Chapter 3

Imputation and Bootstrap Methodologies

3.1 Introduction

In recent years, the human immunodeficiency virus (HIV) has had a considerable influence on the incidence of tuberculosis (TB) in both developed and developing countries (Raviglione et al., 1995; World Health Organization, 1989) In fact, the World Health Organization estimates that of the 8 million new cases recorded in 1990, 4% were HIV-infected, a percentage that is expected to rise to 15% in the year 2000 (World Health Organization, 1994). The high mortality rate among TB HIV-infected patients (TB/HIV), compared with those without HIV infection, suggests that immunodeficiency due to HIV is the cause of such figures (Richter et al., 1995). Thus it is estimated that in 1996 alone some 226 000 TB/HIV patients will have died worldwide, and a third of the HIV fatalities can be directly attributed to TB (World Health Organization, 1996).

It has been demonstrated that in TB/HIV patients, the depletion of T CD4+ lymphocytes constitutes a good marker for the progression of the disease (Jones et al., 1993), which seems to gain ground more rapidly among those individuals with high replication rates of the virus (Bravo et al., 1995; Cao et al., 1995).

Furthermore, many developing countries with high HIV and *Mycobacterium tuberculosis* coinfection rates (De Cock, 1995), lack the necessary resources to carry

out laboratory tests such as the determination of lymphocyte subsets or the viral load. This is one of the reasons why the use of clinical parameters, with inexpensive and simple tests –such as the tuberculin skin test– or straightforward and practical laboratory markers, for estimating survival probability would be desirable.

Survival studies on TB/HIV patients are few and based on complete data; that is to say, without taking into account missing data (Richter et al., 1995; Jones et al., 1993; Caylà et al., 1993a). However, when the problem one wishes to analyze depends on the behavior of covariates with missing data, the treatment of such data is crucial both for conclusions that may be reached and for decisions that must be taken based on those conclusions. The aim of this chapter, therefore, is to show a survival study that it was carried out by analyzing the predictive role played by variables such as the tuberculin skin test and T CD4+ lymphocytes percentage, which both present a large amount of missing data. To this end, the bootstrap methodology (Efron, 1994) and a multiple imputation scheme (Little and Rubin, 1987; Rubin, 1987) will be applied as an initial approach to the methods of treating variables with missing data.

3.2 Methods

In Barcelona, both acquired immunodeficiency syndrome (AIDS) and TB are covered by the active epidemiological surveillance system of the Epidemiology Service of the Municipal Institute of Health. This service gathers data provided by doctors, as well as results of microbiological analysis, hospital discharges, mortality data and linkage between AIDS and TB Registers.

3.2.1 Definitions

Tuberculosis. A patient was considered to be tuberculous if he or she had been receiving chemotherapy treatment for TB, and if such treatment had been administered for a pre-established period of time, unless it had been interrupted by either the appearance of secondary effects or by death of the patient. In cases of pleural TB, the anatomical-pathological and/or enzymatic criteria (Grupo de Trabajo sobre Tuberculosis (FIS-MSC), 1992; Ocaña et al., 1996) were also considered.

Tuberculin skin test. This was performed according to the Mantoux technique, applying 2 TU of PPD RT–23 intradermically, and a reading taken between 48 to 72 hours. Indurations of 5 mm or more were considered positives (Grupo de Trabajo sobre Tuberculosis (FIS-MSC), 1992).

HIV. A patient was considered HIV-infected if the ELISA (enzyme-linked immunosorbent assay) test gave positive at the time of TB diagnosis, and if this result was confirmed by Western blot (Centers for Disease Control, 1993).

3.2.2 Data

The study data belongs to the HIV-infected patients with pulmonary tuberculosis (PTB), with or without extrapulmonary infection, residents in Barcelona city, and diagnosed between January 1st 1992 and December 31st 1994. If the culture identified an atypical mycobacteria, the patient was excluded from the study. Patients with a medical record of TB were included only if they had received no treatment against TB for more than a year.

The survival time for each patient was established as the number of days between the date of TB diagnosis and death –or December 31st 1994 for those patients who were still alive on this date–. Consequently, those patients alive when the study was concluded, or when loss of follow-up occurred, are considered to be right censored because death has not been observed.

3.2.3 Study variables

In addition to the survival time and vital status, other variables were recorded for each patient: sex, age at the TB diagnosis, district of residence, prison history, HIV transmission group, alcohol addiction, homelessness, result of tuberculin test (PPD: negative=0, positive=1), percentage of lymphocyte subsets ($CD4$ and $CD8$), microbiological results, radiological pattern and site of TB. The $CD4$ and $CD8$ variables, measured in percentages, were chosen rather than the respective absolute counts, because they have lower variability and also because it has been shown that a high correlation exists between the T CD4+ absolute number and the T CD4+ percentage (Phair et al., 1990).

3.2.4 Missing data problem

The percentage of missing data was in general lower than 10%. However, there was a large amount of missing data in both the variable $CD4$ (38.9%) and in the PPD variable (50.4%). Only 31.8% of the data is complete in these two variables. The methodology we set forth in the following subsection enables us to make full use of the information found in the other 68.2%, where missing data is present.

