2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Study of the plasma levels of soluble CD14 in HIV infected patients

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Prof. PH KOLH

Nathalie MAES

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Study of the plasma levels of soluble CD14 in HIV infected patients

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Prof. PH KOLH

Nathalie MAES

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Summary

The HIV infection induces an increase of the intestinal permeability leading to the entry of bacteria. Those can be detected by an increase of bacterial lipopolysaccharide (LPS) which accelerates HIV replication by activating immune cells. The protein encoded by CD14 gene is a component of the innate immune system and acts as a co-receptor for the detection of bacterial LPS. Plasma levels of soluble CD14 (sCD14) is a marker for macrophage activity. Recent studies showed that sCD14 levels are higher in HIV infected persons and some favored the use of sCD14 level to predict disease progression and mortality in HIV infection.

In this report, we study the relationship between sCD14 and other markers and patient's characteristics in the light of causality issues, we evaluate the impact of protease inhibitor treatment on sCD14 levels, and we study the profile of patients with high sCD14 level. Therefore, we use a cross-sectional dataset of 443 chronic HIV-infected patients treated at the Department of Infectious Diseases of the University Hospital of Liege in 2011 and 2012.

Structural equation models are used in order to study the complex interrelations between sCD14 levels and other biological, clinical and treatment's information. We show the influence of sCD14 levels on the immune system through its impact on $CD4^+$, $CD8^+$ and viral load levels, we observe his significant impact on the β_2-microglobulin and C-Reactive Protein levels, and we demonstrate that the race of the patient significantly influence the sCD14 value. The impact of protease inhibitor treatment on sCD14 is not showed.

Study of the profile of chronic HIV patients with high sCD14 level (≥ 2000 ng/ml) shows that they are significantly older, have higher β_2-microglobulin, C-reactive Protein and Gamma Glutamyl Transpeptisdase levels. The proportion of Caucasian patients is significantly higher when sCD14 is high.

Thank you to the professors of the master in Biostatistics. I particularly acknowledge Pr Geert Molenberghs who generously proposed his help as master thesis' supervisor. His recommendations and encouragements where precious.

I want to acknowledge the members of the CHU hospital of Liege who gave me the opportunity to follow this biostatistics' master program. Thank you to Mr Pol Louis for his confidence and his understanding, to Pr Philippe Kolh for his support and helpful advices and who accepted to be my external thesis supervisor, to Pr Michel Moutschen who fills up me the possibility to approach the field of infectious diseases, to Adrien Devoeght for his good work in data preparation.

I'm so grateful to my husband and children who accepted the time spend on this education program. Nothing had been possible without their patience and encouragements.

Table of contents

List of abbreviations

AIDS	Acquired Immunodeficiency Syndrome
AGFI	Adjusted Goodness of Fit Index
BMI	Body Mass Index
CRP	C-Reactive Protein
GFI	Goodness of Fit Index
GGT	Gamma Glutamyl Transpeptidase
HDL	High Density Transpeptidase
HDLcholest	HDL cholesterol level
HIV	Human Immunodeficiency Virus
lB2microgl	ln(β_2-microglobulin level)
lsCD14	Ln(sCD14)
lCD4	Ln(CD4+ count)
lCD8	Ln(CD8+ count)
lViralLoad	ln(Viral load)
lCRP	ln(CRP)
lGGT	ln(GGT)
$Race_1$, $Race_2$, $Race_3$	Binary Race indicators
RMR	Root Mean square Residual
sCD14	Soluble sCD14 level
SEM	Structural Equation Modeling
SRMR	Standardized Root Mean Square Residual
TimeDiag	Time since first diagnosis
TimeTreat	Time since first treatment
Treat	Protease inhibitor treatment indicator
VitD	Vitramin D level
VitDdef	Binary Vitamin D deficiency indicator
VL	Binary viral load level indicator
ViralFail	History if viral response failure indicator
WHO	World Health Organization

1 Introduction

Since the apparition of the disease in the beginning of the 1980's, enormous progress has been done in prevention, diagnosis and treatment of Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS). Nevertheless, it remains one of the leading infection in the world and a major challenge for health care systems.

From an epidemiological point of view, crucial points are estimation of prevalence and incidence of HIV and AIDS, build of models describing the epidemic process, identification of risk factors for infection and transmission, elaboration of decision rules for treatment selection and evaluation of new therapies.

According to the World Health Organization (WHO), there were approximately 34 million people living with HIV worldwide in 2010 (Table 1). Its incidence declines and is estimated in 2010 to 2.7 million newly infected. Infected people are spread over the world but most of them (around 60%) live in sub-Saharan Africa. Globally, the estimated number of people dying from AIDS-related causes has been decreasing since 2005.

Table 1 : Key indicators for the HIV epidemic 2002-2010 (WHO, UNAIDS, UNICEF 2011)

	2002	2003	2004	2005	2006	2007	2008	2009	2010
Number of people living with HIV (in millions)	29.5 [27.7-31.7]	30.2 [28.4-32.1]	30.7 [28.8-32.5]	31.0 [29.2-32.7]	31.4 [29.6-33.0]	31.8 [29.9-33.3]	32.3 [30.4-33.8]	32.9 [31.0-34.4]	34.0 [31.6-35.2]
Number of people newly infected with HIV (in millions)	3.1 [3.0-3.3]	3.0 [2.8-3.1]	2.9 [2.7-3.0]	2.8 [2.6-3.0]	2.8 [2.6-2.9]	2.7 [2.5-2.9]	2.7 [2.5-2.9]	2.7 [2.5-2.9]	2.7 [2.4-2.9]
Number of people dying from AIDS-related causes (in millions)	2.0 [1.8-2.3]	2.1 [1.9-2.4]	2.2 [2.0-2.5]	2.2 [2.1-2.5]	2.2 [2.1-2.4]	2.1 [2.0-2.3]	2.0 [1.9-2.2]	1.9 [1.7-2.1]	1.8 [1.6-1.9]
% of pregnant women tested for HIV[a]				8%	13%	15%	21%	26%	35%
Number of facilities providing antiretroviral therapy[a]						7 700	12 400	18 600	22 400
Number of people receiving antiretroviral therapy[a]	300 000	400 000	700 000	1 330 000	2 034 000	2 970 000	4 053 000	5 255 000	6 650 000
Number of children receiving antiretroviral therapy[a]				71 500	125 700	196 700	275 400	354 600	456 000
Coverage of antiretroviral medicines for preventing mother-to-child transmission (%)[a]			9%[b]	14%[b]	23%[b]	33%[b]	43%[b]	48%[b]	48%[c]

a In low- and middle-income countries.
b The coverage data includes provision of single-dose nevirapine which is no longer recommended by WHO.
c This data does not include single-dose nevirapine regimen which is no longer recommended by WHO. It should not be compared with the previous years. When including single-dose nevirapine, the coverage in 2010 is 59%.

We observe (WHO, UNAIDS, UNICEF 2011) that those trends differ across the regions. For example HIV incidence has increased in Middle East and North Africa from 43 000 newly infected people in 2001 to 59 000 in 2010. In Easter Europe and Central Asia, after a decrease in 2000, HIV incidence has been increasing again since 2008.

In Europe in 2010 (European Centre for Disease Prevention and Control/WHO Regional Office for Europe 2011), an incidence of 27 116 newly HIV infected people in 28 countries of the European Union and European Economic Area (EE/EEA) was reported. The highest rates were reported by

Estonia, Latvia, Belgium and United Kingdom. The rate of newly reported cases is stable since 2004. In Belgium, the rate of newly diagnosed per 100 000 population was 9.4 in 2001, 9.6 in 2004, 10.3 in 2008 and 11.0 in 2010.

Despite the global decline of incidence, improvement of treatment and of access to therapy lead to an increase of the number of people living with HIV.

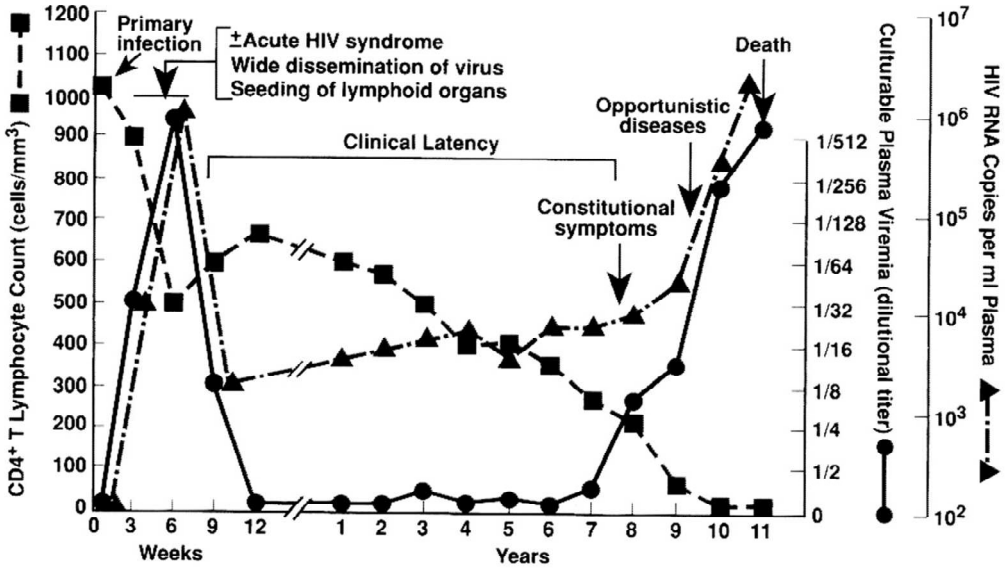
Up to now, no cure has been found but antiretroviral treatments allow controlling the virus. In 2010 around 6.6 million people living in the world with HIV had access to antiretroviral therapy. But it remains too low in some countries. Half of the populaion living with HIV are women and providing antiretroviral prophylaxis to HIV infected pregnant women has prevented a huge number of children from infection.

Human Immunodeficiency Virus (HIV) specifically attacks one group of cells of the immune system, the CD4+ T helper cells, leading to Acquired Immunodeficiency Syndrome (AIDS). This immunodeficiency leads to the development of opportunistic infections.

We observe the following typical course of HIV infection (Figure 1) :

- During the early period after primary infection, there is widespread dissemination of virus and a sharp decrease in the number of CD4+ T cells in peripheral blood.
- An immune response to HIV ensues with a decrease in viremia followed by a prolonged period of clinical latency. During this period, viral replication continues.
- The CD4+ T-cell count gradually decreases during the following years, until it reaches a critical level below a level of 200/microliter which there is a substancial risk of opportunistic diseases.

Figure 1 : Typical course of human immunodeficiency virus (HIV) infection (Fauci A.S. 1996)



The duration of clinical latency period can vary from a person to another. The mean time between primary infection and AIDS is 10 years (Fauci A.S. 1996).

The HIV infection induces an increase of the intestinal permeability leading to the entry of bacteria. Those can be detected by an increase of bacterial lipopolysaccharide (LPS) which accelerates HIV replication by activating immune cells. The protein encoded by CD14 gene is a component of the innate immune system and acts as a co-receptor for the detection of bacterial LPS. Plasma levels of soluble CD14 (sCD14) is a marker for macrophage activity (http://www.ncbi.nlm.nih.gov/gene/929 s.d.). Recent studies showed that sCD14 levels are higher in HIV infected persons and some favored the use of sCD14 level to predict disease progression and mortality in HIV infection (Sandler N.G. 2011) (Lien E. 1998).

