SREE 2014

Abstract Title Page

Title:

What is the Optimal Length of an ELL Program?

Authors and Affiliations:

Guanglei Hong, Joshua Gagne, Andrew West

University of Chicago

Abstract Body

Background / Context:

Language minority students, defined as those who speak a non-English language at home, are a fast-growing subpopulation constituting 21% of the school-age children in the U.S. in 2009 (NCES, 2012). About 75% of these students speak Spanish and are immigrants from Central or South America (NCES, 2009). Many language minority students speak English with difficulty and therefore are lagging behind in their academic performance. The federal law holds districts and schools accountable for students' progress toward English language proficiency and achievement of academic standards. The government allocates funds to support specialized English language learning (ELL) services. An English language leaner is a student whose primary language is not English and whose level of proficiency in English is not sufficient to support learning in a regular English language classroom. Such students are entitled to support in the classroom until they achieve the level of English proficiency needed for full participation. The term "ELL services" encompasses English-as-a-second-language (ESL) programs, bilingual education programs, and other types of specialized programs for ELL students.

There have been intense debates about how many years of ELL services would be optimal for English language learners. Researchers have argued that, although it takes only several years of exposure to English for immigrant children to approach native-like levels in conversational skills, it takes four to nine years (Collier, 1987, 1989) or five to seven years (Cummins, 1981) to become proficient in academic English essential for learning in content areas (August & Hakuta, 1997).

According to the nationally representative Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K) data, however, in practice there has been vast variation across schools in the average years of ELL services provision, with three years being the mode. States and districts vary in how they initially identify students as ELLs and how they determine whether a student is ready to exit from an ELL program (Zehler et al, 2003). Students who are identified as ELLs are assessed annually for continued eligibility for ELL services often on the basis of their demonstration of oral English proficiency, classroom performance, or grades. A student who might be considered not in need of ELL services initially or not in need of continuation of the services under one screening system might possibly be provided with ELL services under a different system. In the meantime, parents have the right to choose among instructional programs and to remove their child from an ELL program (Zehler et al, 2008). The conceivable randomness in ELL services provision to a given student may be attributed to teacher discretion, parental preference, shifting assessment criteria, measurement errors, or temporal changes in staffing and other local resources.

Purpose / Objective / Research Question / Focus of Study:

This study focuses on assessing the contribution of ELL services to Spanish-speaking students' mathematics learning in elementary schools. ELL students tend to have lower average math achievement at school entry and throughout elementary school (Fry 2007; Reardon & Galindo 2009). Only 11 percent of ELL students scored at or above the proficient level on a national assessment of mathematics in comparison with the 36 percent proficient rate among all fourth graders (NCES, 2005). Spanish-speaking ELLs generally achieve lower academically yet demonstrate a faster learning rate than Asian and non-Hispanic white elementary students (Han, 2008). Earlier research has also indicated that ELL services may have differential impacts on Spanish-speaking and non-Spanish-speaking language minority students (Hong, 2012) and that the former group appears to be more vulnerable than the latter to the lack of school resources (Han, 2008). A counterfactual and seemingly politically incorrect question is: Would the Spanish-speaking ELLs be worse off if they were immersed in regular elementary classes instead without ELL services? The answer will supply important empirical evidence shedding light on the potential consequences of the highly inconsistent ELL program prevision across the country.

Specifically, given that a considerable proportion of language minority children receive no ELL services either when they start kindergarten or beyond kindergarten, we first ask whether Spanish-speaking ELL students in kindergarten would be worse off in their math learning if they were deprived of ELL services in kindergarten. The next question is whether these same students would be benefit from a second year of ELL services in first grade in comparison with having one year of ELL services in kindergarten only. A third research question is perhaps of higher theoretical and policy reference: Do more than three years of ELL services benefit Spanish-speaking students' math learning more than three or fewer years of services?

Additionally, we consider heterogeneity within the Spanish-speaking student population. Some students are self-made early bilinguals who may acquire English language skills even without ELL services; some are program-made early bilinguals who may garner benefit from ELL programs and become proficient in English within a relatively short time frame; while some other students are late bilinguals who may require sustained support of ELL services. We develop a strategy for identifying these subpopulations of students and investigating the optimal length of ELL services for each subpopulation.

Setting:

We draw data from the nationally representative ECLS-K data set. The students were followed from the beginning of kindergarten in 1998 through the end of fifth grade in 2004.

