Effect of sample design and loss of follow-up on parameters estimation of a longitudinal study

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva

UFMG / FIOCRUZ

September 24, 2008

Sample Design in Survey Research and Longitudinal Studies

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

- Complex sample design to optimize the sample size
- How to consider it in the analysis?
  - No doubt related to descriptive parameters
  - We are far from an agreement about the better strategy when adjusting an association model
- Longitudinal models $\longrightarrow$ another problem: losses

Effect of sample design and loss of follow-up on parameters estimation of a longitudinal study

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Motivation - ELSA

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

The Longitudinal Study of Adult Health - ELSA Brasil

- Multicenter cohort study
- fifteen thousand civil servants
- six academic institutions in the South, Southeast and Northeast Brazil
- includes men and women aged 35 to 74 years old
- uses a multi-disciplinary approach to assess risk of chronic diseases and their determinants in the cohort

Motivation- ELSA

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

- Reference population has three strata:
  - professional
  - technical
  - administrative
- The strata are related to the socioeconomic position
  $\Rightarrow$ The strata are related to the outcomes.
- The strata are not self-weighted

The Simulation Exercise

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

In this study we simulated a dataset based on ELSA population, to estimate the potential bias of ignoring the sampling design in a few different scenarios, based on effect measures of three exposure variables .
We will call them *smoking* , *hypertension and access (to health system)*.

The Simulation Exercise

- 52,750 observations
- number of people in each stratum $\longrightarrow$ given by the number given by the academic institutions
- indicator of smoking for each individual in the population
- smoking prevalence for each working situation $\longrightarrow$ based on similar population (Pró-saúde/UERJ project)
- indicator of hypertension for each individual in the population
- hypertension prevalence for each working and smoking situation given by Pró-saúde

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

The Simulation exercise

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

- indicator of access to health system
  - related to strata
  - not related to smoking
- time to myocardial infarction based on Spanish data
- time data $\longrightarrow$ Weibull distribution
- 2,000 samples, with 15,000 individuals
- two schemes simple and not self-weighted design

Scenarios

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

We proposed four different scenarios

Table: Parameters used to simulate the datasets

Scenario	Smoking	Administrative	Technical	Hypertension	Access
1	2.00	3.00	1.50	0.00	0.00
2	$3Adm + 2Tech + 1.5Prof$	3.00	1.50	0.00	0.00
3	2.00	3.00	1.50	1.50	0.00
4	2.00	3.00	1.50	0.00	1.50

Data analysis

We analyzed all samples supposing that the true model was the additive model with *smoking* and *strata*, that is, the one given by scenario 1

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

What we expected...

- scenario 1 $\longrightarrow$ no design effect
- scenario 2 $\longrightarrow$ expected design effect
- scenario 3 $\longrightarrow$ expected design effect
- scenario 4 $\longrightarrow$ We are not sure

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

We also simulated the loss of follow up considering all the scenarios.

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

We also simulated the loss of follow up considering all the scenarios.
We considered 2 schemes of losses:

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Schemes of losses

- Scheme 1: random loss 15%
- Scheme 2: differential loss
  - Professional: 8%
  - Technical : 12 %
  - Administrative: 20 %