In this report, we will study the relationship between sCD14 and other markers in the light of causality issues, we will evaluate the impact of protease inhibitor treatment on sCD14 levels, and we will study the profile of patients with high sCD14 level.

2 The crucial role of markers

In the last decades, technological progress provided new biological markers of interest. Markers are variables that indicate the extent of disease progression and the risk of AIDS or death (G. M. Brookmeyer R. 1994). Compared to the classical endpoints given by death and AIDS, markers are more frequently and earlier measurable.

Some well known immunological markers are the CD4+ T cells level, the CD8+ T cells level, the serum β_2-microglobulin and the serum neopterin. Useful viral markers are presence or absence of detectable p24 antigen and plasma viremia. We can add to those some clinical markers such as weight loss, candidiasis, persistent diarrhea, herpes zoster, fatigue and night sweats, persistent fever and oral hairy leukoplakia.

Their use made evaluation of HIV patient's evolution to go from survival (progression to AID or death) to continuous (CD4+ T cells level) and/or binary (viral load above or below a fixed level) measurements.

Their knowledge is useful at all levels of HIV and AIDS studies:

- In the ***estimation of HIV prevalence and incidence***, where biomarkers are used to indicate when a person had been infected (Brookmeyer R. 2010).

- In the ***identification of risk factors*** for infection and transmission to use as prognostic factors for predicting disease progression. It should be emphased that markers are consequences of the infection, by opposition with cofactors (or risk factors), such as age at infection, which are causal agent that affect the disease progression (G. M. Brookmeyer R. 1994).

- In the build of ***models for the epidemic process*** because they characterize the history of HIV infection.

- In the ***evaluation of therapies***, ***to design and analyze clinical trials*** of treatment. Since the mid-1990's, there was interest to use some markers in practice as surrogate endpoints for the progression to the acquired immunodeficiency syndrome (AIDS) or the death. (Burzykowski T. 2005). Surrogate endpoints are measurements that can replace other endpoints. They are useful when they can be measured easier, earlier or more frequently than initial endpoints.

- For optimal ***treatment strategy*** (Lu W. 2011) because the disease progression vary among persons, leading to uncertainty about the therapy choice and when it should begin.

The CD4+ T cells level is decreasing with the progression of the disease (Figure 1) and was one of the first used markers in HIV infection (G. M. Brookmeyer R. 1994) (Mellors J.W. 1997). We note two major problems with the use of this marker: measurement error and within-subject biological variation. This is one of the reasons why the use of other markers was considered.

Viral load is referred to as HIV RNA and it is established that HIV RNA level is strongly predictive of disease progression. For example, Mellors et al (Mellors J.W. 1997) showed that plasma viral load strongly predicts the rate of decrease in CD4+ lymphocyte count and progression to AIDS and death. A high viral load value corresponds to higher risk of disease progression. This marker is often considered as the best predictor of HIV progression.

β_2-microglobulin level is also a widely used marker. Higher values of β_2-microglobulin determine HIV progression (Gupta S.M. 2004). This marker presents the interest to be cheaper than the two previous ones.

3 Material and methods

3.1 Cross-sectional dataset

We use a cross-sectional dataset of chronic HIV-infected patients, treated at the Department of Infectious Diseases of the 'University Hospital of Liege' in 2011 and 2012. For those patients, sCD14 level, CD4+ and CD8+ T cell counts, β_2 microglobulin and viral load are measured in plasma samples. Sex, race, age and BMI at measurement are also collected. Other indicators such as Nadir CD4, Gamma Glutamyl Transpeptidase (GGT), vitamin D levels, C-Reactive Protein (CRP), D-dimer and cholesterol values and disease and treatment history of the patients are also available.

The medical history of patients with a CRP level higher than 3 mg/l was analyzed in details and patients with acute infection are excluded. This selection gives us a final dataset of 443 chronic HIV-infected patients.

The data set contains 497 observations corresponding to the 443 patients (sometimes several observations for the same patient). If we keep only the first biological set of result of each patient, we have 443 observations.

Patients include 206 women (46%) and 237 men (54%). Six (1%) are from asian race, 179 (40%) from caucasian , 6 (1%) from maghrebian and 231(52%) from black race. Race is unknown for 21 patients.

In the models, we will consider the following sex and race indicators:

$$Sex = \begin{cases} 0\ if\ men \\ 1\ if\ women \end{cases},$$

$$Race_1 = \begin{cases} 0\ if\ caucasian, asian\ or\ maghrebian\ race \\ 1\ if\ black\ race \end{cases},$$

$$Race_2 = \begin{cases} 0\ if\ caucasian, maghrebian\ or\ black\ race \\ 1\ if\ asian\ race \end{cases},$$

$$Race_3 = \begin{cases} 0\ if\ caucasian, asian\ or\ black\ race \\ 1\ if\ maghrebian\ race \end{cases}.$$

Patients are aged from 5 to 74 years with a mean age of 43 years.

Mean time since first HIV diagnosis is 105 months. Mean time since first treatment is 83 months and 67/152 patients (44%) have a history of viral response failure leading to change of their treatment. We note that 43% of the patients of the sample received protease inhibitor.

Table 2 shows the patient's characteristics and markers. Histograms of markers" distributions are given in appendix.

Levels of sCD14 range from 674 to 4891 ng/ml, with a mean of 1700 ng/ml and a median of 1609 ng/ml. The laboratory has not yet established references values for those levels but, according to physicians, it seems that these values are relatively high.

Mean value of absolute CD4+ lymphocytes count is 581/mm^3. Almost all patients have a value higher than 200/mm^3 which is the usual cut-off point for a higher risk for AIDS. Nevertheless, a patient below the value of 500/mm³ is already considered as immunodeficient. Nadir CD4+ level is in mean 280/mm³. Mean absolute CD8+ count is 907/mm^3 and is higher than the reference values of the laboratory. CD4+ and CD8+ counts are both of interest because they provide different information about the disease status and are kept separately in the models.

We observe that viral load is not measurable below 20 copies/ml. Since this value is very low, we transform the data by replacing the '<20 copies/ml' value by the 0 to 20 mid-point value of 10 copies/ml in order to have a continuous variable.

Continuous viral load value is often transformed to the binary. In this report, we use viral load level indicator VL:

$$VL = \begin{cases} 0 \; if \; Viral \; load \; < 200 \; copies/ml \\ 1 \; if \; viral \; load \; \geq 200 \; copies/ml \end{cases}.$$

The cut-off point of 200 copies/ml is chosen because some patients may have some small viral load variations between 20 and 200 witch physicians do not consider as medically significant. In our sample, 142 patients have a viral load above 200 copies/ml, 274 a viral load below 200 copies/ml and we don't have the viral load value of the 27 other patients.

Mean β_2 microglobulin level is 2.19 mg/l. Note that, in the initial sample, two patients had very high values of 22.4 and 57.80 mg/l due to renal failure. It was decided in the final sample constitution not to use the β_2 microglobulin levels of those two patients.

Vitamin D levels are low, the median value is 24 ng/ml while normal value is above 32. We observe that vitamin D level is undetectable below 8 ng/ml. Since we are here interested in severe vitamin D deficiency, we will consider the following binary vitamin D deficiency indicator

$$VitDdef = \begin{cases} 1 \; if \; Vitamin \; D \; < 8 \; ng/ml \\ 0 \; if \; Vitamin \; D \; \geq 8 \; ng/ml \end{cases}.$$

And we have in our sample 19 patients with severe vitamin D deficiency, while 226 patients don't have severe vitamin D deficiency and information is not available for the remaining 198 patients.

The mean Body Mass Index (BMI) of the patients is 25.65 kg/m² and they have high mean total cholesterol value of 1.90 g/l since the reference values are between 1.2 and 1.9 g/l. Non HDL cholesterol is in mean 1.34 g/l which is high compared with the reference of <1.30 g/l. Mean HDL cholesterol is 0.56 g/l. We use only that HDL cholesterol value in our models since it is known to be (negatively) correlated with triglycerides and does not require that the patient is fasting during the measurement.

Gamma glutamyl transpeptidase (GGT) level is a liver-health indicator. Its mean value in our sample is near 57 UI/l. This is high compared to the reference between 5 and 50 UI/l but this high mean is due to the high GGT value of one patient. Median value is 34 UI/l.

Table 2 : Patient's characteristics

	Lab. ref values	*Number*	*Min*	*Perc25*	*Med*	*Mean*	*Perc75*	*Max*
age (years)		**442**	**5**	**34**	**42**	**43**	**51**	**74**
BMI* (kg/m²)		**200**	**15.9**	**22.2**	**24.7**	**25.7**	**28.7**	**43.1**
time since first diagnosis (months)		**337**	**0**	**38**	**85**	**105**	**158**	**340**
time since first treatment (months)		**357**	**0**	**25**	**72**	**83**	**126**	**274**
sCD14 (ng/ml)		**443**	**674**	**1385**	**1609**	**1700**	**1952**	**4891**
absolute CD4 (count/mm³)	**300 – 1400**	**372**	**15**	**386**	**537**	**581**	**713**	**2322**
nadir CD4 (count/mm³)		**422**	**1**	**130**	**255**	**280**	**390**	**1311**
absolute CD8 (count/mm³)	**200 – 900**	**372**	**158**	**592**	**835**	**907**	**1144**	**4060**
β_2 microglobulin (mg/l)	**1.02 – 2.46**	**230**	**1.14**	**1.62**	**1.98**	**2.19**	**2.53**	**7.02**
viral load (copies/ml)	**20 – 10000**	**416**	**<20 for 160 patients**	**<20**	**59**	**-**	**506**	**1560000**
vitamin D (ng/ml)	**>32**	**245**	**<8 for 19 patients**	**17**	**24**	**-**	**32**	**61**
GGT* (UI/l)	**5 – 50**	**355**	**7**	**22**	**34**	**56.88**	**57**	**1376**
CRP* (mg/l)	**0.0 – 6.0**	**340**	**<0.2 for 16 patients**	**0.6**	**1.4**	**-**	**2.6**	**42.6**
D-dimer (mg/l)		**45**	**<170 for 5 patients**	**236**	**290**	**-**	**413**	**3856**
total cholesterol (g/l)	**1.20 – 1.90**	**244**	**0.90**	**1.62**	**1.87**	**1.90**	**2.15**	**3.15**
HDL*- cholesterol (g/l)		**237**	**0.21**	**0.42**	**0.53**	**0.56**	**0.66**	**1.28**
non HDL- cholesterol (g/l)	**<1.30**	**237**	**0.37**	**1.04**	**1.32**	**1.34**	**1.58**	**2.81**

* *BMI: Body Mass Index, GGT: Gamma Glutamyl Transpeptidase, CRP: C-Reactive Protein, HDL: High Densoty Lipoprotein*

The C-reactive protein (CRP) level is high is presence of inflammation and is a cardiovascular indicator. In our sample, 16 patients have an undetectable value below 0.2 mg/l. For those patients with very low CRP levels, the value is fixed to 0.1 mg/l which is the mid-point level between 0 and 0.2 mg/l. As explained in the beginning of this section, the medical history of the patients with a CRP level higher than 3 mg/ml was checked in order to exclude non-chronic patients.