3.2.5 Statistical analysis

It is carried out in two stages (Serrat and Gómez, 1995): a) Univariate and multivariate analysis of the complete data, and b) the treatment of missing data and the parametric estimate by multiple imputation.

a) In the complete data univariate analysis, the Kaplan–Meier method is used to estimate the survival curves, and the Cox regression model is used to analyze the role of the various categories within the covariates. For each variable, the model is validated by studying the significance of the interaction of the variable with the time, and with the logarithm of the time, assuming the hypothesis of proportional hazards if such interactions are not significant (Collett, 1994). For the multivariate analysis, the Cox regression model is set with those covariates that are significant from the univariate analysis.

   We define the *observed subsample* as that consisting of the 157 (31.8%) patients for whom the $CD4$ and PPD values are available. We dichotomize $CD4$ variable by assigning a value 0 for those patients with a percentage of T CD4+ cells lower than or equal to 14% –*this value corresponds to 200 T CD4+ cells/mm^3, which is the index below which a patient is judged to have AIDS (Centers for Disease Control, 1993)*; otherwise the $CD4$ variable is assigned a value 1.

   To obtain more precise estimates, we fit a Weibull regression model for each immunosuppression level. The hazard function at moment t for a patient with PPD value in the tuberculin test is given by the expression

$$h(t; PPD) = \frac{1}{\sigma t} \left(e^{-\beta} t\right)^{\frac{1}{\sigma}}$$

with $\beta = \beta_0 + \beta_1 PPD$ and $\sigma = \sigma_0 + \sigma_1 PPD$, where β_0 and σ_0 are the values for the reference group (negative tuberculin). The parameters for the positive tuberculin group are $\beta_0 + \beta_1$ and $\sigma_0 + \sigma_1$, respectively.

The fitting, graphically validated, provides, among others, the relative π-percentile ($0 \leq \pi \leq 1$) estimate for a positive tuberculin patient respect to the reference group,

$$pR(\pi; PPD = 1) = e^{\beta_1} \left[-\log(1-\pi)\right]^{\sigma_1},$$

as a measure of time elongation (*dilation or contraction*) between categories.

b) It is assumed that the response process of the covariates $CD4$ and PPD is Missing At Random (MAR) (Little and Rubin, 1987), that is, that the probability of observing the values of $CD4$ and PPD does not depend on the missing values of these variables. This assumption has been graphically validated.

Let $\mathbf{X}_0$ denote the initial incomplete sample, and θ the relevant parameter. To estimate θ, we proceed according to the following scheme:

$$\mathbf{X}_0 \overset{\text{bootstrap}}{\longrightarrow} \mathbf{X}_b \overset{\text{imputation}}{\longrightarrow} \widehat{\mathbf{X}_b} \overset{\text{estimation}}{\longrightarrow} \widehat{\theta_b}$$

which is repeated 200 times. At each time b, $b = 1, \cdots, 200$, we obtain a bootstrap replica of the initial sample $\mathbf{X}_0$, which we denote by $\mathbf{X}_b$, then, by means of the bilinear imputation method (Efron, 1994), we get a completed sample, $\widehat{\mathbf{X}_b}$, and, finally, using a Weibull regression model, we obtain the corresponding estimate of θ, $\widehat{\theta_b}$. This procedure enable us to obtain 200 estimates $\widehat{\theta_1}, \cdots, \widehat{\theta_{200}}$. The estimator we proposed by multiple imputation is

$$\hat{\theta} = \frac{1}{200} \sum_{b=1}^{200} \widehat{\theta_b}$$

with a variance composed of two parts: the *intra* variance due to the imputation itself and the *between* variance due to the different imputations (Little and Rubin, 1987; Rubin, 1987). So, we can obtain confidence intervals for the relative percentiles between groups.

For the basic statistical analysis, SPSS-PC (Norussis, 1986) and EGRET (EGRET, 1990) software have been used. The bootstrap sampling and the bilinear imputation method have been implemented in PASCAL language (see Appendix I for more details), and carried out on a Sun SPARC work station in a UNIX environment.

3.3 Results

3.3.1 Complete data analysis

The cohort under study was made up of 494 TB/HIV-infected patients. The median age was 32 years, ranging from 17 to 66 years of age. 413 (83.6%) were men. 151 (30.8%) came from the lowest socio-economic district. 146 (29.6%) were or had previously been in prison at the beginning of the study. By HIV transmission category, 358 (72.5%) were exclusively IVDU, 63 (12.8%) were exclusively homosexual men, 31 (6.3%) heterosexual, 11 (2.2%) IVDU and homosexual, only 5 (1.0%) hemophilic hemotransfused and 26 (5.3%) unknown. With regard to TB site, 377 (76.3%) were cases of PTB and the rest were pulmonary with extrapulmonary involvement. 102 (20.6%) patients had TB history. The radiological pattern was normal in 30 (6.2%) cases, anormal non-cavitary in 357 (73.9%) cases, and cavitary in 96 (19.9%). The microbiological result was negative in 82 (18.5%) patients, smear positive in 193 (43.6%), and the culture exclusively positive in 168 (37.9%). The tuberculin skin test was available in 245 (49.6%) cases; of these, 96 (39.2%) were positive. The percentage of T CD4+ cells was available in 302 (61.1%) cases. The median of CD4+ percentages was 11%, oscillating these values between 1% and 93%, the first and third quartile were 5% and 18% respectively, and the mean was 14% (SD: 13.1%). 63.8% (315 cases) were censored and, among them, only 23 (4.7%) corresponded to lost of follow-up patients. The median follow-up time for the censored patients was 455 days.

