

## Table of contents

|                                                            |       |
|------------------------------------------------------------|-------|
| Acknowledgements.....                                      | iii   |
| Résumé et mots clés.....                                   | iv    |
| Summary and keywords.....                                  | vii   |
| Table of contents.....                                     | ix    |
| List of tables.....                                        | xi    |
| List of figures.....                                       | xiii  |
| List of acronyms and abbreviations .....                   | xiv   |
| <br>1 Introduction.....                                    | <br>1 |
| 2 Literature review .....                                  | 4     |
| 2.1 Overview of epidemiological study design.....          | 5     |
| 2.2 Overview of logistic regression .....                  | 7     |
| 2.3 Overview of survival analysis.....                     | 9     |
| 2.3.1 Survival and hazard functions.....                   | 9     |
| 2.3.2 Parametric survival models.....                      | 10    |
| 2.3.3 The Cox semi-parametric model.....                   | 12    |
| 2.3.4 Adaptation of Cox's model to case-control study..... | 15    |
| 3 Objectives .....                                         | 19    |
| 4 Methods.....                                             | 22    |
| 4.1 New proposed Cox models .....                          | 23    |
| 4.2 Simulation study .....                                 | 27    |
| 4.2.1 Overview.....                                        | 27    |
| 4.2.2 Generation of the source population .....            | 27    |
| 4.2.3 Simulation of case-control studies .....             | 30    |
| 4.2.4 Summary of the different scenarios investigated..... | 31    |

|       |                                                                                         |    |
|-------|-----------------------------------------------------------------------------------------|----|
| 4.2.5 | Data analytical models .....                                                            | 33 |
| 4.2.6 | Summary statistics to evaluate the performance of the different analytical models ..... | 34 |
| 4.3   | Real data analysis:.....                                                                | 35 |
| 4.3.1 | Data source.....                                                                        | 35 |
| 4.3.2 | Method to handle missing smoking data .....                                             | 36 |
| 4.3.3 | Description of smoking intensity trajectories over time .....                           | 39 |
| 4.3.4 | Data analytical models .....                                                            | 41 |
| 5     | Results.....                                                                            | 49 |
| 5.1   | Results from simulations.....                                                           | 50 |
| 5.2   | Results from real data analysis .....                                                   | 54 |
| 5.2.1 | Description of real data.....                                                           | 54 |
| 5.2.2 | Patterns of smoking intensity.....                                                      | 56 |
| 5.2.3 | Results from regression models.....                                                     | 58 |
| 6     | Conclusion and discussion .....                                                         | 66 |
| 7     | References.....                                                                         | 71 |
| 8     | Appendix.....                                                                           | 75 |
| 8.1   | Questionnaire on smoking .....                                                          | 76 |
| 8.2   | Simulation results from Leffondré 2003 .....                                            | 77 |