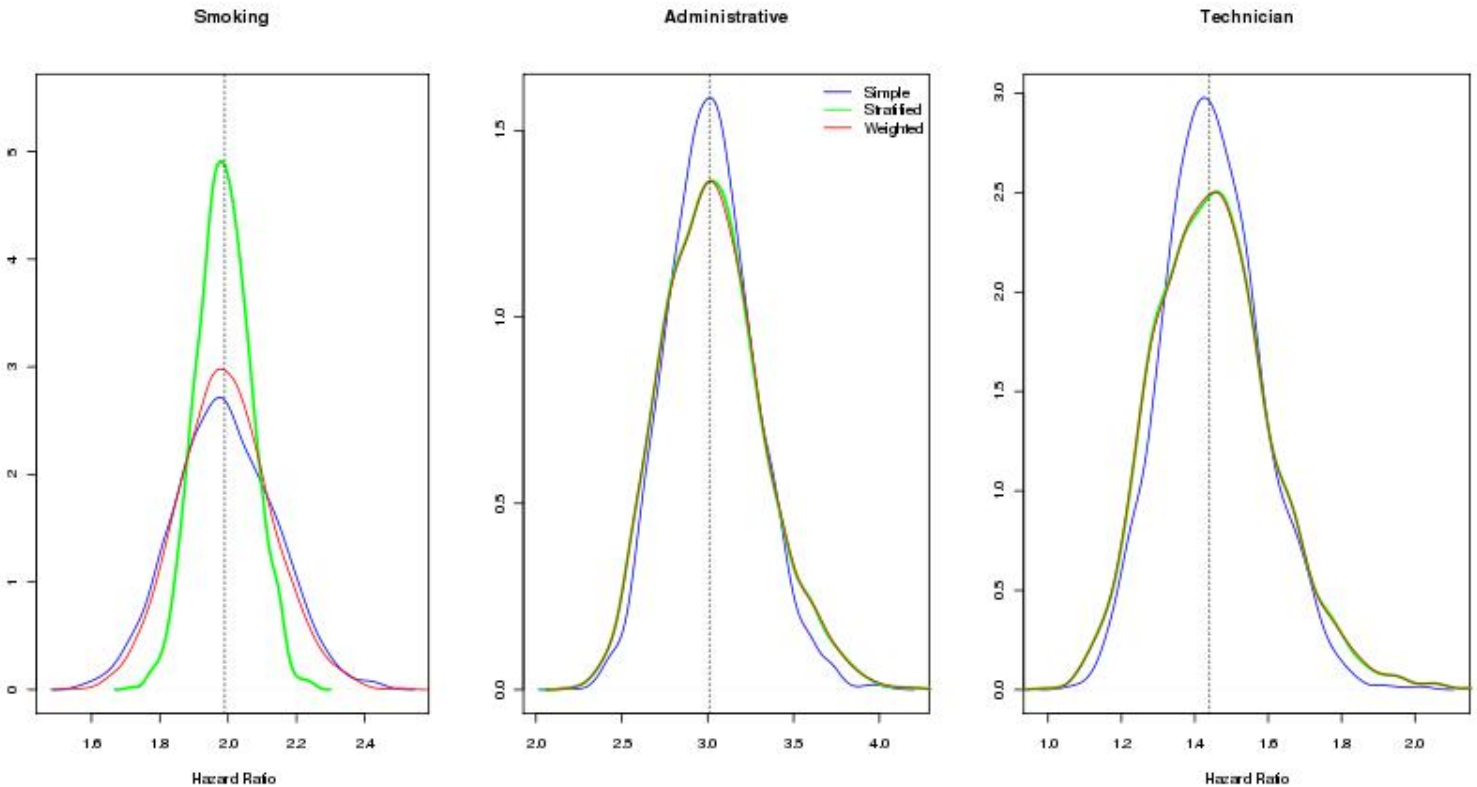
We expected to find design effect in both schemes