D-dimer level, which is an indicator of the formation of a blood clot, is available for 45 patients of our sample. The value in lower than 170 mg/l for 5 patients. For those, this low value is replaced by 85 mg/l which is the mid-point between 0 and 170 in order to have a continuous variable.

3.2 Statistical methods

In order to study the relationship between sCD14 and other markers in the light of causality issues, we will use structural models which are multivariate regression model's tools allowing describing several interrelationships (Shkedy Z. 2011) (Schumacker R.E. 2010) (Kline R.B. 2011) (Raykov T. 2006). The method, based on the analysis of the variance-covariance matrix, consists in the build of several successive models.

We build multivariate regression models (structural models) with the sCD14, CD4, CD8, β_2-microglobulin and viral load markers as response endogenous variables. The endogenous variables can be both a response in an equation and an explanatory variable in another, allowing us to evaluate reciprocal effect between two variables. Age, sex, race, BMI, nadir CD4, time since first diagnosis, time since first treatment, type of treatment (protease inhibitor or not), viral response failure, HDL cholesterol, NadirCD4, Vitamin D deficiency and D-dimer are considered as exogeneous predictors. Since bidirectional relation between sCD14 levels and, respectively, GGT and CRP levels, is also off mean interest, those two characteristics will be considered here has endogenous.

The corresponding theoretical system of structural equations is given by

$$Y = BY + \Gamma X + \zeta$$

Where

- $Y = \begin{pmatrix} lsCD14 \\ lCD4 \\ lCD8 \\ lB2microgl \\ lViralLoad \\ lCRP \\ lGGT \end{pmatrix}$ is the 7x1 vector of response [1]

[1] Note that response variables where log-transformed in the model in order to improve normality of the error terms (see discussion section). Notation refers to: lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT)

- $X = \begin{pmatrix} 1 \\ Age \\ Sex \\ Race_1{}^{2} \\ HDL \\ BMI \\ NadirCD4 \\ TimeDiag^{3} \\ TimeTreat^{4} \\ ViralFail^{5} \\ Treat^{6} \end{pmatrix}$ is the 11x1 vector of predictors (and intercept)

- $B = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} & \beta_{17} \\ \beta_{21} & 0 & \beta_{23} & \beta_{24} & \beta_{25} & \beta_{26} & \beta_{27} \\ \beta_{31} & \beta_{32} & 0 & \beta_{34} & \beta_{35} & \beta_{36} & \beta_{37} \\ \beta_{41} & \beta_{42} & \beta_{43} & 0 & \beta_{45} & \beta_{46} & \beta_{47} \\ \beta_{51} & \beta_{52} & \beta_{53} & \beta_{54} & 0 & \beta_{56} & \beta_{57} \\ \beta_{61} & \beta_{62} & \beta_{63} & \beta_{64} & \beta_{65} & \beta_{66} & \beta_{67} \\ \beta_{71} & \beta_{72} & \beta_{73} & \beta_{74} & \beta_{75} & \beta_{76} & \beta_{77} \end{pmatrix}$ is a 7x7 matrix of coefficients

- $\Gamma = \begin{pmatrix} \alpha_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} & \gamma_{17} & \gamma_{18} & \gamma_{19} & \gamma_{1\,10} \\ \alpha_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} & \gamma_{27} & \gamma_{28} & \gamma_{29} & \gamma_{2\,10} \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} & \gamma_{37} & \gamma_{38} & \gamma_{39} & \gamma_{3\,10} \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & \gamma_{46} & \gamma_{47} & \gamma_{48} & \gamma_{49} & \gamma_{4\,10} \\ \alpha_5 & \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} & \gamma_{56} & \gamma_{57} & \gamma_{58} & \gamma_{59} & \gamma_{5\,10} \\ \alpha_6 & \gamma_{61} & \gamma_{62} & \gamma_{63} & \gamma_{64} & \gamma_{65} & \gamma_{66} & \gamma_{67} & \gamma_{68} & \gamma_{69} & \gamma_{6\,10} \\ \alpha_7 & \gamma_{71} & \gamma_{72} & \gamma_{73} & \gamma_{74} & \gamma_{75} & \gamma_{76} & \gamma_{77} & \gamma_{78} & \gamma_{79} & \gamma_{7\,10} \end{pmatrix}$

  is a 7x11 matrix of coefficients

- $\zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \end{pmatrix}$ is the 7x1 vector of measurements error

The measurement error ζ is assumed$\sim N_7(0, \Psi)$, where Ψ is the 7x7 covariance matrix which is initially assumed to be unstructured.

[2] $Race_1$ (Black or not) is used in the model because most of the patients are from black or Caucasian race
[3] TimeDiag : Time since first diagnosis
[4] Timetreat : Time since first treatment
[5] ViralFail : History of viral response failure indicator (0=no, 1=yes)
[6] Treat : Patient treated with protease inhibitor? (0=no, 1=yes)

$$\Psi = \begin{pmatrix} \psi_{11}^2 & \psi_{12} & \psi_{13} & \psi_{14} & \psi_{15} & \psi_{16} & \psi_{17} \\ \psi_{12} & \psi_{22}^2 & \psi_{23} & \psi_{24} & \psi_{25} & \psi_{26} & \psi_{27} \\ \psi_{13} & \psi_{23} & \psi_{33}^2 & \psi_{34} & \psi_{35} & \psi_{36} & \psi_{37} \\ \psi_{14} & \psi_{24} & \psi_{34} & \psi_{44}^2 & \psi_{45} & \psi_{46} & \psi_{47} \\ \psi_{15} & \psi_{25} & \psi_{35} & \psi_{45} & \psi_{55}^2 & \psi_{56} & \psi_{57} \\ \psi_{16} & \psi_{26} & \psi_{36} & \psi_{46} & \psi_{56} & \psi_{66}^2 & \psi_{67} \\ \psi_{17} & \psi_{27} & \psi_{37} & \psi_{47} & \psi_{57} & \psi_{67} & \psi_{77}^2 \end{pmatrix}$$

We differentiate two types of effects in the structural model: direct and indirect. For example Age has direct structural effect on sCD14. And since CD4+ is an explanatory variable in the equation of sCD14, Age has indirect effect on sCD14 via CD4+. Direct effect of Age on sCD14 is given by γ_{11} and indirect effect is given by $\gamma_{21}\beta_{12}$.

The covariance between Y and X can be decomposed into the sum of products of structural coefficients of all the variables with direct path to Y and the covariance of these variables with X.

In our theoretical model, we have to estimate 147 parameters:

$\beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16}, \beta_{17}, \beta_{21}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{26}, \beta_{27}, \beta_{31}, \beta_{32}, \beta_{34}, \beta_{35}, \beta_{36}, \beta_{37}, \beta_{41}, \beta_{42}, \beta_{43}, \beta_{45}$,
$\beta_{46}, \beta_{47}, \beta_{51}, \beta_{52}, \beta_{53}, \beta_{54}, \beta_{56}, \beta_{57}, \beta_{61}, \beta_{62}, \beta_{63}, \beta_{64}, \beta_{65}, \beta_{67}, \beta_{71}, \beta_{72}, \beta_{73}, \beta_{74}, \beta_{75}, \beta_{76}$,
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \gamma_{11}, \dots, \gamma_{1\,10}, \gamma_{21}, \dots, \gamma_{2\,10}, \gamma_{31}, \dots, \gamma_{3\,10}, \gamma_{41}, \dots, \gamma_{4\,10}, \gamma_{51}, \dots, \gamma_{5\,10}$
$\gamma_{61}, \dots, \gamma_{6\,10}\, \gamma_{71}, \dots, \gamma_{7\,10}$,
$\psi_{11}, \psi_{12}, \psi_{13}, \psi_{14}, \psi_{15}, \psi_{16}, \psi_{17}, \psi_{22}, \psi_{23}, \psi_{24}, \psi_{25}, \psi_{26}, \psi_{27}, \psi_{33}, \psi_{34}, \psi_{35}, \psi_{36}, \psi_{37}, \psi_{44}, \psi_{45}$,
$\psi_{46}, \psi_{47}, \psi_{55}, \psi_{56}, \psi_{57}, \psi_{66}, \psi_{67}$ and ψ_{77}.

They are estimated minimizing 'distance' between the observed (C) and the predicted (S) covariance matrix. We use the Maximum Likelihood method (with Newton-Raphson optimization technique) which minimize the function F given by

$$F = Tr(S\ C^{-1}) - n + \log(detC) - \log(detS),$$

where n is the number of observable variables, and which is chi-squared distributed under the null hypothesis that covariance matrix is S. We use the test statistic X^2=(n-1)F and reject the adequation of the predicted covariance matrix S if X^2 is higher than $\chi^2_{(n-1,\alpha)}$ (in other words, a good model have a high Chi-square p-value).

The associated degrees of freedom are given by the difference of the number of observed variance-covariances and the number of parameters, which has to be higher than one in order to solve the model. Theoretical model includes 147 parameters and we have 153 observed variance-covariance, leading to 153-147=6 degrees of freedom. But some parameters are fixed at zero in the initial model constitution and degrees of freedom are higher in the models.

In order to help us to reduce the number of parameters in the initial model, we first build linear regression models to evaluate separately the relation between each endogenous variable (lsCD14, lCD4, lCD8, lB2microgl,lViralLoad,lCRP, and lGGT) and each patient's characteristic (age, sex, race, BMI, NadirCD4, time since first diagnosis, time since first treatment, type of treatment, viral response failure, HDL cholesterol, Vitamin D deficiency, and D-dimer): $marker = \alpha + \beta\ characteristic + \varepsilon$. And we do not include in the structural models potential relations with a regression t-test p-value

higher than 0.05. Results are presented in section 4.1 and are combined with clinical knowledge of the physicians in order to build the initial structural model.

In practice, we implement the initial structural model on the data and estimate the parameters. Then, we analyze the parameter estimation and test for their significance. Significance is given by the t-value=estimation/standard error, and we consider that a |t| ratio larger than 2 represents a statically significant departure from 0 (corresponding to α=0.05). We build a new model based on those conclusions and repeat the same process until the final 'good' model.

Results are presented in section 4.2. Standardized estimates of the parameters of the final model are presented in order to taken into account the different scales of the measurements. Direct and indirect effects of exogenous variables are given.

A lot of goodness of fit indices are available in order to evaluate the model. We present Goodness of Fit Index (GFI), Adjusted GFI (AGFI), Root Mean square Residual (RMR), and Standardized RMR (SRMR) indices. If we note $W=S^{-1}$, those are given by

$$GFI = 1 - \frac{Tr\left(\left(W(S-C)\right)^2\right)}{Tr\left((WS)^2\right)}$$

$$AGFI = 1 - \frac{n(n-1)}{2df}(1 - GFI)$$

$$RMR = \sqrt{2\sum_{i=1}^{n}\sum_{j=1}^{i}\left(s_{ij} - c_{ij}\right)^2 / n(n+1)}$$

$$SRMR = \sqrt{2\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{\left(s_{ij} - c_{ij}\right)^2}{s_{ii}s_{jj}} / n(n+1)}$$

Model will be considered to be good if GFI and AGFI are close to 0.9, and RMR is as small as possible and SRMR is lower than 0.05.