Figure 3.1 shows the Kaplan-Meier survival estimates according to the immunosuppression level and the result of the tuberculin skin test for the observed subsample. Note that while the first quantile and the median can be evaluated for the most immunosuppressed group and for the negative tuberculin group, for the other two groups the median cannot be explicitly derived.

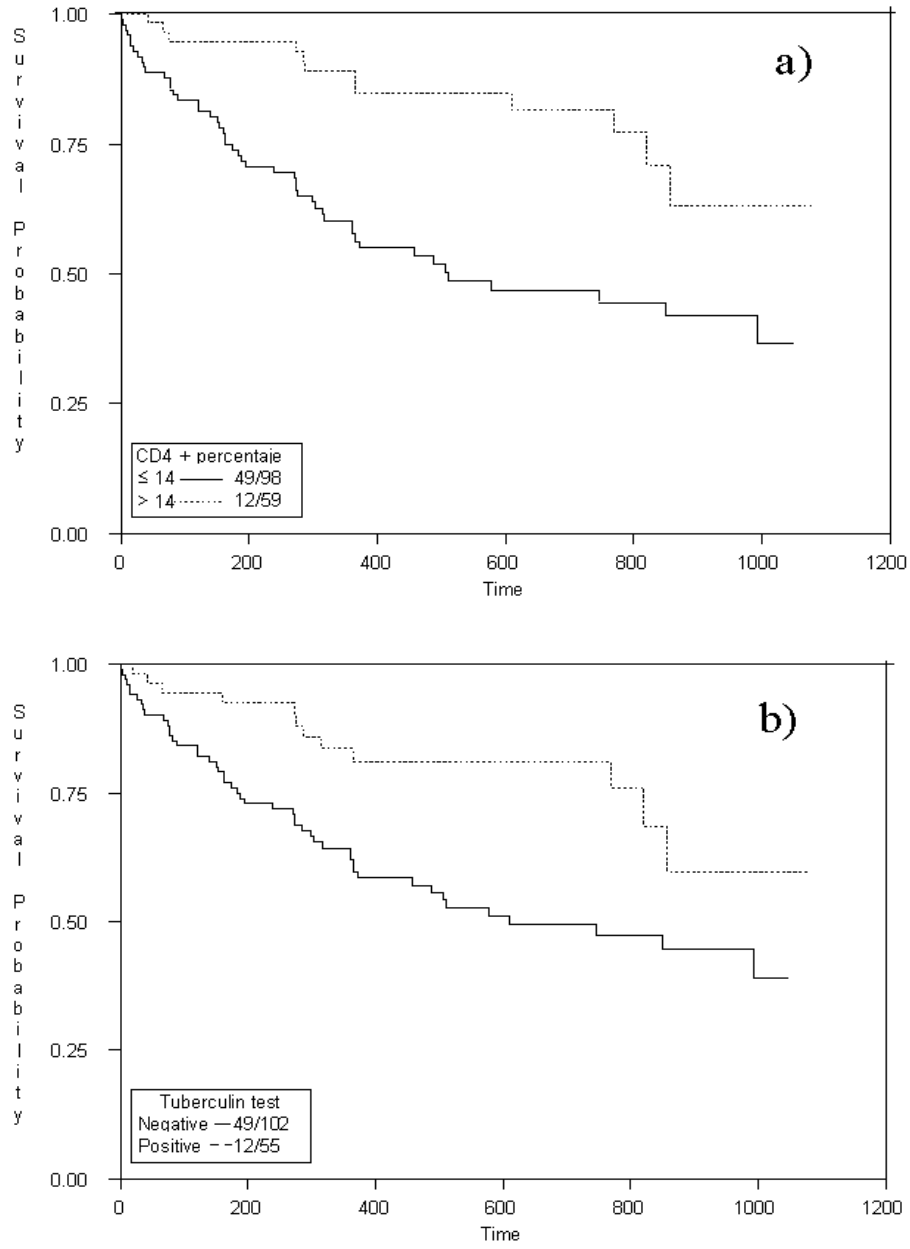

Figure 3.1: *Kaplan–Meier estimates of survival function for Pulmonary TB HIV-infected patients, Barcelona (1992-1994), for which CD4+ lymphocytes % and tuberculin test are available (n=157), stratified by: a) CD4+ lymphocytes %, and b) tuberculin test result. m/n, the proportion of deaths*

