



## A Model-Based Approach To Measure School Achievements In Latent Groups Of Students: A Simulation Study

Giulia Roli & Paola Monari

To cite this article: Giulia Roli & Paola Monari (2014): A Model-Based Approach To Measure School Achievements In Latent Groups Of Students: A Simulation Study, Communications in Statistics - Simulation and Computation, DOI: [10.1080/03610918.2014.944653](https://doi.org/10.1080/03610918.2014.944653)

To link to this article: <http://dx.doi.org/10.1080/03610918.2014.944653>



Accepted author version posted online: 10 Nov 2014.



Submit your article to this journal [↗](#)



Article views: 6



View related articles [↗](#)



View Crossmark data [↗](#)

# ACCEPTED MANUSCRIPT

## A MODEL-BASED APPROACH TO MEASURE SCHOOL ACHIEVEMENTS IN LATENT GROUPS OF STUDENTS: A SIMULATION STUDY

**Giulia Roli and Paola Monari**

Department of Statistical Sciences

University of Bologna

Corresponding author:

Giulia Roli

Dipartimento di Scienze Statistiche “P. Fortunati”

via delle Belle Arti, 41, 40126, Bologna

[g.roli@unibo.it](mailto:g.roli@unibo.it)

Short title:

A model-based approach to measure school achievement in latent groups

ACCEPTED MANUSCRIPT

## **A model-based approach to measure school achievements in latent groups of students: a simulation study**

### **Abstract**

In this paper, we present a model-based framework to estimate the educational attainments of students in latent groups defined by unobservable or only partially observed features which are likely to affect the outcome distribution, as well as being interesting to be investigated. We focus our attention on the case of students in the first year of the upper secondary schools, for which the teachers' suggestion at the end of their lower educational level toward the subsequent type of school is available. **We use this information to develop latent strata according to the compliance behavior of students simplifying to the case of binary data for both counselled and attended school (i.e., academic or technical institute).** We consider a likelihood-based approach to estimate outcome distributions in the latent groups and propose a set of plausible assumptions with respect to the problem at hand. In order to assess our method and its robustness, we simulate data resembling a real study conducted on pupils of the province of Bologna in year 2007/2008 to investigate their success or failure at the end of the first school year.

**Keywords:** student achievement, teachers' suggestions, compliance behavior, principal latent strata.

## **1 Introduction**

The investigation of the factors which may influence the achievement of students in the different levels of their education is a crucial topic in observational studies on individual learning experiences. Latent characteristics may further play a crucial role and their study has drawn the attention of several researchers in the last decades (see, e.g., Glas and Geerlings, 2009).

In this paper, we aim at discerning latent subgroups of students characterized by specific features which can be interesting to be investigated as they may differently affect the educational attainment. In particular, we consider the case of students in the first year of the upper secondary schools and rely on the presence of additional data on teachers' suggestion at the end of the lower educational level toward the subsequent type of school. We exploit the potential outcomes approach (see Rubin, 1974) and Principal Stratification (PS; Frangakis and Rubin, 2002) to identify latent subgroups of pupils according to their compliance behavior with respect to the teachers' recommendations. As a first attempt, both counselled and attended type of school are simplified into binary data (e.g., academic or technical institute).

Despite PS approach was firstly introduced in an experimental setting to address the problem of causal inference in cases of post-treatment complications, here it is used in an observational context to estimate more from the data than average causal effect, as also pointed out by Imbens and Rubin in 1997. In particular, we aim at studying how the teachers' suggestion may affect the outcome distribution of the student performances in terms of success or failure at the end of the first school year. We show that the latent stratum of compliers, that is students always following the teachers' advice, represents an interesting group to evaluate if the suggestion correctly directs students toward the subsequent type of school. In such case, we expect high rates of success at the end of the first year of the upper secondary schools. With this aim, we develop a model-based approach following Instrumental Variables (IV) approach (Angrist *et al.*, 1996) and by relaxing assumptions which may be implausible according to the problem at hand.

In order to assess our method and its robustness, we simulate data to resemble a real study conducted on pupils of the province of Bologna in year 2007/2008 to investigate their success or failure at the end of the school year. Different scenarios corresponding to different assumptions will be also compared.

The paper develops as follows. **Section 2 describes** the method and the main assumptions adopted to identify the latent groups. In section 3, the parametric approach and the distributional

assumptions are introduced. Then, the simulation study and the results are reported (sections 4 and 5). Section 6 concludes with main remarks and further developments of the work.