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Right Model

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

If you adjust the right model, sample plan doesn´t matter

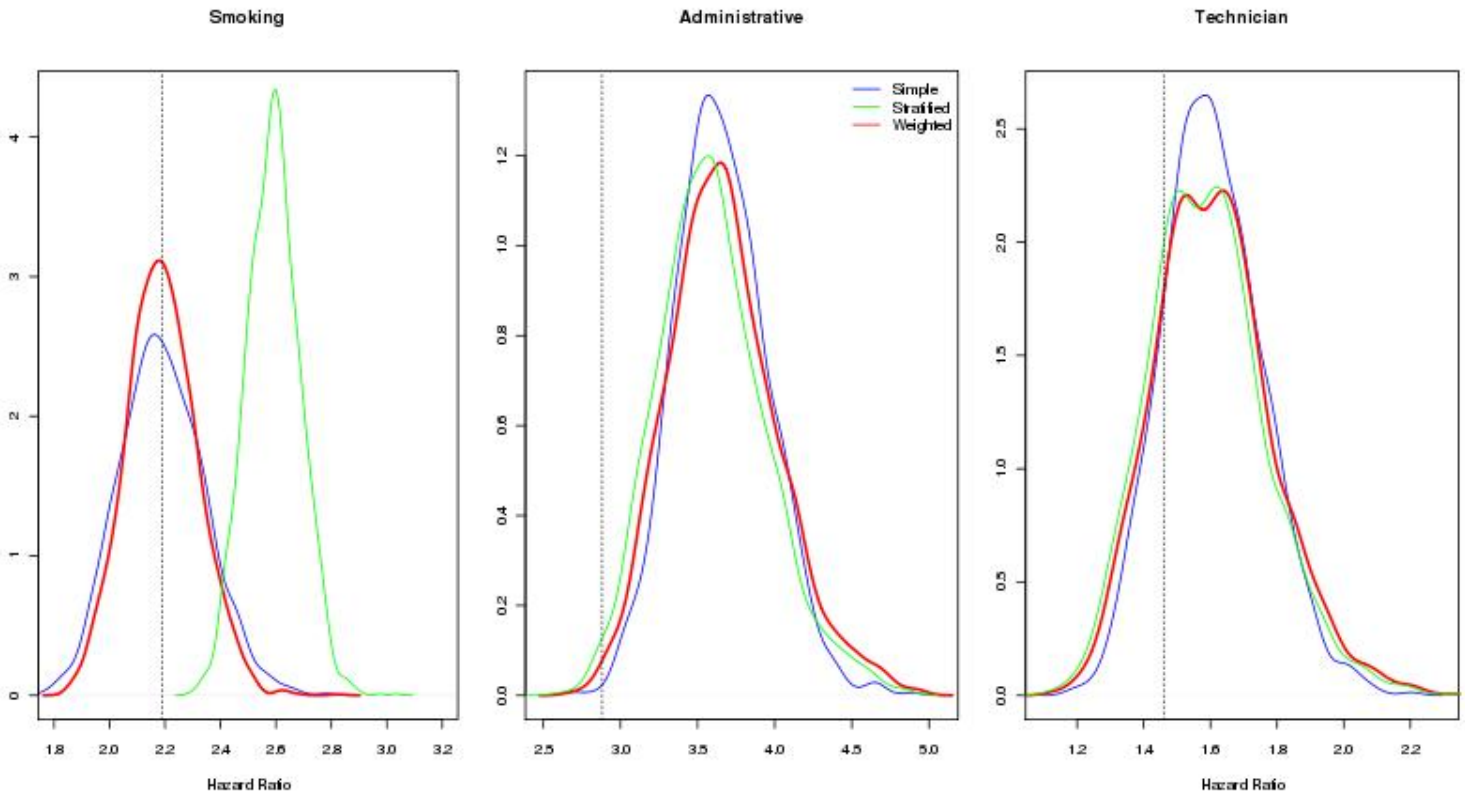


2*Smoking + 3*Administrative + 1.5*Technician

Omitted Interaction

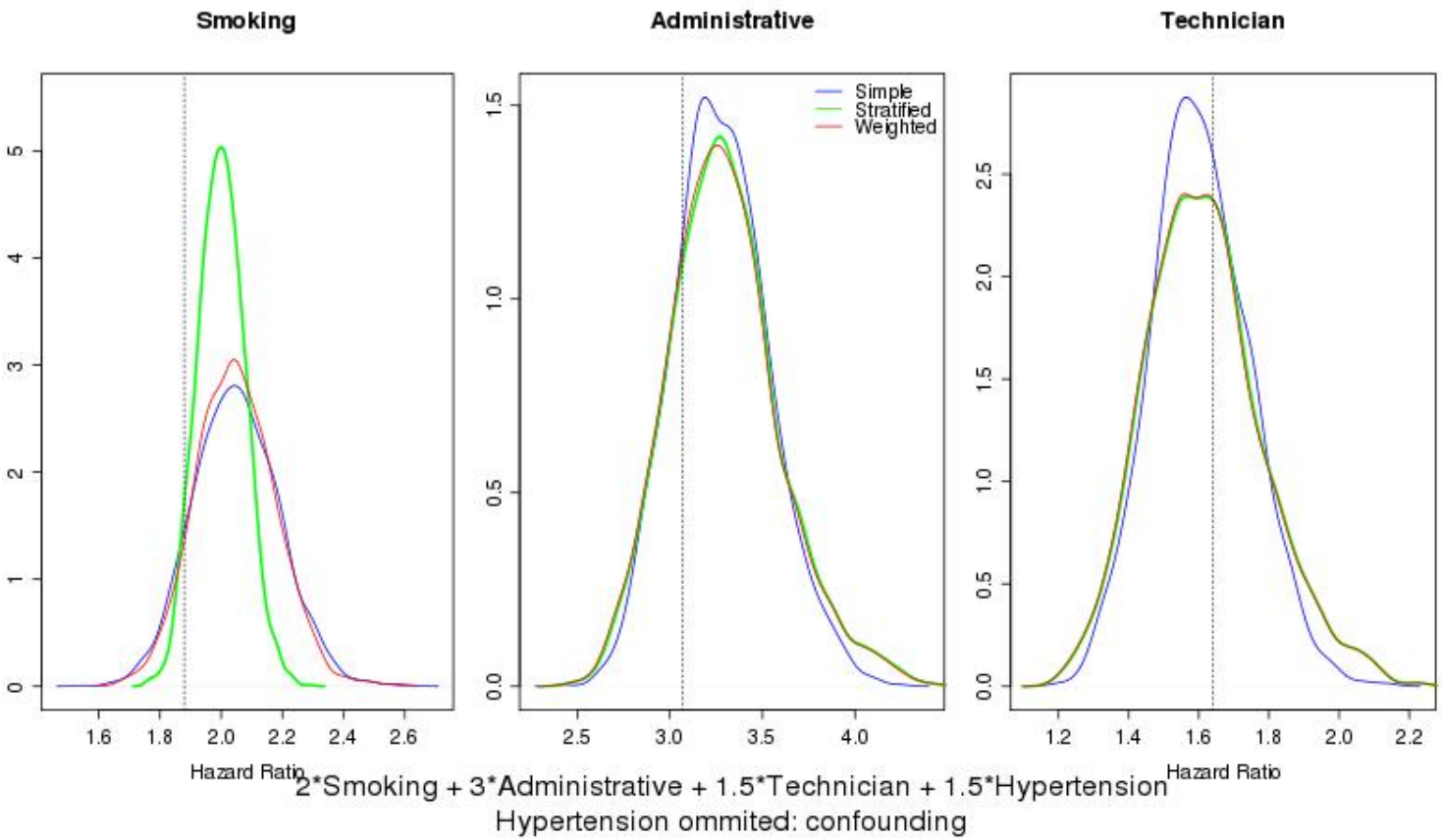
Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

If you have interaction and you don´t consider it in your model, then the sample plan is very important



Omitted covariate associated with the risk factor

If you omit a covariate associated with the risk factor it seems that the sample plan is irrelevant. You always fail!