Population / Participants / Subjects:

We focus our interest on Spanish-speaking language minority students in US elementary schools. We have identified a sample of 2,205 Spanish-speaking language minority students entering kindergarten in 1998. Spanish-speaking students on average are perceived to have relatively low readiness for school (Crosnoe & Turley, 2011; Fuller et al, 2009) and high risk for high school dropout (Driscoll, 1999). In addition to the language barrier and immigration status, many of these students are from poor households and have parents with relatively low education (Crosnoe, 2006: Hernandez, 1999).

Intervention / Program / Practice:

For most Spanish-speaking students in the ECLS-K sample, we have identified their multi-year sequences of ELL services from kindergarten to fifth grade on the basis of teacher report supplemented by school administrator report on ELL program provision. For example, the teachers reported on whether every sampled student received in-class or pull-out ESL instruction, bilingual education, or Title I ESL or bilingual education, and whether the teacher used Spanish for instruction (see the technical report by West (2013) for details).

Research Design:

We conduct secondary analysis of large-scale longitudinal survey data evaluating ELL services as time-varying treatments. A major challenge to the evaluation of the causal effects of time-varying treatments was well-documented in the epidemiology literature over 100 years ago

(Arrighi & Hertz-Picciotto, 1994). In this context, students who display a higher level proficiency in English at kindergarten entry are less likely to be assigned to an ELL program; subsequently at the end of each school year, those who have gained more proficiency in English are more likely to exit the ELL program. Therefore, comparing the average level of academic outcome of the treated students with that of the untreated, or comparing the average outcome of those who have remained in an ELL program with those who have already exited, one would likely underestimate the potential benefit of ELL services. These dynamic selection processes result in the endogeneity problem in causal inference that cannot be easily addressed by most methods for statistical adjustment (Hong & Raudenbush, 2008).

We develop and apply a non-parametric marginal mean weighting through stratification (MMWS) strategy to remove selection bias associated with baseline and time-varying covariates. Our goal is to approximate a sequential randomized experiment in which students are hypothetically assigned at random to either an ELL program or a control condition at the beginning of each school year. The causal validity of the results depends primarily on the richness of the observed covariates, which we will describe below. Our key identification assumption is that, given the past observed covariate history, treatment history, and outcome history, a student's current treatment assignment is assumed independent of all the future potential outcomes such that the treatment assignment could be viewed as if randomized. When this assumption holds, the weighted mean observed outcome of a group experiencing a given treatment sequence consistently estimates the population mean outcome associated with that treatment sequence (Hong & Raudenbush, 2008; Robins, 1999). By employing a non-parametric procedure to estimate the weight, MMWS overcomes important limitations of the inverse-probability-of-treatment weighting (IPTW) method well-known to epidemiologists (Hernan, Brumbeck, & Robins, 2000) and generates relatively more robust and efficient estimates despite possible misspecification of the statistical models for predicting the probability of treatment assignment (Hong, 2010). We conduct a similar weighted analysis within each subpopulation of students to approximate a sequential randomized block design.

Data Collection and Analysis:

One outcome measure is math direct assessment scores. The assessment was administered in English if a student was proficient in English and was administered in Spanish if the student was proficient in Spanish but not in English. A second outcome measure is teacher rating of student math achievement. Teachers were asked to consider a student's math skills demonstrated in his or her native language if the student did not demonstrate skills in English. The direct assessment scores and the teacher ratings both have high reliability and have both been vertically scaled over multiple waves of observations. The correlation between the two outcome measures is about .5. This is partly because, unlike the math direct assessment, the teacher ratings measure not only the products but also the process of a student's math learning and reflect a broader sampling of the most recent math curriculum standards and guidelines. We use Y_t to denote a student's math outcome at time t for $t = -1, 0, 1, 3, 5$ representing the time at kindergarten entry, the spring of kindergarten, the spring of first grade, the spring of third grade, and the spring of fifth grade, respectively (there were no data collection in second grade and fourth grade). Here the value of t corresponds to the grade level.

Individual and contextual characteristics measured at the beginning of kindergarten, denoted by $\mathbf{X}_{-1}$, include student demographic characteristics, family characteristics, preschool experience, baseline oral English proficiency, and baseline academic and social-emotional skills.