Variable categories	Observed cases	Relative hazard*	95% CI	p-value†
Sex	494			
Female		1		
Male		0.9603	0.66–1.41	0.836
Age (years)	494			
Continuous variable		1.035	1.02–1.05	< 0.001
Age (years)	494			
≤ 28		1		
29–32		1.137	0.75–1.73	0.546
33–36		0.9368	0.59–1.48	0.778
≥ 37		1.671	1.12–2.48	0.011
District of residence	491			
Others		1		
Inner city		1.049	0.76–1.44	0.768
Transmission group	468			
No IVDU‡		1		
IVDU		0.7962	0.57–1.12	0.188
Alcohol addiction	494			
No		1		
Yes		1.107	0.82–1.49	0.508
Homelessness	494			
No		1		
Yes		0.7014	0.33–1.49	0.358
Tuberculin test result	**245**			
Negative		1		
Positive		0.4064	0.24–0.68	< 0.001
CD4+ lymphocytes %	**302**			
Continuous variable		0.9493	0.93–0.97	< 0.001
CD4+ lymphocytes %	**302**			
≤ 14		1		
> 14		0.3537	0.23–0.55	< 0.001
CD8+ lymphocytes %	**299**			
Continuous variable		0.9929	0.98–1.01	0.273
Smear	443			
Negative		1		
Positive		0.9220	0.60–1.42	0.712
Positive culture only		1.324	0.86–2.03	0.197
Radiological pattern	483			
Normal		1		
Cavitary		0.8605	0.39–1.89	0.709
Non cavitary		1.359	0.67–2.77	0.399
Site TB	494			
Pulmonary		1		
Mixte		1.051	0.75–1.47	0.771

Table 3.1: *Estimated relative hazards in Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*
* *Relative hazards are obtained using a validated univariate Cox proportional hazards model for each variable.* † *Wald test.* ‡ *Intravenous drug users*

Table 3.1 presents the results of the univariate Cox model. It can be observed that only the T CD4+ percentage, the result of the tuberculin test and age are significant.

For the multivariate analysis in the observed subsample ($n = 157$), we only consider the covariates $CD4$ (continuous and categorized), PPD and age, and only the dichotomized $CD4$ variable is found to be significant (see Table 3.2). The estimated relative hazard is 0.3198 (95% CI: 0.17–0.60) for the least immunosuppressed group with respect to the most immunosuppressed group.

If we divide the observed subsample into two groups according to the immunosuppression level, and if in each group we analyze the effect of the PPD, $CD4$ and age covariates, we observe that in the most immunosuppressed group there is a significant relative hazard 0.3657 (95% CI: 0.13–1.02; $p = 0.054$) for patients with positive tuberculin, against those with negative tuberculin. However, in the least immunosuppressed group, and after adjusting for all the mentioned covariates, there is no significant difference between the cases (see Table 3.3).

After fitting a Weibull model, that has been graphically validated, for the whole sample the maximum likelihood estimate of β and σ, before including the covariates in the model, is 7.364 and 1.411, respectively. Analogously, for the observed subsample we obtain 7.318 and 1.326, respectively. Table 3.3 also shows the β and σ estimates according to the level of immunosuppression, both before and after recording the effect of the PPD covariate. For the most immunosuppressed group we verify the protective effect of positivity in the tuberculin skin test by estimating the PPD coefficient of β, $\beta_1 = 8.100 - 6.681 = 1.419$ (SE: 0.739, $p = 0.055$) and a constant relative percentile value of $pR = 4.1329$ (95% CI: 0.97–17.59), whereas for the PPD covariate in the least immunosuppressed group, no significance is obtained. The corresponding percentile estimates, expressed in days, for the different relevant subsamples can be seen in Table 3.4. So, for example, the 20% percentile is estimated at 190 days for the whole sample, and at 206 days for the observed subsample, before any covariate adjustments are made. Nevertheless, if we consider the most immunosuppressed group, and we adjust for the tuberculin test, we estimate the 20% percentile at 101 days for the negative tuberculin patients, and at 417 days for the positive tuberculin ones. Finally, in the least immunosuppressed group, the information furnished by the tuberculin test is not significant, and we are able

	Variables in the model	
	None	Dichotomized CD4% *
Variable	p-value†	p-value†
Age (years)		
Continuous variable	0.140	0.657
Age (years)	0.586‡	0.732‡
≤ 28	–	–
29–32	0.710	0.363
33–36	0.733	0.706
≥ 37	0.315	0.826
Tuberculin test result	0.004‡	0.129‡
Negative	–	–
Positive	0.004	0.129
CD4+ lymphocytes %		
Continuous variable	0.002	0.468
CD4+ lymphocytes %	< 0.001‡	–
≤ 14	–	–
> 14	< 0.001	–

Table 3.2: *p-values on fitting a multivariate Cox proportional hazards model (n=157) to the Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*
* *This variable is selected in the multivariate Cox model because it has the lowest p-value.* † *Wald test.* ‡ *Overall score test on all degrees of freedom of the variable*