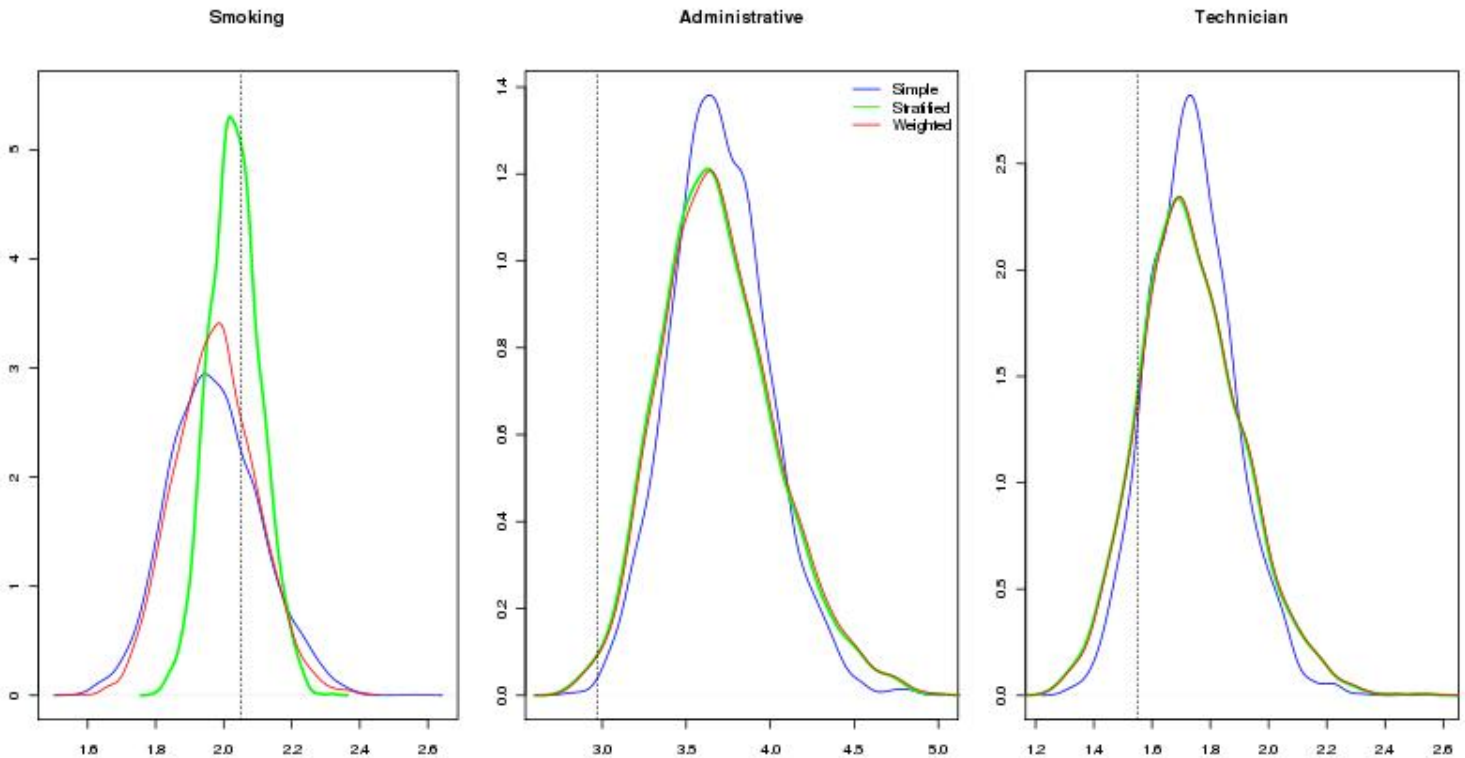


Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Omitted covariate not related to the risk factor