Baseline covariates also include kindergarten class composition, school composition, and teacher characteristics. Time-varying covariates measured in each subsequent wave, denoted by $\mathbf{X}_t$ for $t = 0, 1, 3, 5$, include the evolving status of student oral English proficiency, academic and social-emotional skills, class and school compositions, and teacher characteristics.

Let Z_t denote the treatment received by a student in year t for $t = 0, 1, 3, 5$. Let $z_t = 1$ if a student received ELL services in year *t* and 0 otherwise. Table B.1 explains how we compute MMWS to adjust for bias due to selective treatment assignment at each time point. Figure B.1 provides more details on the sequential stratification of two years of data.

We analyze a two-level model and apply MMWS at level 1 for student *i* at time *t*. For example, with the first two years of data, the level 1 model is specified as

$$Y_{ti} = \beta_{-1i} + \beta_{0i}L_{0i}I_{ti}(t \geq 0) + \beta_{1i}L_{1i}I_{ti}(t = 1) + \delta_0^{(1)}I_i(Z_0 = 1)I_{ti}(t = 0) + \left[\delta_1^{(1,0)}I_i(Z_0 = 1)I_i(Z_1 = 0) + \delta_1^{(0,1)}I_i(Z_0 = 0)I_i(Z_1 = 1) + \delta_1^{(1,1)}I_i(Z_0 = 1)I_i(Z_1 = 1)\right]I_{ti}(t = 1) + \varepsilon_{ti},$$

$\varepsilon_{ti} \sim N(0, \sigma^2)$. At level 2, we have that $\beta_{-1i} = \gamma_{-1} + u_{-1i}$, $\beta_{0i} = \gamma_0 + u_{0i}$, and $\beta_{1i} = \gamma_1 + u_{1i}$. The random coefficients u_{-1i}, u_{0i}, and u_{1i} are assumed to be multivariate normal. Here $\delta_0^{(1)}$ estimates the effect of having ELL services in kindergarten on the kindergarten outcome; $\delta_1^{(1,0)}$, $\delta_1^{(0,1)}$, and $\delta_1^{(1,1)}$ estimate the respective effects on the first grade outcome of having ELL services in kindergarten only, in first grade only, and in both kindergarten and first grade. Table B.2 illustrates the level-1 data structure with four hypothetical students. We analyze a model with an analogous structure for multiple years of data.

Additionally, we consider three subpopulations of students empirically identified according to their potential trajectories of English language development if never treated or always treated in kindergarten and first grade. Table B.3 provides a summary. We combine MMWS with an extension of the Peters-Belson-Prognostic score method (Belson, 1956; Hansen, 2008; Peters, 1941) to identify subpopulation membership.

Findings / Results:

Table B.4 summarizes the observed frequency of multi-year treatment sequences. We have identified at least four years of treatment information for 1,070 students. Among them, 335 never received ELL services throughout the elementary school years while 440 received more than three years of services starting from kindergarten. Among the 1,689 students who have treatment information in kindergarten and first grade, about 50% of them received ELL services in both years while more than a third of them never received ELL services. More results later…

Conclusions:

The findings from this study will inform the theoretical discussion with regard to whether four or more years of ELL services on average are necessary to enable Spanish-speaking elementary students to become proficient in academic English essential for math learning. Yet a one-size-for-all recipe is practically naïve and often wasteful. Identifying the optimal length of ELL services for subpopulations of students therefore has immediate implications for ELL resource allocation. Nonetheless, empirical results from quasi-experimental data such as ECLS-K may contain remaining selection bias. To replicate the findings and draw more conclusive decisions, experimental designs contrasting the subpopulation-specific optimal treatment sequences identified in this study with the dominant sequences in the current practice may be employed as a natural next step in research.

Appendices

Appendix A. References

Arrighi, H. M., & Hertz-Picciotto, I. (1994). The evolving concept of the healthy worker survivor effect. *Epidemiology*, *5*(2), 189-196.

August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: a research agenda*. Washington, DC: National Academy Press.

Belson, W. A. (1956). A technique for studying the effects of a television broadcast. *Applied Statistics*, *5*, 195 –202.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, *21*, 617-641.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, *23*, 509-531.

Crosnoe, R. (2006). Health and the Education of Children from Racial/Ethnic Minority and Immigrant Families. *Journal of Health and Social Behavior*, *47*(1), 77-93.

Crosnoe, R. & Turley, R. L. (2011). K-12 educational outcomes of immigrant youth. *The Future of Children*, *21*(1), 129-152.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, *2*, 132-149.