## 2 Identifying latent groups

As a streamlined example, we consider a set of  $N$  students where **each unit**  $i$  may be counselled by their lower education teachers at only two types of upper secondary schools  $Z$ , say academic ( $Z_i = 1$ ) or technical ( $Z_i = 0$ ) institute. The extension to a **polytomous** case with more than two types of school is straightforward. Then, let  $\mathbf{Z}$  be the  $N$ -dimensional vector of this assignment with  $i$ -th element  $Z_i$ , and let  $D_i(\mathbf{Z})$  be the binary indicator for the upper secondary school that unit  $i$  **chooses** given the allocated vector of teachers' assignment. Thus,  $D_i(\mathbf{Z})$  equals 1 if the attended school is academic and  $D_i(\mathbf{Z})$  equal 0 if it is technical.

Within the framework of the Rubin Causal Model (RCM; Rubin, 1974),  $D_i(\mathbf{Z})$  is a potential outcome because we only observe the choice of students given the vector of teachers' assignments. In a perfect compliance environment,  $D_i(\mathbf{Z})$  would equal  $Z_i$  for all units  $i$ , that is all students would attend the school assigned by teachers, **thus following** the teachers' suggestions. In real cases, students can be non-compliers for several reasons, such as personal, cultural, family or logistic causes.

Similar to the definition of  $D_i(\mathbf{Z})$ , we can define the potential outcome  $Y_i(\mathbf{Z}, \mathbf{D})$  to be the response measuring the achievement of student  $i$  given the vectors  $\mathbf{Z}$  and  $\mathbf{D}$  of assignments and attendances, respectively. In our example, let consider  $Y$  as the response indicator of whether students would pass ( $Y_i = 1$ ) or not ( $Y_i = 0$ ) to the second year at the end of the first year of the chosen upper secondary school.

In order to simplify the structure of notation, the Stable Unit Treatment Value Assumption (SUTVA; Rubin, 1980) is usually invoked. This guesses that no interference between units is allowed and both levels of the assignment and actual attendance define a unique outcome for each

unit:

*Assumption: SUTVA.*

- If  $Z_i = Z'_i$ , then  $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$ .
- If  $Z_i = Z'_i$  and  $D_i = D'_i$ , then  $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(\mathbf{Z}', \mathbf{D}')$ .

Thus, SUTVA allows us to simply write  $D_i(\mathbf{Z})$  and  $Y_i(\mathbf{Z}, \mathbf{D})$  as  $D_i(Z_i)$  and  $Y_i(Z_i, D_i)$ , respectively.

As formulated, both potential outcomes  $D_i(Z_i)$  and  $Y_i(Z_i, D_i)$  are fixed but only partially observable. Indeed, the observed data for each unit  $i$  are denoted by  $(Z_i, D_i^{\text{obs}}(Z_i), Y_i^{\text{obs}}(Z_i, D_i^{\text{obs}}))$ , indicating that we only observe the outcome  $Y_i$  given the solely chosen observable school  $D_i^{\text{obs}}(Z_i)$  and the assignment  $Z_i$ .

As introduced above,  $D_i(Z_i)$  describes the compliance behavior of each unit. We can suppose that this could be critical and interesting to be investigated in terms of yielded differences in the school performances of pupils at least at the end of the upper secondary first year. Thus, we exploit it to define an indicator partitioning the population of students into four latent subgroups, named Principal Latent Strata (PLS): compliers (i.e., students always following teachers' suggestion), always takers (i.e., students always choosing academic school, independently on the suggestion), never-takers (i.e., students always choosing technical institute, independently on the suggestion) and defiers (i.e., students who choose the opposite school with respect to the proposed one). We denote the stratum membership with  $C_i$ , where:

$$C_i = \begin{cases} c \text{ (i.e., unit is a complier),} & \text{if } D_i(z) = z \text{ for } z = 0, 1 \\ a \text{ (i.e., unit is a always-taker),} & \text{if } D_i(z) = 1 \text{ for } z = 0, 1 \\ n \text{ (i.e., unit is a never-taker),} & \text{if } D_i(z) = 0 \text{ for } z = 0, 1 \\ d \text{ (i.e., unit is a defier),} & \text{if } D_i(z) = 1 - z \text{ for } z = 0, 1 \end{cases} \quad (1)$$