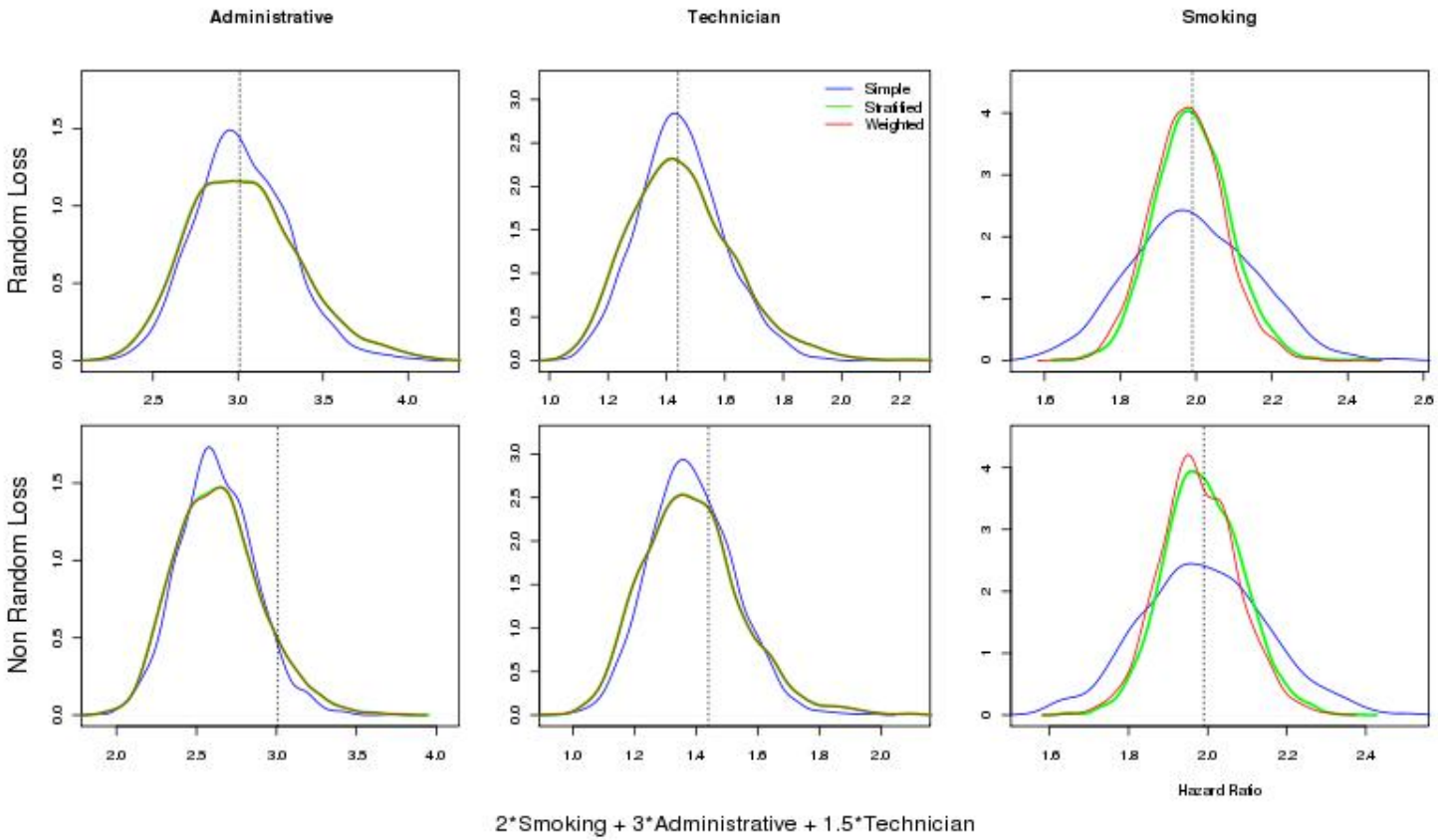
Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

If you omit a covariate not related to the risk factor it seems that the stratified plan does a better job