Driscoll, A. (1999). Risk of high school dropout among immigrant and native Hispanic youth," *International Migration Review*, 33(4), 857–876.

Fry, R. (2007). *How far behind in reading and math are English Language Learners?* Washington, DC: Pew Hispanic Center.

Fuller, B., Bridges, M., Bein, E., Jang, H., Jung, S., Rabe-Hesketh, S., Halfon, N. & Kuo, A. (2009). The health and cognitive growth of latino toddlers: At risk or immigrant paradox? *Maternal and Child Health Journal*, 13 (6), 755-768.

Han, W.-J. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology*, *44*(6), 1572-1590.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, *95*(2), 481-488.

Hernan, M. A., Brumbeck, B., & Robins, J. M. (2000). Marginal structural mdoels to estimate the causal effect of zidovudine on the survival of HIV-positive mean. *Epidemiology*, *11*(5), 561-570.

Hernandez, D. J. (1999). *Children of immigrants: Health, adjustment, and public assistance*. Washington, DC: National Academies Press.

Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *Journal of Educational and Behavioral Statistics*, *35*(5), 499-531.

Hong, G. (2012). Marginal mean weighting through stratification: A generalized method for evaluating multi-valued and multiple treatments with non-experimental data. *Psychological Methods*, *17*(1), 44-60.

Hong, G., & Raudenbush, S. W. (2008) Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics*, *33*(3), 333-362.

National Center for Education Statistics. (2005). *Nation's report card for math*. Washington, DC: U.S. Department of Education, Institute of Educational Sciences.

National Center for Education Statistics. (2009). *The condition of education 2009, Indicator 8: language minority school-age children* (NCES 2009-081). Washington, DC: U.S. Department of Education, Institute of Educational Sciences.

National Center for Education Statistics. (2012). *The Condition of education 2011* (NCES 2011-045). Washington, DC: U.S. Department of Education, Institute of Educational Sciences.

Peters, C. C. (1941). A method of matching groups for experiments with no loss of populations. *Journal of Educational Research, 34*, 606–612.

Reardon, S. & Galindo, C. (2009). The Hispanic-white achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853-891.

Robins, J. (1999). Marginal structural models versus structural nested models as tools for causal inference. In M. E. Halloran & D. Berry (Eds.), *Statistical models in epidemiology, the environment, and clinical trials* (pp. 95–134). New York: Springer.

West, A. (2013). *Technical report on ELL services for Spanish-speaking English language learners: K-5*. Chicago, IL: University of Chicago.

Zehler, A. M., Adger, C., Coburn, C., Arteagoitia, I., Williams, K., & Jacobson, L. (2008). *Preparing to serve English language learner students: School districts with emerging English language learner communities*. Washington, DC: U.S. Department of Education, Institute of Education Sciences National Center for Education Evaluation and Regional Assistance.

Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Stephenson, T. G., Pendzick, M. L., & Sapru, S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities. Volume 1: Research Report.* Report submitted to the U.S. Department of Education, Office of English Language Acquisition. Arlington, VA: Development Associates, Inc.