Variable categories	Relative hazard*	95% CI	p-value†	β‡	σ‡
Observed subsample⋆ (n=157)					
CD4+ lymphocytes %					
≤ 14	1			6.872	1.383
> 14	0.3198	0.17–0.60	< 0.001	7.737	0.8816
CD4+ % ≤ 14 subsample (n=98)					
Tuberculin test result					
Negative	1			6.681	1.378
Positive	0.3657	0.13–1.02	0.054	8.100	1.378
CD4+ % > 14 subsample (n=59)					
Tuberculin test result					
Negative	1			7.737	0.8816
Positive	1.3032	0.39–4.33	0.666	7.737	0.8816

Table 3.3: *Estimated relative hazards and parameters estimates on fitting a Weibull model to the Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*
* *Relative hazards estimates after adjusting a multivariate Cox proportional hazards model.* † *Wald test.* ‡ β *and* σ *Weibull maximum likelihood estimates.* ⋆ *The observed subsample is integrated by the patients for whom CD4+ lymphocytes % and tuberculin test result are available*

to estimate this percentile at 611 days. Figure 3.2 shows both the estimate of the survival function for each of the different relevant subsamples, and the time elongation between categories; one may observe that the group showing greatest survival is that made up of the most immunosuppressed patients with positive tuberculin.

			$CD4\% \leq 14$ subsample		
Percentile (%)	All the	Observed	Negative	Positive	$CD4\% > 14$
(100 π)	sample	subsample	PPD	PPD	subsample
10	66	76	36	148	315
20	190	206	101	417	611
30	368	384	193	796	923
40	612	618	316	1306	1268
50	941	927	481	1988	1659
60	1395	1342	707	2921	2122
70*	2051	1928	1029	4255	2699

Table 3.4: *Estimated percentiles (in days) of the distributions of survival times on fitting a Weibull model to the Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*
* *The Weibull regression model has been only validated until the 70th percentile of survival time*

3.3.2 Missing data analysis

Figure 3.3a shows the survival curves according to whether or not the CD4+ percentage has been observed. Analogously, Figure 3.3b shows the survival curves according to the observed value of the PPD variable.

After obtaining 200 completed bootstrap replicas ($n = 494$) and estimating the coefficients (β_1 and σ_1) of the PPD variable for each one, and for each immunosuppressed level, while the scale parameter, β_1, is found to be significant for most of the replicas in the most immunosuppressed group, the shape parameter, σ_1, is not. It was therefore decided to fit a Weibull model dependent on a scale β_1 parameter for each of the 200 samples. In this way, 200 β_1 estimates were obtained and the average proposed as estimator. An analogous process was carried out for the least

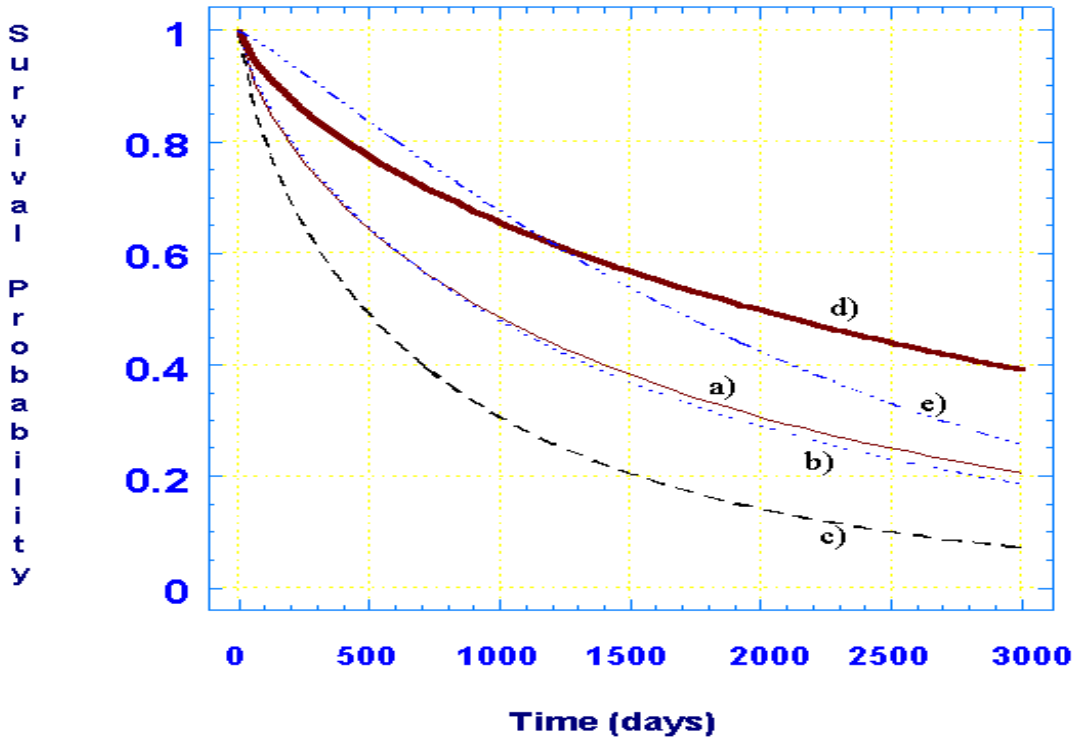

Figure 3.2: *Estimated survival functions for Pulmonary TB HIV-infected patients, Barcelona (1992-1994), on fitting a Weibull model to: a) all the sample (n=494), b) cases for whom CD4+ % and tuberculin test are available (n=157), c) cases with CD4+ % $\leq$ 14 and negative tuberculin (n=80), d) cases with CD4+ % $\leq$ 14 and positive tuberculin (n=18), and e) cases with CD4+ % $>$ 14 (n=59)*

