1473	1384	1378	4976	5371

Note: The outcome of interest is the difference between endline and baseline for oral reading fluency. All regressions include dummies for baseline reading skills quartiles. Standard errors are clustered at the community level for SPS/South Africa and at the school level for all other experiments. Asterisks denote significance at 0.1% (***), 1% (**) and 5% (*) level.