All the observed groups, characterized by the observed value of  $Z$  and  $D$ , thus result from a mixture of a number of principal strata. Table 1 **shows** the different strata included in each observed group ( $O(z, D^{\text{obs}}(z))$ ), with  $z = 0, 1$ . Group  $O(1, 1)$  identifies students who have been advised to and actually attended academic school. Those can belong to the latent stratum of compliers or always-takers. The observed group of students advised to academic school but attending technical institute,  $O(1, 0)$ , is a mixture of never-takers and defiers. Conversely, students advised to technical institute but attending academic school,  $O(0, 1)$ , can belong to the stratum of always-takers or defiers. Finally, the observed group  $O(0, 0)$  identifies students who are advised to and actually attend technical institute and results from a mixture of compliers and never-takers.

Discerning the latent stratum of compliers, that is students always following the teachers advice, represents an interesting topic as it allows to evaluate if the suggestion correctly directs students toward the subsequent type of school. In particular, if students are properly advised we expect high performances at the end of the school year. Moreover, we can evaluate the success rate for compliers separately by type of school, as the latent stratum is involved in the two observed groups of students advised and attending academic ( $O(1, 1)$ ) or technical institute ( $O(0, 0)$ ).

Nevertheless, the distributions of always- and never- takers can be further interesting. In particular, the performance of students attending a school different from that advised from teachers can provide crucial information, reflecting, e.g., personal motivation of students toward the chosen school and capability of teachers to directs students, independently on the actual choice.

### 3 Likelihood approach

In order to estimate the outcome distributions of the strata, both parametric and nonparametric strategies can be carried out, depending on the set of assumptions that can be reasonably maintained. Under a nonparametric perspective, a set of sufficient assumptions are specified and invoked in order to identify (some features of) the distribution of the outcome in the latent groups

(Frolich, 2007). A parametric approach is here proposed by introducing some distributional hypothesis that are consistent with the data and exploiting the results on finite mixture distribution theory (McLachlan and Peel, 2000).

By considering the correspondence between the observed groups and the latent strata (Table 1) and further considering the effect of a set of covariates  $\mathbf{X}$ , we obtain the likelihood function to be maximized according to the vector of parameters  $\theta$ , which is shown to result in a finite mixture of distributions:

$$\begin{aligned}
 L(\theta|Z, D^{\text{obs}}(Z), Y^{\text{obs}}(Z, D^{\text{obs}})) \propto & \\
 & \prod_{i \in O(1,1)} P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = c) \times P(C_i = c | \mathbf{X}_i = \mathbf{x}_i) + \\
 & + P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = a) \times P(C_i = a | \mathbf{X}_i = \mathbf{x}_i) \times \\
 & \prod_{i \in O(1,0)} P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = n) \times P(C_i = n | \mathbf{X}_i = \mathbf{x}_i) + \\
 & + P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = d) \times P(C_i = d | \mathbf{X}_i = \mathbf{x}_i) \times \\
 & \prod_{i \in O(0,1)} P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = a) \times P(C_i = a | \mathbf{X}_i = \mathbf{x}_i) \\
 & + P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = d) \times P(C_i = d | \mathbf{X}_i = \mathbf{x}_i) \times \\
 & \prod_{i \in O(0,0)} P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = c) \times P(C_i = c | \mathbf{X}_i = \mathbf{x}_i) \\
 & + P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i = n) \times P(C_i = n | \mathbf{X}_i = \mathbf{x}_i) \tag{2}
 \end{aligned}$$

**In order to form the likelihood function, we have to specify probability distributions for the potential outcomes,  $P(Y(z, D_i(z)) = y_i | \mathbf{X}_i = \mathbf{x}_i, C_i)$ , as well as for the principal strata membership  $C$ ,  $P(C_i | \mathbf{X}_i = \mathbf{x}_i)$ . These choices depend on the type of data and application at hand (see Section 4).**

Some additional assumptions can be further imposed to reduce the number of strata or stating the equivalence of the outcome distribution across strata. These assumptions are often invoked on the ground that they may facilitate the convergence of the likelihood optimization process, as well



as being reasonable in most part of studies. Angrist *et al.* in 1996 deeply discussed this issue by connecting the approach to that on Instrumental Variables (with variable  $Z$  which can be viewed as the instrument), which has a long usage and tradition especially in econometric field. These assumptions are (weak) Exclusion-Restriction (ER) and Monotonicity. The former would guess the assignment value  $z$  is unrelated to the potential outcomes, formally:

$$Y(0, 0) = Y(0, 1) \text{ and } Y(1, 0) = Y(1, 1).$$

This can be true only for strata of always- and never- takers, but, despite it is not directly verifiable from the data, it seems to be implausible with respect to the problem at hand, unless teachers' suggestions are considered completely unrelated to the student performance.