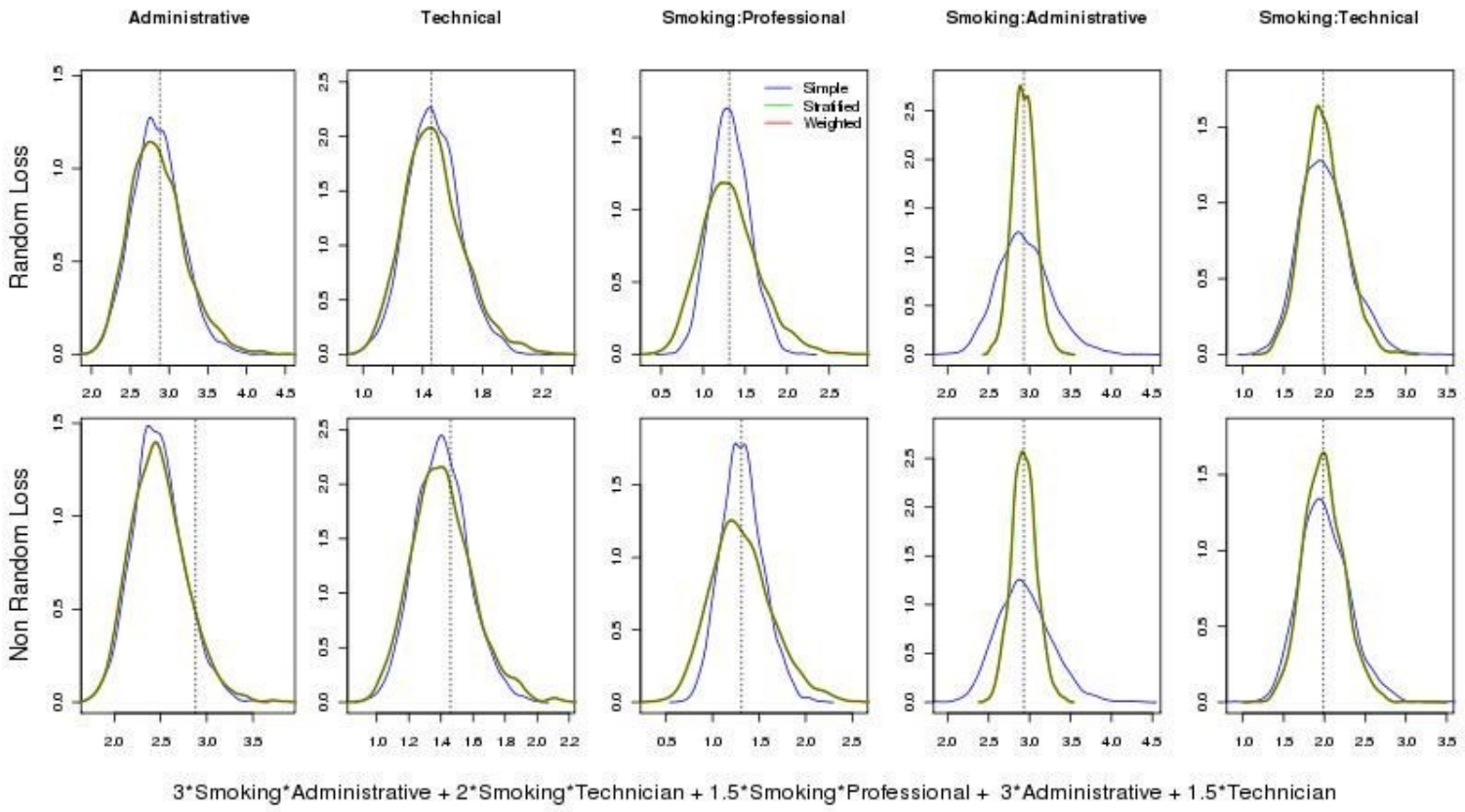
Loss - Right Model

If you adjust the right model, you don´t have bias, even if you have loss



Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Loss - Scenario 2 - Right Model