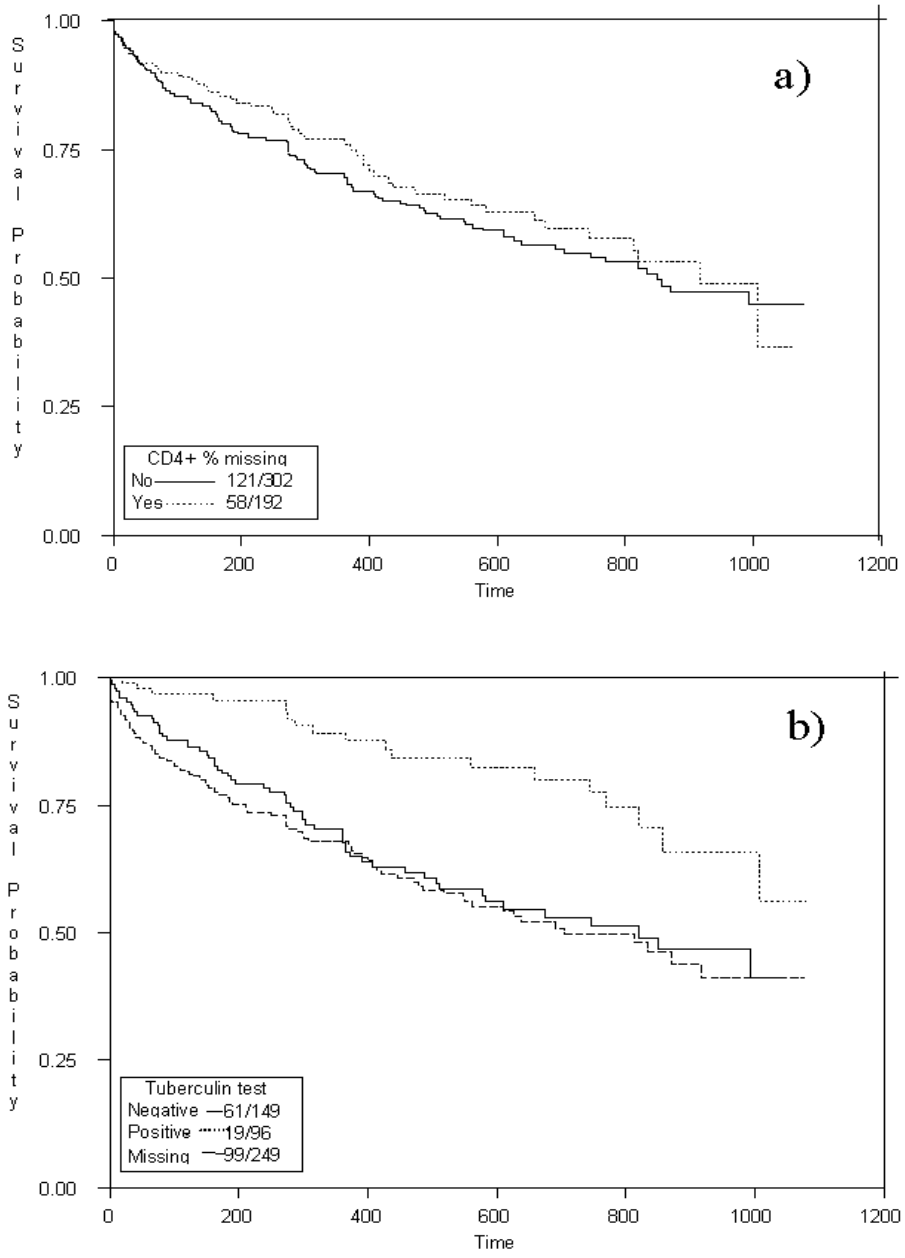

Figure 3.3: *Kaplan–Meier estimates of survival function for Pulmonary TB HIV-infected patients, Barcelona (1992-1994), stratified by: a) whether or not CD4+ lymphocytes % is available, and b) tuberculin test result. m/n, the proportion of deaths*

immunosuppressed group. In this case the shape parameter σ_1 was the parameter of greatest interest. Likewise, the average of the 200 σ_1 estimates are proposed as the shape parameter estimator. Table 3.5 shows the results of this methodology.