Conversely, monotonicity assumption is more reasonable, as it states the non-existence of defiers:

$$D_i(0) \leq D_i(1) \quad \forall i.$$

Indeed, it seems quite implausible that choices of students are **based** on the opposite of teachers' advices and, even if it can be possible, we can suitably assume this group of students would be very small. In the simulation exercise we consider both scenarios, in order to assess the model performances.

## 4 Data simulation

In order to test the proposing method, we generate an artificial data **set** based on a real study conducted on the province of Bologna, through the *Osservatorio della Scolarità* agency, to experiment a more detailed collection of data on enrolled students. In particular, family socio-cultural and economic conditions, in addition to the educational and demographic characteristics of the students (including lower education grade and teachers' suggestion toward upper schools) have been gathered through a specific questionnaire during the registration at the upper secondary school in year

2007/2008.

Referring to this motivating example, we simulate a sample of about 5,000 individuals resembling the proportions of the assignment and attendance regarding the two types of school to those observed in the real data. Data are generated under two main scenarios:

- **Scenario (1):** existence of stratum of defiers, i.e. relaxing the monotonicity assumption;
- **Scenario (2):** non-existence of defiers, i.e. under the monotonicity assumption.

The observable features of the two generated samples are resumed in table 2 and 3.

**The potential outcome variable indexes the success of students and, thus, its distribution has a logistic regression form conditional on each stratum and covariates:**

$$P(Y(z, D_i(z)) = 1 | \mathbf{X}_i = \mathbf{x}_i, C_i) = \pi_i^{z, D(z), C} = \frac{\exp(\alpha_0^{z, D(z), C} + \alpha_1^{z, D(z), C} \mathbf{x}_i)}{1 + \exp(\alpha_0^{z, D(z), C} + \alpha_1^{z, D(z), C} \mathbf{x}_i)} \quad (3)$$

The distribution of the principal latent strata is modeled as a multinomial logit where

$$P(C_i | \mathbf{X}_i = \mathbf{x}_i) = \pi_i^{C^*} = \frac{\exp(\delta_0^{C^*} + \delta_1^{C^*} \mathbf{x}_i)}{1 + \sum_{C=c, a, n} \exp(\delta_0^C + \delta_1^C \mathbf{x}_i)} \quad \forall C^*, C^* = c, a, n$$

$$\pi_i^d = 1 - \sum_{C=c, a, n} \pi_i^C$$

under scenario (1) and

$$P(C_i | \mathbf{X}_i = \mathbf{x}_i) = \pi_i^{C^*} = \frac{\exp(\delta_0^{C^*} + \delta_1^{C^*} \mathbf{x}_i)}{1 + \sum_{C=c, a} \exp(\delta_0^C + \delta_1^C \mathbf{x}_i)} \quad \forall C^*, C^* = c, a$$

$$\pi_i^n = 1 - \sum_{C=c, a} \pi_i^C$$

under scenario (2).

We specify parameters which reproduce credible results in terms of percentage of success/failure at the end of the first school year. To simplify, a univariate covariate  $X$  is considered which can be viewed as a summary of several variables. In this first attempt, we suppose a common effect of the univariate covariate on the outcome across the strata. Table 4 reports the values of the parameters used to generate the artificial data under both scenarios.

## 5 Results

During the estimation process, carried out using R (version 2.13.1), different assumptions under the two scenarios are compared, following this scheme:

- **Scenario (1a)**: estimation with assumptions of data generation.
- **Scenario (1b)**: estimation by wrongly assuming exclusion restriction.
- **Scenario (2a)**: estimation under no monotonicity, i.e. existence of defiers.
- **Scenario (2b)**: estimation with assumptions of data generation.

The results of our analysis are in Tables 5 and 6. These report the true and estimated parameters of strata proportions and outcome distribution under the different constraints and scenarios described above. Standard errors are computed by using the delta method.