Note that we are in the case of a mixture model, involving the analysis of observed variables that are categorical and continuous. Observed variance-covariance between continuous variables is calculated using Pearson correlations. We use Polychoric correlations for binary variables and Polyserial for correlation between continuous and binary variables. The obtained observed correlation matrix is converted into a variance-covariance matrix by the software and used as input to the model estimation. Polychoric and polyserial correlations assume that a latent normal continuous process underlies each observed binary variable. The value of the binary variable depends on if the continuous value is smaller or larger than a certain threshold point.

The laboratory has not yet established references values for sCD14 levels. Following their observations, physicians want to propose a cut-off point at 2000 ng/ml as reference level. Section 4.3 presents the profile of the patients with high sCD14 level, therefore the patients are divided in two subgroups with sCD14 level higher or equal to 2000 ng/ml or lower than 2000 ng/ml and we compare their characteristics in those two subgroups. For continuous variables, a student t test is used in order to test equality of the means. For binary variables, a Chi-square test is used in order to test

equality of proportions. In order to avoid problems with multiple tests, we choose the significance level is fixed at 0.05/17tests=0.003 (Bonferroni).

Analyses are done with SAS 9.2, using PROC CALIS (Covariance Analysis of Linear Structural Equations) in order to implement structural model (SAS Institute Inc. 2008). The 'PROC CALIS' SAS program is given in appendix. Correlation matrix is calculated using LISREL 8.80 Student Edition.

4 Results

4.1 Influence of patient's characteristic's: Simple regressions

If we build the simple regression models in order to link variation of marker levels in function of patients characteristics (age, sex, race, BMI, nadir CD4, time since first diagnosis, time since first treatment, type of treatment, viral response failure, HDL cholesterol, VitD deficiency, and D dimer), we observe in the following Table 3 that (with a level of significance of 0.05) age as an impact on sCD14, CD8+, β_2 microglobulin, Viral Load, CRP, and GGT levels but not on the CD4+ marker. Sex has an impact on the CD8+, β_2 microglobulin, Viral Load and GGT levels.

Concerning the race of the patient, we observe the impact of the 'Black-race' indicator on the markers, except on the Viral Load, CRP and GGT levels.

We will consider the impact of the HDL cholesterol level on the CD8+, β_2 microglobulin and Viral Load markers and of BMI on CD4+ and CRP levels.

Following those simple regression models, severe Vitamin D deficiency indicator and D dimer will not be considered as exogenous predictors.

Time since first diagnosis will be considered as predictor for CD4+, CD8+ and Viral Load levels, and time since first treatment for CD4+, β_2 microglobulin and Viral Load. The type of treatment (protease inhibitor or not) is a significant predictor for Viral Load and GGT levels. And a history of viral response failure may have an impact on CD8+ level.

Table 3 : Simple regression models $\mathbf{Y = \alpha + \beta X + \varepsilon}$ of markers in function of patient's characteristic: p-value of F test of null value of parameter

Y:	***lsCD14***	***lCD4***	***lCD8***	***lB2microgl***	***lViralLoad***	***lCRP***	***lGGT***
X:							
Age	**<0.0001**	0.3648	**0.0352**	**0.0445**	**<0.0001**	**0.0016**	**0.0005**
Sex	0.4693	0.5183	**<0.0001**	**0.0067**	**0.0346**	0.3736	**<0.0001**
Race1,	***<0.0001***	**0.0008**	**0.0032**	**0.0009**	0.3741	0.2581	0.0974
Race2,	*0.2034*	0.9562	0.0521	0.3668	0.8814	0.1040	0.1073
Race3	*0.0899*	0.6637	**0.0193**	0.2286	0.8679	0.3681	0.4855
HDL	0.2275	0.0878	**0.0008**	**0.0143**	**0.0038**	0.6670	0.2363
BMI	0.0717	**0.0059**	0.1479	0.1497	0.3067	**<0.0001**	0.3446
NadirCD4	**0.0065**	**<0.0001**	**0.0052**	0.5703	**<0.0001**	0.6794	0.1968
TimeDiag	0.0989	**0.0101**	**0.0421**	0.2539	**0.0001**	0.1868	0.6890
TimeTreat	0.4690	**0.0038**	0.4257	**0.0237**	**0.0135**	0.4039	0.4813
Viralfail	0.1133	0.8406	**0.0243**	0.1680	0.4980	0.1816	0.6001
Treat	0.7660	0.0786	0.643	0.2094	**0.0076**	0.4959	**0.0009**
VitDdef	0.0957	0.1656	0.9371	0.9939	0.1438	0.9854	0.1184
Ddimer	0.1779	0.6554	0.7365	0.7339	0.8216	0.8216	0.3389

lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes), VitDdef: Severe Vitamin D deficiency.

4.2 Relationship between sCD14 and other characteristics: Structural equations

Initial model

In order to build initial model, two considerations where take into account: physician knowledge and results of single linear regressions presented in section 4.1.

In the initial model, bidirectional interactions between sCD14, CD4+, CD8+, and Viral Load markers are studied. Following their experience, physicians recommend to study only potential causal relation of sCD14, CD4+, CD8+, and Viral Load markers on β2-microglobulin level. We are interested in the bidirectional interrelation between sCD14 and GGT levels and between sCD14 and CRP levels and we take into account that it is known that CRP and GGT can be related, that BMI influence CRP, that Viral Load is related with CRP and that high level of β2-microglobulin can lead to high CRP level.

On the light of those aspects and on the results presented in point 4.1., we will consider the initial model represented by the path diagram in Figure 2 , corresponding to the following initial matrices:

- $$B = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} & 0 & \beta_{15} & \beta_{16} & \beta_{17} \\ \beta_{21} & 0 & \beta_{23} & 0 & \beta_{25} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 & \beta_{35} & 0 & 0 \\ \beta_{41} & \beta_{42} & \beta_{43} & 0 & \beta_{45} & 0 & 0 \\ \beta_{51} & \beta_{52} & \beta_{53} & 0 & 0 & \beta_{56} & 0 \\ \beta_{61} & 0 & 0 & \beta_{64} & \beta_{65} & 0 & \beta_{67} \\ \beta_{71} & 0 & 0 & 0 & 0 & \beta_{76} & 0 \end{pmatrix}$$

- $$\Gamma = \begin{pmatrix} \alpha_1 & \gamma_{11} & 0 & \gamma_{13} & 0 & 0 & \gamma_{16} & 0 & 0 & 0 & 0 \\ \alpha_2 & 0 & 0 & \gamma_{23} & 0 & \gamma_{25} & \gamma_{26} & \gamma_{27} & \gamma_{28} & 0 & 0 \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & 0 & \gamma_{36} & \gamma_{37} & 0 & \gamma_{39} & 0 \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & 0 & 0 & 0 & \gamma_{48} & 0 & 0 \\ \alpha_5 & \gamma_{51} & \gamma_{52} & 0 & \gamma_{54} & 0 & \gamma_{56} & \gamma_{57} & \gamma_{58} & 0 & \gamma_{5\,10} \\ \alpha_6 & \gamma_{61} & \gamma_{62} & \gamma_{63} & \gamma_{64} & \gamma_{65} & 0 & \gamma_{67} & \gamma_{68} & \gamma_{69} & \gamma_{6\,10} \\ \alpha_7 & \gamma_{71} & \gamma_{72} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{7\,10} \end{pmatrix}$$

We also have to postulate a diagonal error variance-covariance matrix in order to reduce the number of parameters (discussion on that hypothesis is developed in section 5):

- $$\Psi = \begin{pmatrix} \psi_{11}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_{22}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi_{33}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_{44}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_{55}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_{66}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_{77}^2 \end{pmatrix}$$

And the initial model includes 78 parameters. Observed covariance matrix contains 18(18+1)/2=171 different values and is given in appendix (Appendix A.2). The number of degree of freedom of 171-78=93 and model is over identified.

Initial model's goodness of fit indexes are given in Table 4. We observe that the initial model fits quite well.

Table 4 : Goodness of fit indices of the initial model

Goodness of Fit Index (GFI)	0.9670
Adjusted Goodness of Fit Index (AGFI)	0.7912
Root Mean Square Residual (RMR)	1.2720
Standardized Root Mean Square Residual (SRMR)	0.0077
Chi-square	140.8221 (p-value<0.0001)

From the estimation of the parameters, we reduce step by step the structural model by removing the non-significant parameters (details given in appendix A.3 in Table 12). Final model is given in Figure 2 and parameter's estimates and standard errors are given in Table 5.

All the parameters of the final model are significantly different from zero. Note that, since variables of interest are differently scaled, we have to look at the standardized values of the parameters if we want to interpret the coefficients. They are given in Table 5.

Since markers are interrelated, the total effect of each characteristic on the others and its decomposition into direct and indirect effect provides us information of interest on the percentage of direct and indirect effects due to structural effect. Those are given in Table 13 in appendix.

We first observe that race has an impact on sCD14 level; chronic HIV patients of black race have lower sCD14 count than caucasian patients. The impact of race on sCD14 levels is in our model principally direct causal effect (88%). Race as also a (direct) effect on $CD4^+$ and $CD8^+$ counts; chronic HIV patients of black race have lower $CD4^+$ count, and higher $CD8^+$ count, than Caucasian patients.

We observe no direct causal relation between sCD14 and $CD4^+$ levels. But since $CD8^+$ count and sCD14 value are interrelated and $CD8^+$ and $CD4^+$ are highly interrelated, sCD14 and $CD4^+$ levels are indirectly linked. Patients with high sCD14 level have low $CD4^+$ level.

Concerning the Viral Load, we show an interrelation with sCD14 levels. Since causal relation coefficient of sCD14 on Viral Load is high (standardized estimate regression coefficient: 5.91) compared with causal relation coefficient of Viral Load on sCD14 (standardized estimate regression coefficient: -0.0388), we consider that the interrelation is meanly a causal relation of sCD14 on Viral Load. And patients with high sCD14 level have high Viral Load.

We show a significant impact of sCD14 level on β_2-microglobulin value. High sCD14 level leads to high β_2-microglobulin value. This causal relation is mainly direct (79%).