Subsample	Parameters and Relative percentile*	Estimate ± SE	95 % CI	p-value†
$CD4\% \leq 14$	β_1	1.7012 ± 0.9680	−0.20 to 3.60	0.079
	pR (constant)	5.48	0.82–36.55	
$CD4\% > 14$	σ_1	−0.6658 ± 0.5527	−1.75 to 0.42	0.228
	pR $100\pi = 30$	1.99	0.65–6.07	
	$100\pi = 50$	1.28	0.86–1.90	
	$100\pi = 70$‡	0.88	0.72–1.08	

Table 3.5: *Parameters estimate and estimated relative percentiles of the distributions of survival times on fitting a Weibull model to the Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*

* *Parameters estimates are obtained via multiple imputation using a Weibull regression model for each subsample in each imputated bootstrap replica and relative percentiles are for a positive tuberculin patient with respect to a negative one.* † *Wald test.* ‡ *The Weibull regression model has been only validated until the 70th percentile of survival time*

Specifically, the β_1 estimate differs significantly from 0 for the most immunosuppressed patients, and we can estimate a constant relative percentile 5.48 ($p = 0.079$) for the positive tuberculin group compared with the negative tuberculin group. Furthermore, the σ_1 estimate does not differ significantly from 0 for the least immunosuppressed patients. Finally, Figure 3.4 shows the confidence intervals for the aforementioned relative percentiles. The results obtained for the two immunosuppression levels are presented separately in Table 3.6.

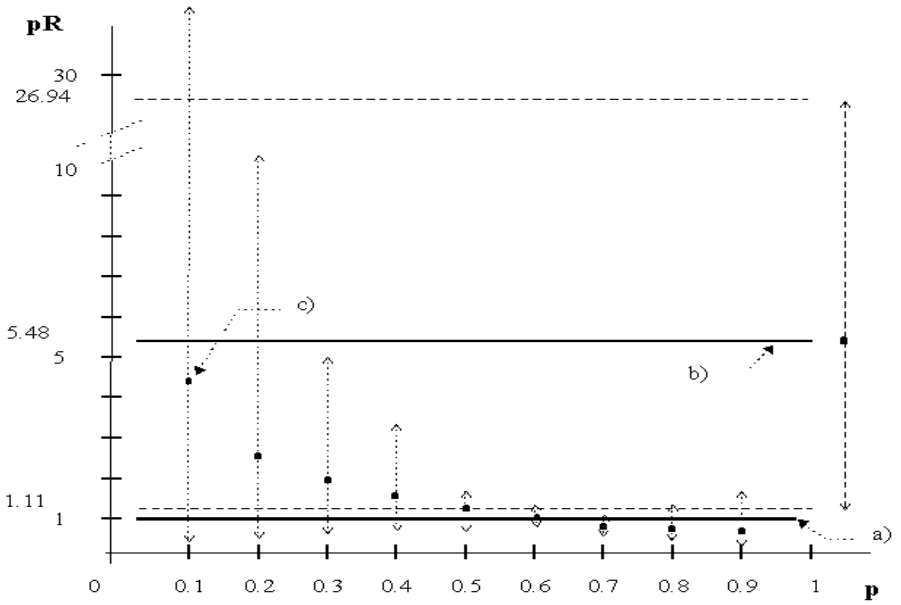

Figure 3.4: *90 % confidence intervals for the relative percentiles estimates for a positive tuberculin patient with respect to a negative one in Pulmonary TB HIV-infected patients, Barcelona (1992-1994): a) negative tuberculin group (reference group), b) cases with CD4+ % $\leq$ 14 and positive tuberculin, and c) cases with CD4+ % $>$ 14 and positive tuberculin. Relative percentiles estimates were obtained by multiple imputation using a Weibull regression model for each subsample in each imputated bootstrap replica*

3.4 Discussion

Missing data constitute a frequent and on occasions unavoidable problem in epidemiological studies, and therefore recourse must be made to sophisticated methodologies. That is the case in this study. For this reason it is worthwhile considering the following two elements: the use of the MAR hypothesis, and the precision of the corresponding estimators.

In our study, if we analyze survival from the point of view of the response or non-response of the $CD4$ variable, Figure 3.3a suggests, graphically and validated according a Cox model, that the distribution of patients for whom the $CD4$ variable is available is basically the same as that for those whom it is unavailable. In this way we would thus obtain a MCAR response pattern.

If the survival study is carried out assuming that the tuberculin test is positive, negative or even unavailable, Figure 3.3b and the corresponding Cox modeling show that there is no significant differences between the negative tuberculin group and the not available tuberculin group. However, differences do exist between these two groups and the positive tuberculin group. In this case, the PPD observed values are not a random subsample of the potential PPD values, although they are random within the categories defined by the survival times. In other words, since the missing data distribution in the PPD variable depends on the survival time, the response probability in the PPD variable is the same for all patients for whom the same survival time has been observed. In this case, the response mechanism can be considered *ignorable* (Little and Rubin, 1987). Thus, inferences based directly on the observed values of the PPD variable cannot be made since they are not representative. Nevertheless, inferences can be made from the conditional probabilities of the PPD variable, given the observed variables. In particular, methods based on maximum likelihood, which is the case of those used in this work, are valid, and the said inferences take into account the information referring to the non-observation of the PPD variable at some levels of the survival time variable.