The estimates of the strata proportions seem to be precise and generally robust to misspecifications of the potential outcomes distributions in the principal strata. Moreover, when the monotonicity is wrongly relaxed (scenario (2a)), our method properly identifies the non-existence of defiers. When proper assumptions (scenarios (1a) and (2b)) are compared with misspecified ones, Aikaike information criterion (AIC) rightly directs toward the true model.

A quite good performance is further observed in the estimation of the outcome distribution, even if disentangling the outcome distribution from the mixture seem to be more difficult, being even more sensitive to the stratum and the observed group sizes. Thus, for instance, under scenario (1) estimates seem less robust in the stratum of defiers, i.e. the group with the smallest sample size, especially when exclusion-restriction is wrongly assumed, despite the stratum is not directly involved (scenario (1b)). In such cases, additional assumptions can be useful to be imposed in order to facilitate the convergence of the likelihood optimization process. In our example, a more realistic scenario assumes non-existence of defiers, as stated and justified above. In such case,

**this** actually reduces bias in the estimates as shown by results of scenario (2b), especially for the outcome distribution parameters.

## 6 Final remarks

We presented a likelihood-based approach to estimate the performances of students in latent groups defined by their compliance behavior with respect to the teachers' advices. We use the PS stratification approach and the finite mixture distribution theory to identify and disentangle the outcome distribution in the latent groups by counselled and attended type of school. In particular, the group of compliers can be considered notably interesting to be evaluated, because it contains information on effectiveness of teachers' suggestions, controlling for the different type of school.

We implemented a simulation study based on real data to assess our method and its robustness with respect to different scenarios and related misspecifications. We proposed a set of plausible assumptions regarding the problem at hand and facilitating the convergence of the optimization process, as well as yielding more accurate estimates.

Even if we considered a simplification of data with only two possible types of suggested and attended school, the method we propose can be easily extended to the multivalued case. Moreover, as a further development we aim at applying the method to real data.

## References

Angrist, J. D., Imbens, G. W., Rubin, D. B.: Identification of causal effects using instrumental variables (with discussion). *J. Amer. Statist. Assoc.* **91**, 444–472 (1996).

Frangakis, C. E., Rubin, D. B.: Principal stratification in causal inference. *Biometrics*, **58**, 191–199 (2002).

Frolich, M.: Nonparametric IV estimation of local average treatment effects with covariates, *J.*

Econometrics, **139**, 35–75 (2007).

Glas, C. A. W., Geerlings, H.: Psychometric aspects of pupil monitoring systems. *Stud. Educ. Eval.*, **35**, 83–88 (2009).

Imbens, G. W., Rubin, D. B.: Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann. Statist.* **25**, 305–327 (1997).

McLachlan, G., Peel, D.: *Finite Mixture Models*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons (2000).

Rosenbaum, P., Rubin, D. B.: The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).

Rubin, D. B.: Estimating causal effects of treatments in randomized and nonrandomized Studies. *J. Educat. Psychol.* **66**, 688–701 (1974).

Rubin, D. B.: Discussion of Randomization analysis of experimental data: the Fisher randomization test by D. Basu. *J. Amer. Statist. Assoc.* **75**, 591–593 (1980).

## Tables

Table 1: Correspondence between observed **data** and latent strata

$O(z, D^{\text{obs}}(z))$	Latent strata
$O(1, 1)$	$c, a$
$O(1, 0)$	$n, d$
$O(0, 1)$	$a, d$
$O(0, 0)$	$c, n$

Table 2: Observed data **by teacher assignment  $Z$  and actual choice  $D^{\text{obs}}(Z)$**  - Scenario (1)

Number of observations	$Z = 0$	$Z = 1$
Tot.	1764 (35.3%)	3236 (64.7%)
$D^{\text{obs}}(Z) = 0$	840 (47.6%)	890 (27.5%)
$D^{\text{obs}}(Z) = 1$	924 (52.4%)	2346 (72.5%)
Outcome $Y=1$ (rate of success)	$D^{\text{obs}}(Z) = 0$	$D^{\text{obs}}(Z) = 1$
Tot.	1520 (87.9%)	2941 (90.0%)
$Z = 0$	783 (93.2%)	720 (80.9%)
$Z = 1$	737 (79.8%)	2221 (94.7%)