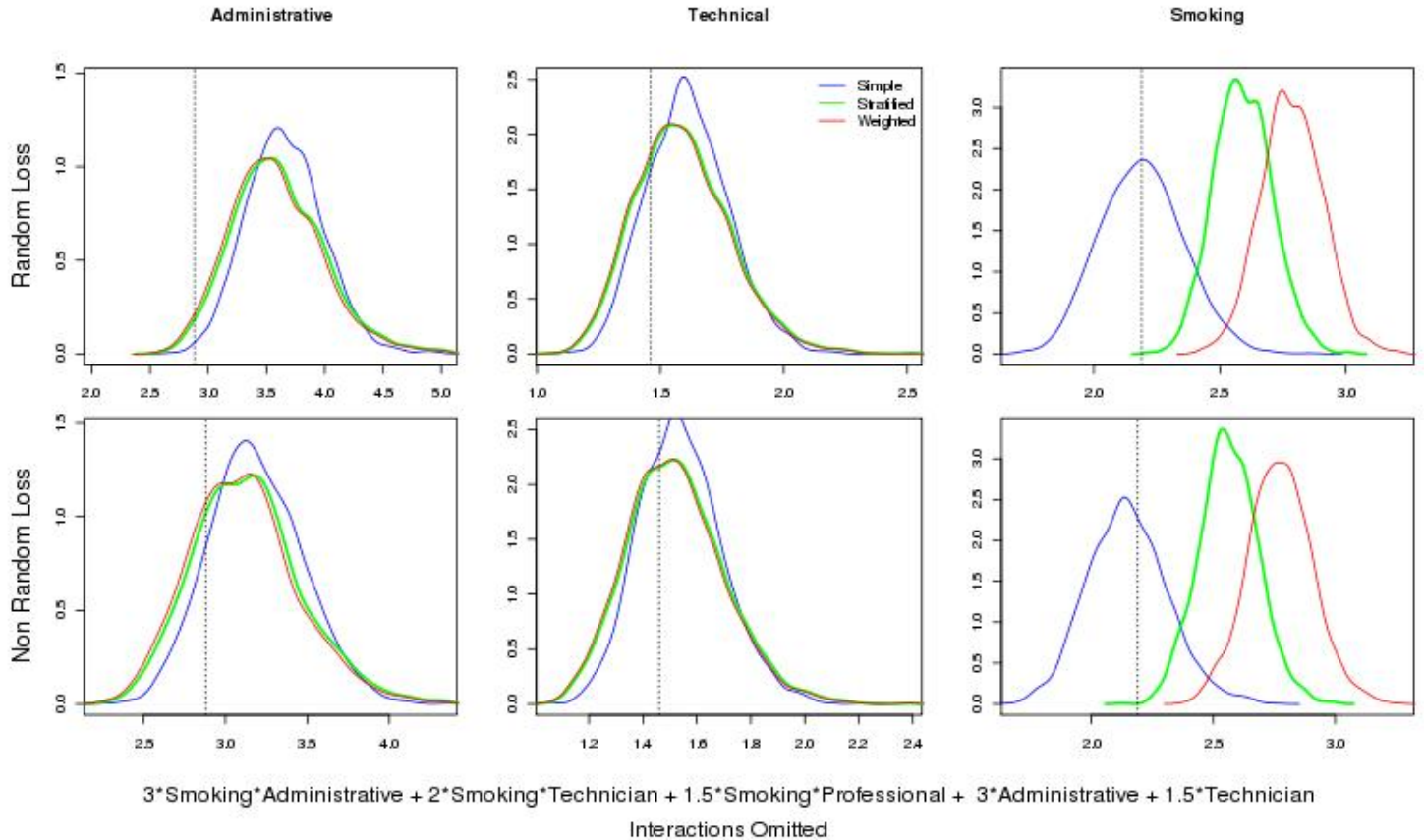


Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Loss - Scenario 2 -Omitted Interaction

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

If you have loss, even if you weigh the sample, you get biased results



Final Conclusions

Cibele Comini César
Marilia Sá Carvalho
Cosme Marcelo Furtado Passos da Silva
UFMG / FIOCRUZ

Which strategy is the best?

- There isn´t one strategy that is always the best
- It depends on the situation
- The stratified sampling gives the least MSE in a lot of situations
- It seems that the stratified sampling is less robust
- If you have loss $\longrightarrow$ use a proper model to analyse the data