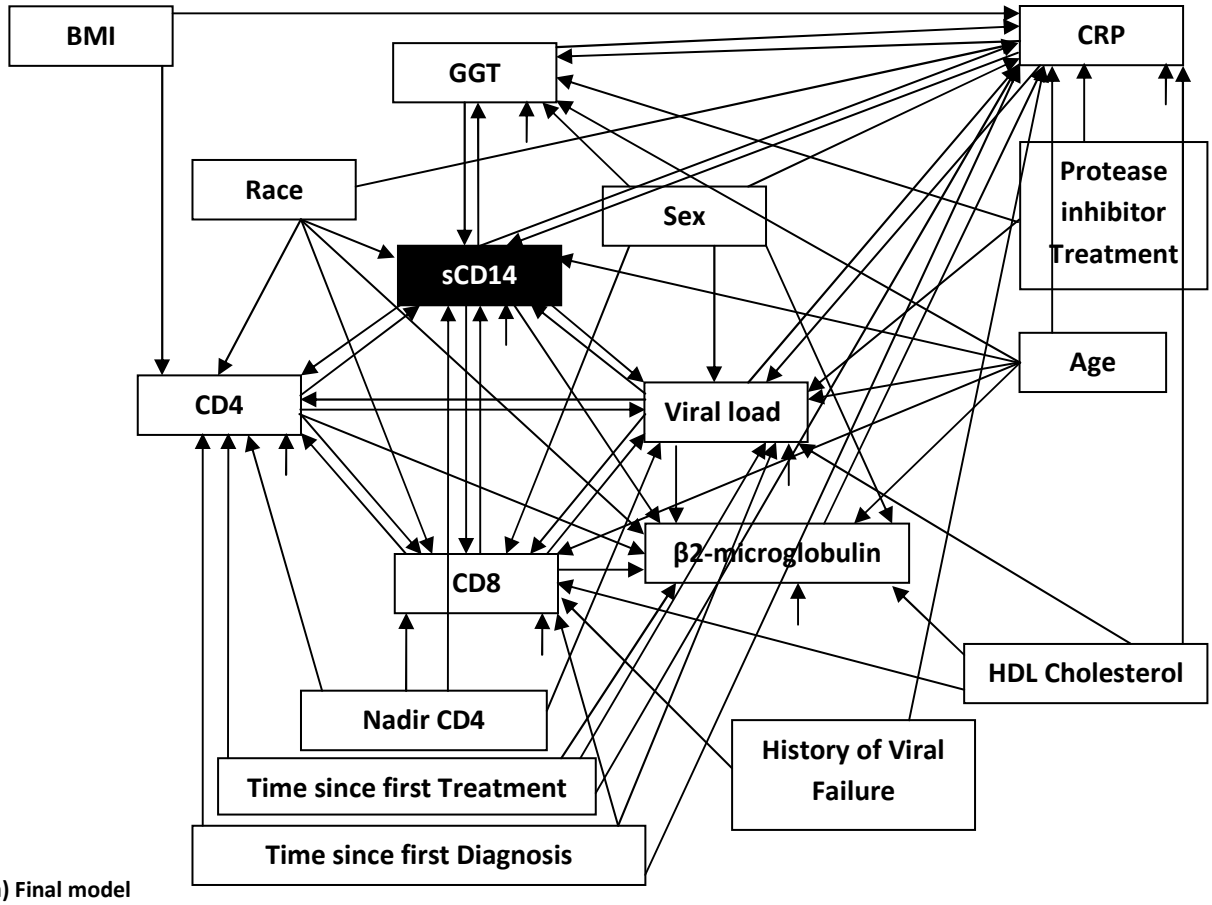
Table 5 shows that CRP level is (mainly directly) influenced by BMI and sCD14 levels. High BMI leads high CRP level and high sCD14 value leads high CRP level. We do not observe significant direct interrelation between sCD14 and GGT levels, and indirect effect in very low.

We observe the known relation between $CD4^+$ count and Viral Load. Chronic HIV patients with low $CD4^+$ count have a high Viral Load value.

The impact of protease inhibitor treatment on sCD14 levels is not showed here. We only observe a small significant indirect effect on sCD14 (Table 13). But patients who received protease inhibitor treatment have a lower Viral Load and a lower GGT value.

Figure 2 : Path diagram of the (a) initial model, (b) final model

(a) Initial model



(a) Final model

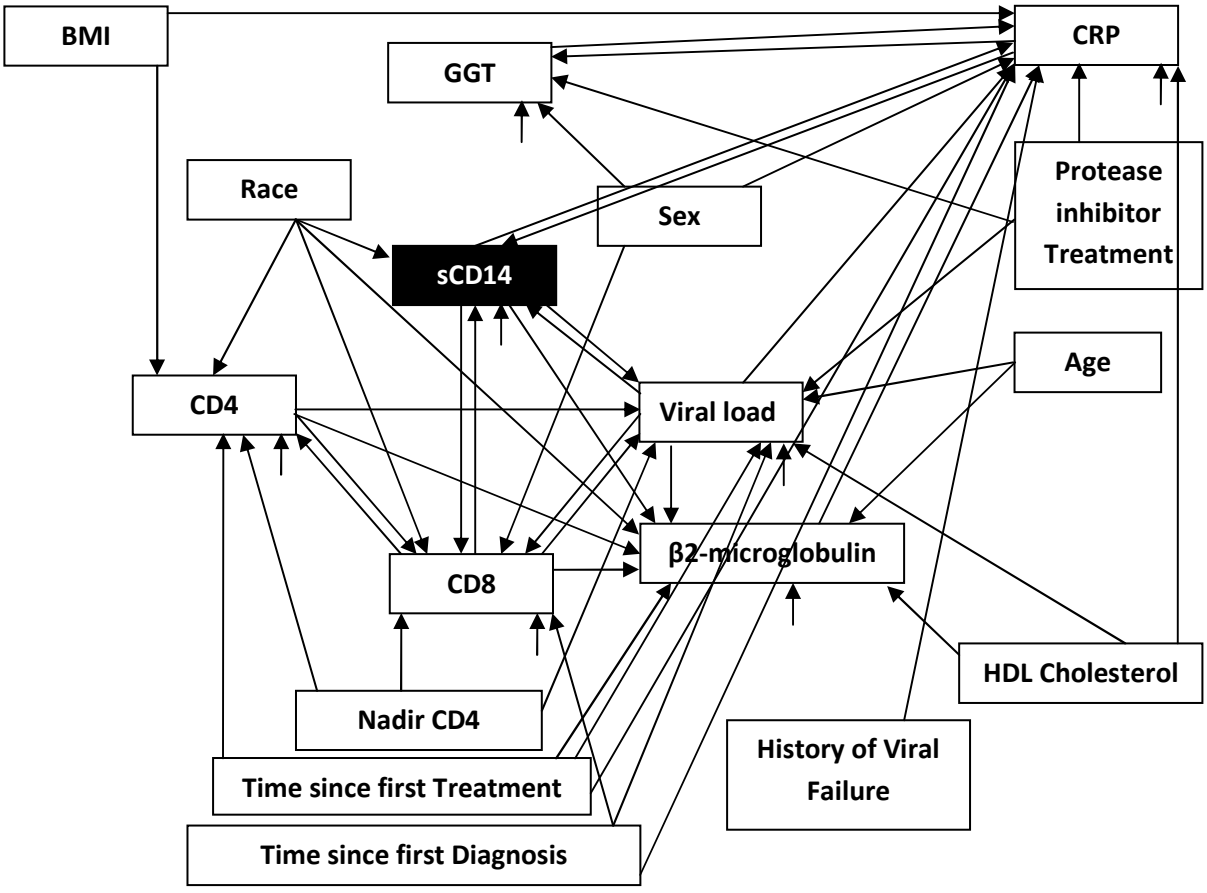


Table 5 : Final model: Parameters estimates, standard errors, t-values and standardized estimates

Parameter	*Path*	*Estimate*	*Std Error*	*tValue*	*Standardized estimate*
α_1	*lsCD14 intercept*	*10.2323*	*0.7247*	*14.1197*	***1.3817***
β_{12}	lCD4→lsCD14	-	-	-	**-**
β_{13}	lCD8→lsCD14	-0.3882	0.1083	-3.5857	**-0.3521**
β_{15}	lViralLoad→lsCD14	-0.0489	0.0125	-3.9257	**-0.0388**
β_{16}	lCRP→lsCD14	0.0558	0.0229	2.4385	**0.0087**
β_{17}	lGGT→lsCD14	-	-	-	**-**
γ_{11}	Age→lsCD14	-	-	-	**-**
γ_{13}	Race1→lsCD14	-0.0976	0.0193	-5.0441	**-0.0132**
γ_{16}	NadirCD4→lsCD14	-	-	-	**-**
α_2	*lCD4 intercept*	7.4618	0.8890	8.3933	**1.1928**
β_{21}	lsCD14→lCD4	-	-	-	**-**
β_{23}	lCD8→lCD4	-0.3841	0.1394	-2.7555	**-0.4123**
β_{25}	lViralLoad→lCD4	-	-	-	**-**
γ_{23}	Race1→lCD4	-0.1449	0.0290	-4.9950	**-0.0231**
γ_{25}	BMI→lCD4	0.0324	0.0059	5.5144	**0.1352**
γ_{26}	NadirCD4→lCD4	0.0014	0.0001	9.7101	**0.0739**
γ_{27}	TimeDiag→lCD4	-	-	-	**-**
γ_{28}	TimeTreat→lCD4	0.0016	0.0004	3.6819	**0.0262**
α_3	*lCD8 intercept*	-7.1822	1.5395	-4.6654	**-1.0695**
β_{31}	lsCD14→lCD8	1.2751	0.2004	6.3618	**1.4060**
β_{32}	lCD4→lCD8	0.6704	0.0826	8.1163	**0.6245**
β_{35}	lViralLoad→lCD8	0.0667	0.0176	3.7977	**0.0583**
γ_{31}	Age→lCD8	-	-	-	**-**
γ_{32}	Sex→lCD8	-0.2521	0.0367	-6.8776	**-0.0375**
γ_{33}	Race1→lCD8	0.2835	0.0432	6.5591	**0.0422**
γ_{34}	HDL→lCD8	-	-	-	**-**
γ_{36}	NadirCD4→lCD8	-0.0006	0.0002	-2.7907	**-0.0276**
γ_{37}	TimeDiag→lCD8	0.0009	0.0003	2.5655	**0.0172**
γ_{39}	ViralFail→lCD8	-	-	-	**-**
α_4	*lB2microgl intercept*	-1.603	0.4196	-3.8194	**-2.0131**
β_{41}	lsCD14→ *lB2microgl*	0.3058	0.0481	6.3530	**2.8438**
β_{42}	lCD4→ *lB2microgl*	-0.1134	0.0240	-4.7152	**-0.8909**
β_{43}	lCD8→ *lB2microgl*	0.1049	0.0289	3.6280	**0.8850**
β_{45}	lViralLoad→ *lB2microgl*	0.0285	0.0046	6.2595	**0.2103**
γ_{41}	Age→ *lB2microgl*	0.0030	0.0012	2.4340	**0.1654**
γ_{42}	Sex→ *lB2microgl*	-	-	-	**-**
γ_{43}	Race1→ *lB2microgl*	-0.0621	0.0147	-4.2131	**-0.0779**
γ_{44}	HDL→ *lB2microgl*	-0.2020	0.0749	-2.6952	**-0.1492**
γ_{48}	Timetreat→ *lB2microgl*	-0.0011	0.0002	-5.1687	**-0.1401**
α_5	*lViralLoad intercept*	-22.4552	7.5550	-2.9722	**-3.8239**
β_{51}	lsCD14→ *lViralLoad*	4.6903	1.0064	4.6605	**5.9147**
β_{52}	lCD4→ *lViralLoad*	-2.3950	0.3540	-6.7651	**-2.5515**
β_{53}	lCD8→ *lViralLoad*	1.5482	0.6003	2.5789	**1.7705**
β_{56}	lCRP→ *lViralLoad*	-	-	-	**-**
γ_{51}	Age→ *lViralLoad*	-0.0549	0.0133	-4.1293	**-0.4127**
γ_{52}	Sex→ *lViralLoad*	-	-	-	**-**
γ_{54}	HDL→ *lViralLoad*	-3.5038	0.8036	-4.3599	**-0.3509**
γ_{56}	NadirCD4→ *lViralLoad*	0.0066	0.0008	7.8636	**0.3802**
γ_{57}	TimeDiag→ *lViralLoad*	-0.0167	0.0332	-5.0145	**-0.3744**
γ_{58}	TimeTreat→ *lViralLoad*	0.0178	0.0039	4.5516	**0.3211**
$\gamma_{5\,10}$	Treat→ *lViralLoad*	-2.2993	0.1391	-2.1519	**-0.0509**