Appendix B. Tables and Figures

Table B.1. MMWS Computation

Time	Propensity Score	Stratification on	MMWS
0	$\theta_0^{(Z_0=z_0)} = pr(Z_0 = z_0 \mid Y_{-1}, \mathbf{X}_{-1})$	$\hat{\theta}_0^{(Z_0=z_0)}$	$W^{(z_0)} = \dfrac{pr(Z_0 = z_0)}{pr\left(Z_0 = z_0 \mid S_0^{(z_0)} = s_0\right)}$
1	$\theta_1^{(Z_1=z_1 \mid Z_0=z_0)} = pr(Z_1 = z_1 \mid z_0, Y_{-1}, Y_0, \mathbf{X}_{-1}, \mathbf{X}_0)$	$\hat{\theta}_1^{(Z_1=z_1 \mid Z_0=z_0)}$	$W^{(z_0.z_1)} = W^{(z_0)} \times W_1^{(z_1 \mid z_0)},$ $W_1^{(z_1 \mid z_0)} = \dfrac{pr(Z_1 = z_1 \mid Z_0 = z_0)}{pr\left(Z_1 = z_1 \mid Z_0 = z_0, S_1^{(z_1 \mid z_0)} = s_1\right)}$
3	$\theta_3^{(Z_3=z_3 \mid \bar{z}_{0\sim1})} = pr(Z_3 = z_3 \mid \bar{z}_{0\sim1}, \bar{Y}_{-1\sim1}, \bar{\mathbf{X}}_{-1\sim1})$	$\hat{\theta}_3^{(Z_3=z_3 \mid \bar{z}_{0\sim1})}$	$W^{(\bar{z}_{0\sim3})} = W^{(z_0)} \times W_1^{(z_1 \mid z_0)} \times W_3^{(z_3 \mid z_0, z_1)},$ $W_3^{(z_3 \mid z_0, z_1)} = \dfrac{pr(Z_3 = z_3 \mid Z_0 = z_0, Z_1 = z_1)}{pr\left(Z_3 = z_3 \mid Z_0 = z_0, Z_1 = z_1, S_3^{(z_3 \mid z_0, z_1)} = s_3\right)}$
5	$\theta_5^{(Z_5=z_5 \mid \bar{z}_{0\sim3})} = pr(Z_5 = z_5 \mid \bar{z}_{0\sim3}, \bar{Y}_{-1\sim3}, \bar{\mathbf{X}}_{-1\sim3})$	$\hat{\theta}_5^{(Z_5=z_5 \mid \bar{z}_{0\sim3})}$	$W^{(\bar{z}_{0\sim5})} = W^{(z_0)} \times W_1^{(z_1 \mid z_0)} \times W_3^{(z_3 \mid z_0, z_1)} \times W_5^{(z_5 \mid z_0, z_1, z_3)},$ $W_5^{(z_5 \mid z_0, z_1, z_3)} = \dfrac{pr(Z_5 = z_5 \mid Z_0 = z_0, Z_1 = z_1, Z_3 = z_3)}{pr\left(Z_5 = z_5 \mid Z_0 = z_0, Z_1 = z_1, Z_3 = z_3, S_5^{(z_5 \mid z_0, z_1, z_3)} = s_5\right)}$

Note: Here $\bar{z}_{0\sim1}$ denotes the observed treatment history (z_0, z_1) and $\bar{z}_{0\sim3}$ denotes the observed treatment history (z_0, z_1, z_3). Propensity score estimation and stratification are always conducted within a subsample of students who shared the same treatment history. Additionally, $\bar{Y}_{-1\sim1}$ and $\bar{Y}_{-1\sim3}$ denote the outcome history (Y_{-1}, Y_0, Y_1) and (Y_{-1}, Y_0, Y_1, Y_3), respectively; and $\bar{\mathbf{X}}_{-1\sim1}$ and $\bar{\mathbf{X}}_{-1\sim3}$ denote the covariate history (X_{-1}, X_0, X_1) and (X_{-1}, X_0, X_1, X_3), respectively.

Table B.2

Repeated Observations at Level 1 for a Two-Level Model for Evaluating Two-Year Treatment Sequences

Student	Y_t	W_t	$L_0 I_t(t \geq 0)$	$L_1 I_t(t = 1)$	$I(Z_0 = 1)I_t(t = 0)$	$I(Z_0 = 1)I_t(t = 1)$	$I(Z_1 = 1)I_t(t = 1)$	$I(Z_0 = 1)I(Z_1 = 1)I_t(t = 1)$
1	22	1	0	0	0	0	0	0
1	32	$W^{(0)}$	8	0	0	0	0	0
1	42	$W^{(0.0)}$	8	13	0	0	0	0
2	20	1	0	0	0	0	0	0
2	30	$W^{(1)}$	8	0	1	0	0	0
2	45	$W^{(1.0)}$	8	13	0	1	0	0
3	24	1	0	0	0	0	0	0
3	28	$W^{(0)}$	9	0	0	0	0	0
3	48	$W^{(0.1)}$	9	12	0	0	1	0
4	18	1	0	0	0	0	0	0
4	30	$W^{(1)}$	9	0	1	0	0	0
4	40	$W^{(1.1)}$	9	12	0	1	1	1

Table B.3

Definition of Three Student Subpopulations

Subpopulations	Would Become Orally Proficient in English in Two Years	Under Treatment	
		Kindergarten	1st Grade
Self-Made Early Bilinguals	Yes	No	No
Program-Made Early Bilinguals	Yes	Yes	Yes
Late Bilinguals	No	Yes	Yes