If we consider the precision of the estimators resulting from the new methodology, two points are worthy of note; on one hand, at each step of the imputation we recover the size of the initial sample ($n = 494$), and this facilitates greater precision in the estimates. However, the added variance due to the methodology in each bootstrap

replica, and that due to the multiple imputation process itself, must be taken into account.

In recent years the relation existing between a series of demographical, clinical and immunological factors in AIDS patient prognosis has become increasingly clear (Macroft et al., 1996). Nevertheless, studies based on TB/HIV are scarce, and to the present date somewhat inconclusive, since the number of cases available for study has been small. In this study, it is important to point out the high number of patients with PTB studied and the survival time (a median of 28.8 months) greater than that recorded in other studies (median between 12 and 24 months) (Macroft et al., 1996). This difference could well be accounted for by the fact that TB was detected in the early stages of HIV infection (Stoneburrer et al., 1992), which would represent a lead time between 6 and 9 months in AIDS diagnosis (Ellner, 1990; Poznasky et al., 1995). Moreover, younger AIDS patients have a greater chance of survival (Macroft et al., 1996; Darby et al., 1996; Vella et al., 1995). It is well worth noting that more than 75% of all patients studied were IVDU, and it is well known that IVDU patients are often younger than persons belonging to other HIV transmission groups (Caylà et al., 1993b; Marco et al., 1995; Caylà et al., 1995).

Furthermore, after fittings for different variables such as HIV transmission group, presence of opportunistic diseases and therapy against HIV, TB is an accelerating factor in the progression of HIV/AIDS, and patients with TB have a shorter survival than HIV patients who do not have TB (Whalen et al., 1995). However, TB preventive therapy delays the appearance of other illnesses in HIV-infected patients, while at the same time prolonging survival (Pape et al., 1993). Consequently, it is important to note that early diagnosis and treatment of TB is not only beneficial for the community at large, preventing further cases of infection, but also improves the quality of life and extends the survival time of patients already infected. To this end, in communities where HIV/*Mycobacterium tuberculosis* coinfection is high, such as the IVDU group (de Maron and García, 1993), it is necessary to promote programs of chemoprophylaxis and chemotherapy together with the administration of methadone to ensure effective treatment and to bring down the incidence of TB and AIDS, as well as to reduce contagion period of TB patients (Curtis et al., 1994; Gourevitch et al., 1996; Raistrick et al., 1996).

Previous survival studies show that patients with a high percentage of T CD4+

have a better prognosis (Jones et al., 1993; Milles and Jones, 1993). In this study, a positive result in the tuberculin test proved to be a better marker for survival prediction than the CD4+ percentage in patients with a T CD4+ lymphocyte percentage lower than 14%, which would indicate that the activity of T CD4+ lymphocytes, evaluated indirectly by means the tuberculin test, is more important than the T CD4+ percentage. In this regard, various authors have suggested that the presence of lymphocytes with Th1 citokines pattern (IL-2 and IFN-γ, among others) of production would increase chances of survival for HIV positive patients, since these Th1 lymphocytes are the ones involved in immunity-delay phenomena such as the tuberculin test (Levy, 1993). Before HIV epidemic, 90% of PTB had a positive tuberculin test (Fernández-Nogués et al., 1983). This suggests that the high prevalence of negative tuberculin test in TB HIV-infected patients observed in our study is due to the HIV immunodepression. Then, a negative tuberculin test without information of $CD4$ in these patients is indicative of poor prognosis (see Figure 3.3b).

These results presented here concur with those of other studies which point to the possibility of carrying out an estimate of survival probability in TB patients without necessarily using T CD4+ lymphocyte counts (Richter et al., 1995; Whalen et al., 1994). As can be seen from this study, the tuberculin skin test can play an important role, particularly when lymphocyte subsets or viral load, for example, are still unavailable in countries with the highest levels of HIV/*Mycobacterium tuberculosis* coinfection; countries which, moreover, due to their delicate economic situation, are unable to draw upon the technical resources necessary for specific laboratory tests.

Subsample	Complete data analysis ($n = 157$)		Bootstrap & Imputation ($n = 494$)	
parameters*	Estimate (95% CI)	VC† %	Estimate (95% CI)	VC %
$CD4\% \leq 14$				
β_1	1.419 (−0.03 to 2.87)	52.08	1.7012 (−0.20 to 3.60)	56.90
pR	4.1329 (0.97–17.59)		5.48 (0.82–36.55)	
$CD4\% > 14$	No significant differences		No significant differences	

Table 3.6: *Comparative study between complete data analysis and the bootstrap & imputation new methodology in Pulmonary TB HIV-infected patients, Barcelona (1992-1994)*

* *New methodology is based in multiple imputation using a Weibull regression model for each subsample in each imputated bootstrap replica and relative percentiles are for a positive tuberculin patient with respect to a negative one.* † *Variation coefficient*