Table 5 (cont)

Parameter	*Path*	*Estimate*	*Std Error*	*tValue*	*Standardized estimate*
α_6	*lCRP intercept*	-7.4894	1.8754	-3.9936	**-6.5268**
β_{61}	lsCD14→ *lCRP*	0.8751	0.2518	3.4757	**5.6472**
β_{64}	lB2microgl→ *lCRP*	0.4635	0.1952	2.3750	**0.3216**
β_{65}	lViralLoad→ *lCRP*	-0.1084	0.0225	-4.8149	**-0.5546**
β_{67}	lGGT→ *lCRP*	-0.6399	0.1401	-4.5687	**-2.0633**
γ_{61}	Age→ *lCRP*	-	-	-	**-**
γ_{62}	Sex→ *lCRP*	-0.4806	0.0861	-5.5816	**-0.4184**
γ_{63}	Race1→ *lCRP*	-	-	-	**-**
γ_{64}	HDL→ *lCRP*	1.1220	0.3937	2.8500	**0.5751**
γ_{65}	BMI→ *lCRP*	0.1331	0.0140	9.4883	**3.0240**
γ_{67}	TimeDiag→ *lCRP*	-0.0063	0.0013	-4.8261	**-0.7198**
γ_{68}	TimeTreat→ *lCRP*	0.0043	0.0015	2.7927	**0.3975**
γ_{69}	ViralFail→ *lCRP*	0.2849	0.0640	4.4534	**0.2480**
$\gamma_{6\,10}$	Treat→ *lCRP*	-0.2063	0.0651	-3.1671	**-0.1796**
α_7	*lGGT intercept*	3.5147	0.0416	84.4648	**0.9499**
β_{71}	lsCD14→ *lGGT*	-	-	-	**-**
β_{76}	lCRP→ *lGGT*	0.4232	0.0658	6.4329	**0.1312**
γ_{71}	Age→ *lGGT*	-	-	-	**-**
γ_{72}	Sex→ *lGGT*	-0.1809	0.0395	-4.5753	**-0.0488**
$\gamma_{7\,10}$	Treat→ *lGGT*	-0.1262	0.0395	-3.1971	**-0.0341**
Ψ_{11}^2	Var lsCD14	0.1369	0.0273	5.01	**-**
Ψ_{22}^2	Var lCD4	0.2951	0.0371	7.96	**-**
Ψ_{33}^2	Var lCD8	0.2445	0.0296	8.26	**-**
Ψ_{44}^2	Var lB2microgl	0.0662	0.0045	14.86	**-**
Ψ_{55}^2	Var lViralLoad	7.5392	0.8262	9.13	**-**
Ψ_{66}^2	Var lCRP	1.1722	0.1368	8.57	**-**
Ψ_{77}^2	Var lGGT	0.6760	0.0641	10.55	**-**

Race1:Black(1) or not(0), lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

Goodness of fit of this model is given in Table 6.

Table 6 : Goodness of fit indices of the final model

Goodness of Fit Index (GFI)	0.9614
Adjusted Goodness of Fit Index (AGFI)	0.8499
Root Mean Square Residual (RMR)	1.9250
Standardized Root Mean Square Residual (SRMR)	0.0075
Chi-square	166.2871 (p-value<0.0001)

4.3 Patients with high sCD14 levels

In our sample, 93 patients have a sCD14 level higher or equal to 2000 ng/ml and 350 have a sCD14 level lower than 2000 ng/ml.

Table 7 gives the profile of the patients in those two subgroups.

Patients with high (≥ 2000 ng/ml) sCD14 values are older ($p=0.0007<0.003$[7]), have higher β_2-microglobulin level ($p<0.0001$), higher GGT level ($p=0.0004<0.003$) and higher CRP level ($p=0.0032\approx0.003$). The proportion of Caucasian patients is higher when sCD14 is high. The difference is statistically significant ($p=0.0009<0.003$).

Patients with high sCD14 values have higher absolute CD8 counts and higher viral load marker but difference is not statistically significant. The proportion of patients with high viral load (>200 copies/ml) is the same in the two subgroups. The proportion of patients with severe vitamin D deficiency is higher in the subgroup with high sCD14 level but difference is not statistically significant.

Mean D-dimer value is higher in the subgroup with high sCD14 level, the difference is not statistically significant but borderline and since the number of patients with that measure is low, it is possible that the significance will arise with a larger data set.

Absolute CD4 counts, BMI and the cholesterol measures do not differ significantly between the two subgroups.

[7] Significance level fixed at 0.05/17 tests=0.003 (Bonferroni)

Table 7 : Patient's characteristics in sCD14 level's subgroups

	sCD14<2000 ng/ml	***sCD14≥2000 ng/ml***	***Stat. test***
Number of patients	350	93	
Mean age (years)	42 (SD:12, N:349)	46 (SD:12,N:93)	t-test:-3.41, df:440 p-value: 0.0007
Sex (M:W)	1.2:1	1.1:1	Chi-square:0.1683 p-value:0.6816
Race			
% Caucasian	40%	58%	Chi-square:10.931 p-value:0.0009
% Black	58%	42%	
% Asian or Maghrebian	2%	0%	
Mean BMI*(kg/m²)	*25.9 (SD:5.0, N:157)*	24.8 (SD:3.7, N:43)	t-test: 1.37, df:198 p-value: 0.1726
Mean absolute CD4 (count/mm³)	579 (SD:282, N:299)	591 (SD:308, N:73)	t-test:-0.34,df:370 p-value:0.7347
Mean nadir CD4 (count/mm³)	285 (SD:188, N:337)	260 (SD:219, N:85)	t-test: 1.02,df:420 p-value: 0.3067
Mean absolute CD8 (count/mm³)	878 (SD:423, N:299)	1024 (SD: 553, N:73)	t-test:-2.47,df:370 p-value: 0.0138
Mean β_2 microglobulin (mg/l)	2.04 (SD:0.60, N:182)	2.77 (SD:1.35, N:48)	t-test:-5.53,df:228 p-value<0.0001
Mean viral load (copies/ml)	18109 (SD:97097, N:331)	44865 (SD:164507, N:85)	t-test:-1.93,df:414 p-value: 0.0544
% high viral load level*	34%	34%	Chi-square:0.00 p-value:0.997
% severe vitamin D deficiency*	6%	14%	Chi-square:3.2093 p-value:0.0732
Mean GGT* (UI/l)	48 (SD:55, N:284)	93 (SD:184, N:71)	t-test:-3.56,df:353 p-value: 0.0004
Mean CRP* (mg/l)	2.0 (SD:2.6, N:275)	3.5 (SD:6.3, N:65)	t-test:-2.97,df:338 p-value: 0.0032
Mean D-dimer	317 (SD:180, N:36)	771 (SD:1166, N:9)	t-test:-2.30,df:43 p-value: 0.0263
Mean total cholesterol (g/l)	1.87 (SD:0.41, N:190)	1.98 (SD:0.40, N:54)	t-test:-1.65,df:242 p-value: 0.1003
Mean HDL*-cholesterol (g/l)	0.55 (SD:0.18, N:186)	0.58 (SD:0.18, N:51)	t-test:-1.01,df:235 p-value: 0.3119
Non HDL-cholesterol (g/l)	1.32 (SD:0.43, N:186)	1.42 (SD:0.41, N:51)	t-test:-1.46,df:235 p-value: 0.1454

** BMI: Body Mass Index, GGT: Gamma Glutamyl Transpeptidase, CRP: C-Reactive Protein, HDL: High Density Lipoprotein*

*** High viral load level if higher than 200 copies/ml, severe vitamin D deficiency if vitamin D level lower than 8 ng/ml.*

5 Discussion

Epidemiologists know well that correlation rarely mean causation and it would be very desirable if there is a methodology which can discover causes and effects amongst variables or, at least, confirm or proposed causal relationship (Tu Y.K. 2009).

This is the field of Structural equation modeling (SEM) which is an important statistical tool for evaluating complex relations (Amorim L.D.A.F. 2010).

In particular, SEM is useful to study the relation between observed measurements and unobserved latent variables. Therefore, they are often used in social and human sciences. Structural equation models combine two types of models: a multivariate regression model called structural model and a confirmatory factor analysis. Its main purpose is the improvement of the model's structure and parameters estimators are also produced (Shkedy Z. 2011). Is this report, all variables are observed and we only use structural model (also called path analysis). Structural models present the interesting characteristic that a variable can be both a response in an equation and an explanatory variable in another, allowing us to evaluate reciprocal effect between two variables.

In the past, structural equation modeling was not so frequently used in epidemiology as in the social sciences. Under-utilization of SEM in epidemiology was probably due to restriction in the assumptions of variables (continuous outcome) and difficulties in testing non-linear relationship (Tu Y.K. 2009). But during the past 20 years, we have seen considerable research on the behavior of methods of estimation under various conditions. The most crucial conditions are characterized by the lack of multivariate normality (Schumacker R.E. 2010). In addition, software programs has been developed and improved, providing tools to larger use of SEM methodology.

In recent years, SEM methodology has been larger used in health research. A literature search in PubMed published by Amorim in 2010 (Amorim L.D.A.F. 2010) in six leading periodicals in the field of Epidemiology showed that 24 articles used SEM from 2001 to 2008, and that 62.5% of these had been published since 2006.

The ability of structural equations methodology to deal with complex models is its strength and its weakness. Dealing with complex models is complex! Sources of errors are multiple: initial model specification, normality and linearity assumptions, treatment of dichotomous variables, impact of outliers, treatment of missing values, … can lead to criticisms. Those points are discussed here.

Initial model specification:

Initial model choice is of first importance in structural equation modeling. The exclusion or inclusion of unimportant variables will produce misspecified models and may lead to biased parameter estimates (specification error) (Schumacker R.E. 2010). We used medical knowledge an exploratory analysis conducted in point 4.1 to guide us in the initial model specification.

A kind of 'sensibility analysis' has been performed with the introduction of small changes in the initial model and we observe similar results for the variables of interest.

Observed correlations:

Since we have a mixture model, involving the analysis of observed variables that are dichotomous and continuous, we used Pearson, Tetrachoric or Polyserial correlation coefficients, in function of the two implied variable's type. Table 9 in appendix gives the observed correlations with the method used.

Polychoric and polyserial correlations assume that a latent normal continuous process underlies each observed binary variable. The value of the binary variable depends on if the latent continuous value is smaller or larger than a certain threshold point. Goodness of fit of polychoric models can be improved by comparing observed crossclassification frequencies to model-predicted frequencies. In our case only 5.2% of the tests reject goodness of fit of the latent normal continuous process approximation (α=0.005).

Missing values:

We have a lot of missing values in our dataset. Dataset constitution allows us to postulate that data are missing unrelated statistically to their values and the values of other variables (Missing Completely At Random).