Table B.4

Frequency of Multi-Year Treatment Sequences

Six Years (K-5)		Four Years (K-3)		Two Years (K-1)		One Year (K)	
Treatment Sequences	N	Treatment Sequences	N	Treatment Sequences	N	Treatment Sequences	N
0-0-0-0	324	0-0-0	335	0-0	609	0	941
1-0-0-0	110	1-0-0	117	1-0	140	1	1097
1-1-0-0	98	1-1-0	117	1-1	849		
1-1-1-0	53	1-1-1	440				
1-1-1-1	231						
0-1-0-0	24	0-1-0	30	0-1	91		
0-1-1-0	3	0-1-1	31				
0-1-1-1	16						
Total	859	Total	1,070	Total	1,689	Total	2,038

Figure B.1

Sequential Stratification of Two Years of Observations

Stratify on $\hat{\theta}_0^{(z_0)}$

$S_0^{(z_0)} = s_0$	$Z_0 = 0$	$Z_0 = 1$
1		
2		
3		
4		
5		
6		

Stratify on $\hat{\theta}_1^{(z_1|z_0=0)}$

$S_1^{(z_1|z_0=0)} = s_1$	$Z_1 = 0$	$Z_1 = 1$		$Z_1 = 0$	$Z_1 = 1$	$S_1^{(z_1|z_0=1)} = s_1$
1						1
2						2
3						3
4						4
5						5
6						6

Stratify on $\hat{\theta}_1^{(z_1|z_0=1)}$

Note:

We first stratify the year-0 data on the estimated propensity score $\hat{\theta}_0^{(z_0)}$ for the year-0 treatment and use $s_0 = 1, \ldots, 6$ to denote the six strata. We then compute $W^{(0)} = \frac{pr(Z_0=0)}{pr\left(Z_0=0|S_0^{(z_0)}=s_0\right)}$ for a student in stratum s_0 whose year-0 treatment is $Z_0 = 0$. The weight is $W^{(1)} = \frac{pr(Z_0=1)}{pr\left(Z_0=1|S_0^{(z_0)}=s_0\right)}$ for a student in stratum s_0 whose year-0 treatment is $Z_0 = 1$.

For students who had the same year-0 treatment $Z_0 = 0$, we stratify their year-1 observations on the estimated propensity score for the year-1 treatment $\hat{\theta}_1^{(z_1|z_0=0)}$ and use $S_1^{(z_1|z_0=0)} = s_1$ for $s_1 = 1, \ldots, 6$ to denote the six strata. The weight for the year-2 observation of a student in stratum s_1 with treatment sequence $(0,0)$ is $W^{(0.0)} = W^{(0)} \times W_1^{(0|0)}$ where

$W_1^{(0|0)} = \frac{pr(Z_1=0|Z_0=0)}{pr\left(Z_1=0|Z_0=0,S_1^{(z_1|z_0=0)}=s_1\right)}$. The weight for a student in the same stratum with treatment sequence $(0,1)$ is $W^{(0.1)} = W^{(0)} \times W_1^{(1|0)}$ where $W_1^{(1|0)} = \frac{pr(Z_1=1|Z_0=0)}{pr\left(Z_1=1|Z_0=0,S_1^{(z_1|z_0=0)}=s_1\right)}$.

The year-1 observations of students who had the same year-0 treatment $Z_0 = 1$ are stratified on the estimated propensity score $\hat{\theta}_1^{(z_1|z_0=1)}$ instead. This creates a different set of strata denoted by $S_1^{(z_1|z_0=1)} = s_1$ for $s_1 = 1, \ldots, 6$. The weight for the year-2 observation of a student in stratum s_1 with treatment sequence $(1,0)$ is $W^{(1.0)} = W^{(1)} \times W_1^{(0|1)}$ where $W_1^{(0|1)} = \frac{pr(Z_1=0|Z_0=1)}{pr\left(Z_1=0|Z_0=1,S_1^{(z_1|z_0=1)}=s_1\right)}$. The weight for the year-2 observation of a student in stratum s_1 with treatment sequence $(1,1)$ is $W^{(1.1)} = W^{(1)} \times W_1^{(1|1)}$ where $W_1^{(1|1)} = \frac{pr(Z_1=1|Z_0=1)}{pr\left(Z_1=1|Z_0=1,S_1^{(z_1|z_0=1)}=s_1\right)}$.