Table 3: Observed data **by teacher assignment  $Z$  and actual choice  $D^{\text{obs}}(Z)$**  - Scenario (2)

Number of observations	$Z = 0$	$Z = 1$
Tot.	1753 (35.1%)	3247 (64.9%)
$D^{\text{obs}}(Z) = 0$	1090 (62.2%)	447 (13.8%)
$D^{\text{obs}}(Z) = 1$	663 (37.8%)	2800 (86.2%)
Outcome $Y=1$ (rate of success)	$D^{\text{obs}}(Z) = 0$	$D^{\text{obs}}(Z) = 1$
Tot.	1404 (91.3%)	3099 (89.5%)
$Z = 0$	1002 (91.9%)	469 (70.7%)
$Z = 1$	402 (89.9%)	2630 (93.9%)

Table 4: Values of the parameters to generate artificial data **by scenario**

Parameters	Scenario (1)	Scenario (2)
Strata proportions:		
$\delta_0^c$	1.2	1.2
$\delta_1^c$	3	3
$\delta_0^a$	1	1
$\delta_1^a$	1.1	1.1
$\delta_0^n$	1	-
$\delta_1^n$	1	-
Outcome distribution:		
$\alpha_1$	0.8	0.8
$\alpha_0^{1,1,c}$	2.6	2.6
$\alpha_0^{1,1,a}$	2.5	2.5
$\alpha_0^{1,0,n}$	2.7	2.7
$\alpha_0^{1,0,d}$		-
$\alpha_0^{0,1,a}$	1.6	1.6
$\alpha_0^{0,1,d}$	1.4	-
$\alpha_0^{0,0,n}$	3.2	3.2
$\alpha_0^{0,0,c}$	2.7	2.7

Table 5: True and estimated parameters (se=standard error) - Scenario (1)

Strata	True parameters		Estimates (se) - Scenario (1a)		Estimates (se) - Scenario (1b)	
	proportion	% successes	proportion	% successes	proportion	% successes
$C_i = c$	0.397		0.376 (0.016)		0.406 (0.024)	
$Z_i = 1, D_i(Z_i)=1$		95.3		95.0 (0.010)		96.3 (0.008)
$Z_i = 0, D_i(Z_i)=0$		92.5		93.2 (0.019)		98.3 (0.077)
$C_i = a$	0.218		0.240 (0.017)		0.222 (0.025)	
$Z_i = 1, D_i(Z_i)=1$		94.3		95.6 (0.015)		88.1 (0.018)
$Z_i = 0, D_i(Z_i)=1$		72.5		73.8 (0.027)		
$C_i = n$	0.227		0.246 (0.015)		0.212 (0.022)	
$Z_i = 1, D_i(Z_i)=0$		92.1		93.7 (0.056)		93.0 (0.014)
$Z_i = 0, D_i(Z_i)=0$		92.8		94.9 (0.049)		
$C_i = d$	0.157		0.137 (0.016)		0.159 (0.023)	
$Z_i = 1, D_i(Z_i)=0$		77.7		71.0 (0.033)		89.5 (0.039)
$Z_i = 0, D_i(Z_i)=1$		71.6		68.9 (0.129)		61.0 (0.042)
AIC			13016.5		13251.12	

Table 6: True and estimated parameters (se=standard error) - Scenario (2)

Strata	True parameters		Estimates (se) - Scenario (2a)		Estimates (se) - Scenario (2b)	
	proportion	% successes	proportion		proportion	% successes
$C_i = c$	0.479		0.495 (0.012)		0.482 (0.011)	
$Z_i = 1, D_i(Z_i)=1$		94.8				94.0 (0.009)
$Z_i = 0, D_i(Z_i)=0$		92.6				93.1 (0.015)
$C_i = a$	0.308		0.283 (0.013)		0.302 (0.010)	
$Z_i = 1, D_i(Z_i)=1$		91.9				90.2 (0.019)
$Z_i = 0, D_i(Z_i)=1$		70.7				75.9 (0.014)
$C_i = n$	0.213		0.212 (0.008)		0.216 (0.006)	
$Z_i = 1, D_i(Z_i)=0$		89.9				85.9 (0.019)
$Z_i = 0, D_i(Z_i)=0$		91.4				93.7 (0.037)
$C_i = d$	-		0.01 (0.090)		-	
AIC			12135.0		12063.79	