Variances and Covariances were computed from all observations that have values present for the pair of variable involved (pairwise deletion). We use therefore different sample size in the calculations of the correlation elements. Effective sample sizes (given in Table 10 of the appendix) is quite large for most of the variables combinations, except for BMI's variable combinations which go to only 91 patients in the evaluation of correlation between BMI and HDL cholesterol measures.

The deletion of the patients with missing values (listwise deletion) do not have that problem of different sample size but is not applicable in our data set which would be too small. One other alternative is the use of imputation methods (Schumacker R.E. 2010) (Molenberghs G. 2011) in which we replace missing values with an estimate. But since 'data creation' can always be source of errors and since our pairwise sample sizes are acceptable, we do not use imputation methods in this report.

Note that D-dimer level was not included in the initial model. But due the small number of patients with D-dimer level measurement, we are maybe not able to detect statistically causal relations with the other measures.

Estimation and validation of the final model:

In the build of the final model, we successively put at zero the parameters when |t| value is smaller than 2, which correspond to a statistically significant departure from 0 (α=0.05). This step is the succession of small steps, removing parameters one by one because, in this complex interrelation model, removing one parameter can affect the other relations. Different criteria for the order of the remove of the parameters where tested and they all lead to similar final model.

There is not sufficient procedure for model's validation. We use the combination of different methods:

- Goodness of fit indices give us indications:
  Goodness of Fit Index (GFI) measures the amount of observed variance-covariance that is predicted by the estimated matrix. A model with a GFI, or a AGFI fit, higher than 90 can be considered as acceptable. For our final model we have a GFI of 0.9614 and a AGFI of 0.8499.

The Chi² fit-criterion is not in favor of the model (p-value<0.0001) but, since it depends on the sample size, large datasets tend to indicate even small deviations from a good fit.

- Another method is to examine the residuals which should be small and should not be larger for one variable than another. Large values overall indicate general model misspecification and large values for a single variable indicate misspecification for that variable only. Large standardized residual (>2.00) indicate that a particular covariance structure is not well explained by the model. Note that because of the differential scaling of the variables it is usually more useful to examine the standardized residuals.

  In our case, residuals are overall small. The 10 largest asymptotically standardized residuals (Table 9) inform us that the highest standardized residual is 1.37 and that their do not correspond to one variable in particular.

Table 8: Rank order of the 10 largest asymptotically standardized residuals

		Std Residual
ViralFail	lB2microgl	1.3701
ViralFail	lCRP	0.8049
lCRP	lB2microgl	-0.7415
lCRP	Race1	-0.7024
Treat	lB2microgl	0.6701
lViralLoad	Race1	-0.5222
Treat	lViralLoad	-0.4637
lGGT	Race1	0.4547
lCRP	lViralLoad	0.4546
Timediag	lCRP	-0.3082

Race1:Black(1) or not(0), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), TimeDiag: Time since first diagnosis, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

All those observations lead us to conclude that our model is valid in terms of goodness of fit.

Nevertheless, it could be of interest to conduct validation of the model by replication using another similar dataset and to use bootstrapping to determine the bias in the parameter estimates.

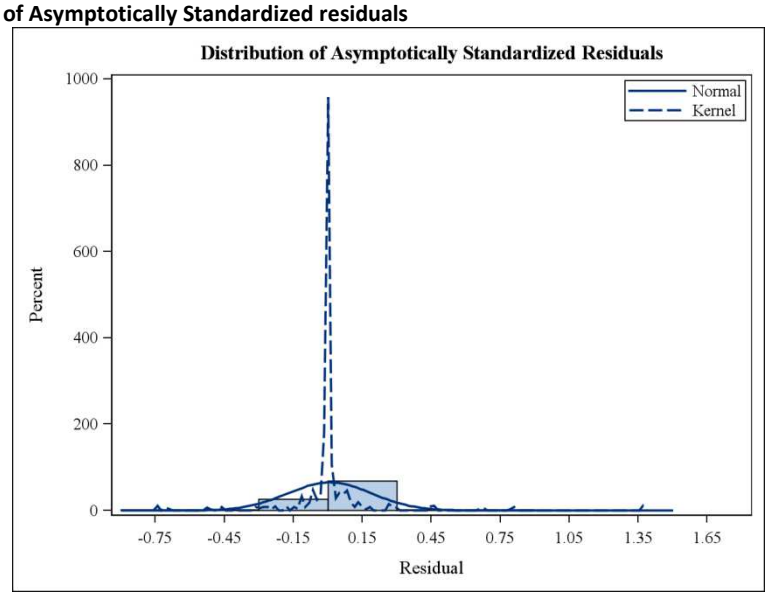
Normality:

When the normality of the error terms assumption is violated, parameter estimates and standard error are suspect. When the data are generated from non normally distributed population and/or represent discrete variables, the normal theory estimators of standard errors and the model-fit indices could be suspect (Schumacker R.E. 2010) (Kline R.B. 2011). However, recent simulations research by Lei and Lomax in 2005 indicated that the maximum likelihood estimators are quite comparable in the case of small to moderate non-normality for interval data. Following Kline (Kline R.B. 2011), severe non-normality leads to relatively accurate parameter estimates in large samples but too low estimated standard errors and the null hypothesis that the corresponding parameter is zero is rejected more often than is correct.

One option to avoid this problem is to normalize the continuous variables with transformations. That is the reason why log-transformed variables where used. Nevertheless, the Viral Load's distribution stays not normally distributed after this transformation.

If we look at the residuals, the following histogram of the asymptotically standardized residuals shows a quite symmetric distribution.

Figure 3 : Distribution of Asymptotically Standardized residuals



Note that we use a diagonal variance-covariance matrix of the error term in order to reduce the number of parameters of the model. This assumption is very restrictive but the use of several more simple initial models with unstructured variance-covariance lead all to the conclusion that covariance parameters do not differ significantly from zero.

Outliers:

Correlation coefficients are affected by outliers (Schumacker R.E. 2010). Since the GGT and CRP measures contain outlier observations, we have tested the sensibility of our model when those outliers are removed from the dataset. We have observed that, in our case, those outliers do not affect the final model.

Linearity:

Structural Equation Models used in this report assume that the variables are linearly related to one another (Schumacker R.E. 2010). The presence of curvilinear data reduces the magnitude of the Pearson correlation coefficient, even resulting in the presence of zero correlation.

Longitudinal aspect:

Since second measurements of the sCD14 levels are available for some patients, we could take into account the longitudinal aspect of the data. But since the two measurements are very close to each other and are available for few patients, those measurements do not provide information of interest on the disease process. That is the reason why we choose to work only with the first biological set of result of each patient.

6 Conclusion

In this report, we use structural equations models in order to study the complex interrelations between sCD14 and other markers and patient's characteristics in Chronic HIV patients treated at the Department of Infectious Diseases of the 'University Hospital of Liege' in 2011 and 2012.

Immune activation is a critical component of HIV diseases pathogenesis. We analyze the influence of sCD14 levels on the immune system through its impact on CD4, CD8 and viral load levels. Results show that this impact is significant and confirm the published observations (Lien E. 1998).

Following our final model, a significant direct causal relation of sCD14 levels on Viral Load is present. Chronic HIV patients with high sCD14 level have high Viral Load. And we observe indirect causal relation between sCD14 and $CD4^+$ levels. Patients with high sCD14 level have low $CD4^+$ level.

Model shows a significant impact of sCD14 level on β_2-microglobulin value, high sCD14 level leading to high β_2-microglobulin value, and on C-Reactive Protein (CRP) level, high sCD14 value leading to high CRP level. It also confirms the significant impact of Body Mass Index (BMI) on CRP level.

Armah et al (Armah K.A. 2012) showed that ongoing HIV replication and immune depletion significantly contribute to increased prevalence of elevated biomarkers of inflammation, altered coagulation (D-dimer), and monocyte activation (sCD14). Our results agree with those observations, except for the D-dimer for which we have too few measurements.

Influence of patient's characteristics is explored and we observe that the race has a significant direct impact on sCD14 levels. Chronic HIV patients of black race have lower sCD14 count than Caucasian patients.

The impact of protease inhibitor treatment on sCD14 levels is not showed here. But patients who received protease inhibitor treatment have significantly lower viral load and lower Gamma Glutamyl Transpeptidase (GGT).

Studying the profile of chronic HIV patients with high sCD14 level (≥ 2000 ng/ml) shows that they are older ($p=0.0007$), have higher β_2-microglobulin level ($p<0.0001$), higher GGT level ($p=0.0004$) and higher CRP level ($p=0.0032$). The proportion of Caucasian patients is higher when sCD14 is high. The difference is statistically significant ($p=0.0009$).

We finally conclude that we confirm in this report that soluble CD14 (sCD14) can be a marker of interest in the study of HIV process. Most of the patients studied in this report are still followed in the Department of Infectious Diseases and physicians are collecting longitudinal observations. Those longitudinal data will give the opportunity to go further in the analysis.

References

Amorim L.D.A.F., Fiaccone R.L., Santos C.A.S.T., dos Santos T.N., de Moraes L.T.L.P., Oliveira N.F., Barbosa S.O., dos Santos D.N., dos Santos L.M., Matos Sh.M.A., Barreto M.L. «Structural equation modeling in epidemiology.» *Métodos*, 2010: 26(12):2251-2262.

Armah K.A., McGinnis K., Baker J., Gilbert C., Butt A.A., Bryant K.J., Goetz M., Tracy R., Oursler K.A.K., Rimland D., Crothers K., Rodriguez-Barradas M., Crystal S., Gordon A., Kraemer K., Brown Sh., Gerschenson M., Leaf D.A., Deeks St.G., Rinaldo Ch., K. «HIV Status, Burden of Comorbid Disease and Biomarkers of Inflammation, Altered Coagulation and Monocyte Activation.» *Clinical Infectious Diseases*, 2012: 55(1):126-36.

Brookmeyer R. «Measuring the HIV/AIDS Epidemic: Approaches and Challenges.» *Epidemiologic Reviews*, 2010: 32:26-37.

Brookmeyer R., Gail M.H. *AIDS Epidemiology: A Quantitative Approach.* New York: Oxford University Press, 1994.

Burzykowski T., Molenberghs G., Buyse M. *The Evaluation of Surrogate Endpoints.* New York: Springer, Statistics for Biology and Health, 2005.

European Centre for Disease Prevention and Control/WHO Regional Office for Europe. *HIV/AIDS surveillance in Europe 2010.* Stockholm: European Centre for Disease Prevention and Control, 2011.

Fauci A.S., Pantaleo G., Stanley Sh., and Weissman Dr. «Immunopathogenic Mechanisms of HIV Infection.» *Annals of Internal Medicine*, 1996: 124,7:654-663.

Gupta S.M., Ray K., Bala M. «Evaluation of beta2 microglobulin level as a marker to setermine HIV/AID progression.» *J Commun Dis.*, 2004: 36(3), 166-70.

http://www.ncbi.nlm.nih.gov/gene/929 . (accès le 12 2011).

Kline R.B. *Principles and Practice of Structural Equation Modeling.* New York: The Guilford Press, 2011.

Lien E., Aukrust P., Sundan A., Muller F., Frøland S.S., and Espevik T. «Elevated Levels of Serum-Soluble CD14 in Human Immunodeficiency Virus Type 1 (HIV-1) Infection: Correlation to Disease Progression and Clinical Events.» *Blood*, 1998: 92:2084-2092.

Lu W., Zhang H.H., and Zeng D. «Variable selection dor optimal treatment decision.» *Statl Methods Med Res*, 2011: 1-12.

Mellors J.W., Munoz A., Giorgi J.V., Margolick J.B., Tassoni Ch.J., Gupta Ph., Kingsley L.A., Todd J.A., Saah A.J., Detels R., Phair J.P., and Rinaldo Ch.R. «Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection.» *Annals of internal medicine*, 1997: 126,12:946-954.

Molenberg G., Faes Ch. *Introduction to Multivariate Data Analysis.* Universiteit hasselt, 2011.

Molenberghs G., Verbeke G. *Introduction to Longitudinal Analysis, course notes.* Hasselt: University Hasselt, 2011.

Raykov T., Marcoulides G.A. *A First Course in Structural Equation Modeling.* London: Psychology Press, Taylor & Francis Group, 2006.

Sandler N.G., Wand H., Roque A., Law M., Nason M.C., Nixon D.E., Pedersen C., Ruxrungtham K., Lewin Sh.R., Emery S., Neaton J.D., Brenchley J.M., Deeks St.G., Sereti I., and Douek D.C. «Plasma Levels of Soluble CD14 Independently Predict Mortality in HIV Infection.» *The Journal of Infectious Diseases*, 2011: 203:780-790.

SAS Institute Inc. *SAS/STAT® 9.2 User's guide.* Cary, NC: SAS Institute Inc., 2008.

Schumacker R.E., Lomax R.G. *A Beginner's Guide to Structural Equation Modeling.* New York: Routledge, Taylor & Francis Group, 2010.

Shkedy Z. *Structural Equation Modeling: An Introduction: Path analysis and Confirmatory Factor Analysis.* Limburg Universitair Centrum, 2011.

Tu Y.K. «Commentary: ls structural equation modelling a step forward for epidemiologists?» *International Journal of Epidemiology*, 2009: 38:549-551.

WHO, UNAIDS, UNICEF. *GLOBAL HIV/AIDS RESPONSE - Epidemic update and health sector progress towards Universal Access.* Progress Report 2011, Geneva: WHO Library Cataloguing-in-Publication Data, 2011.

Appendix

A1. Histogram of the distributions of the markers en other endogeneous variables

Figure 4 : Histograms of sCD14 and ln(sCD4) level distributions

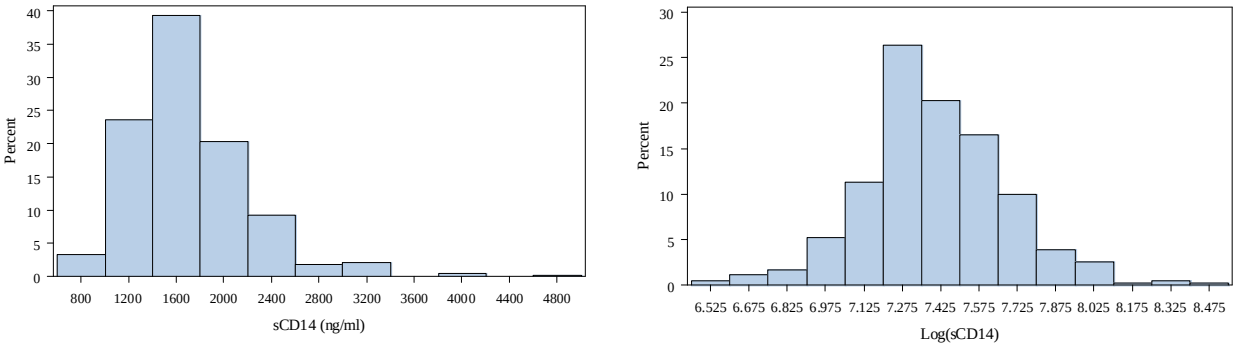


Figure 5 : Histograms of CD4+ level and ln(CD4) distributions

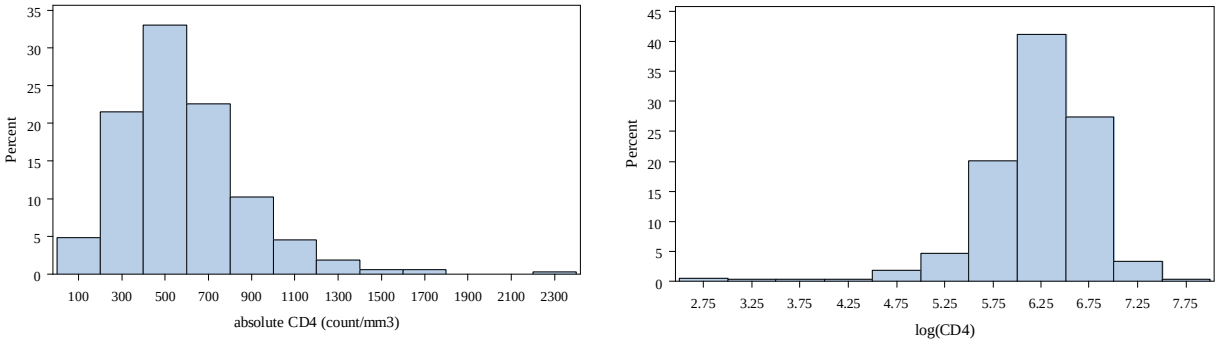


Figure 6 : Histograms of CD8+ level and ln(CD8) distributions

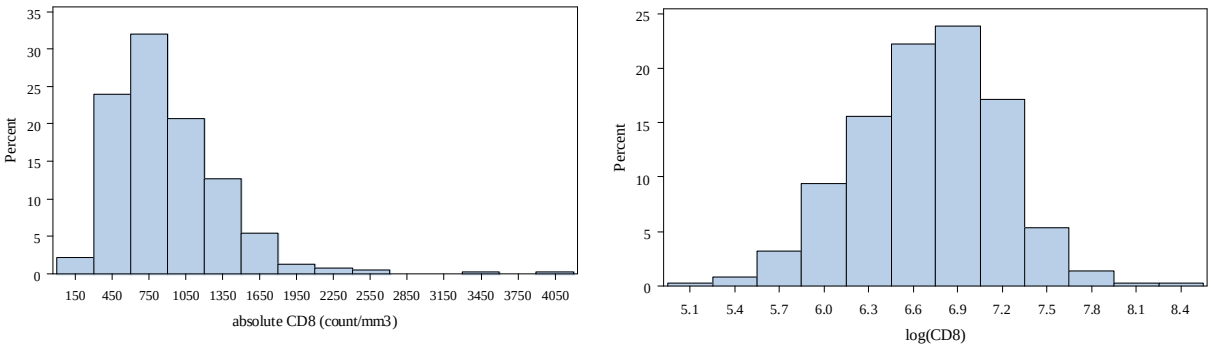


Figure 7 : Histograms of β_2 microglobulin levels and ln(β_2 microglobulin) distributions

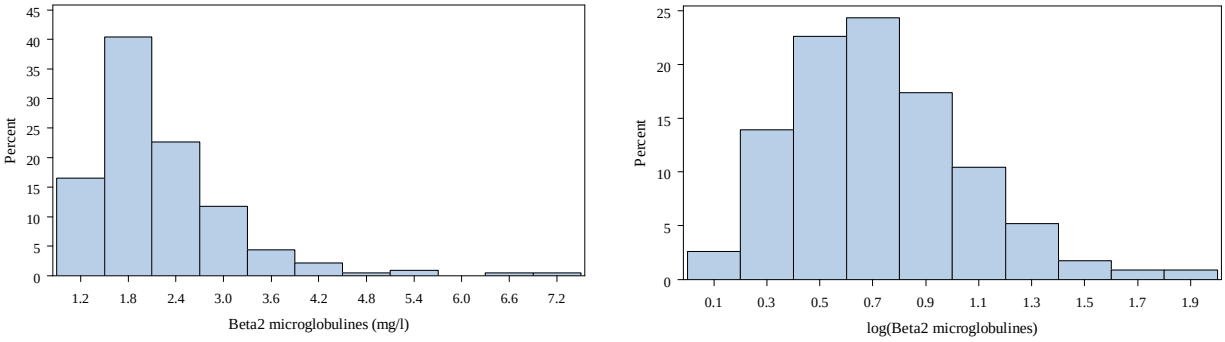


Figure 8 : Histograms of viral load and ln(viral Load) distributions

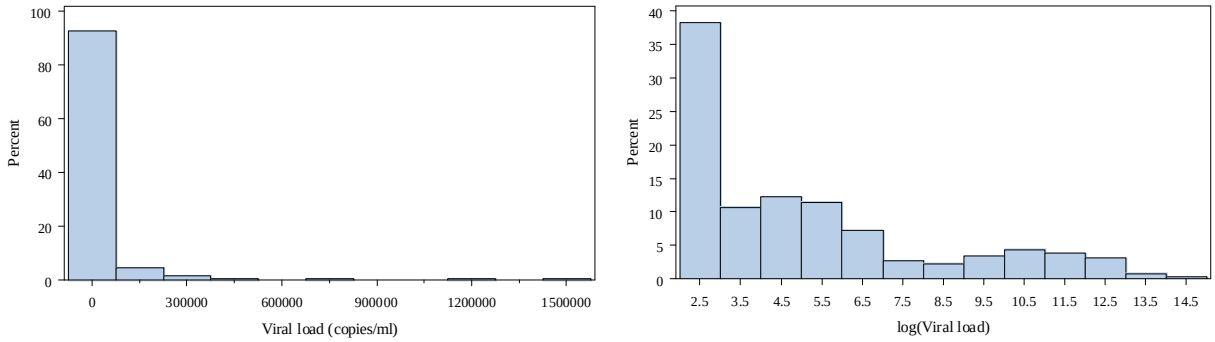


Figure 9 : Histograms of CRP and ln(CRP) distributions

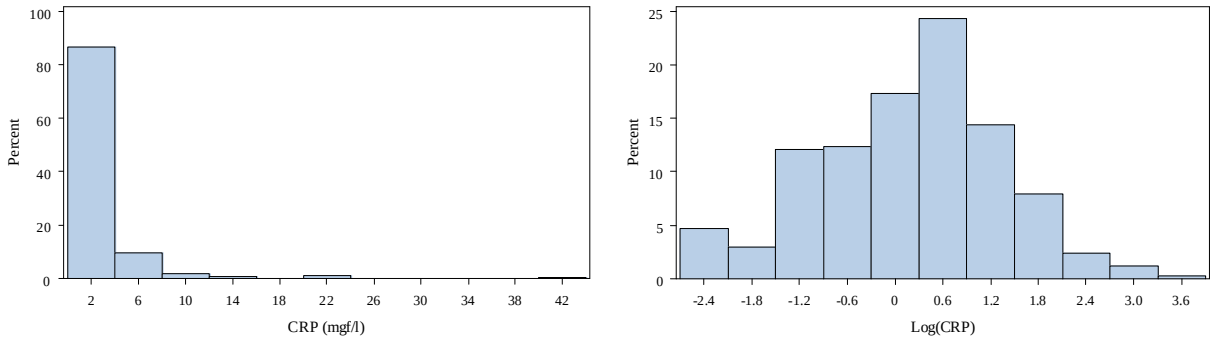
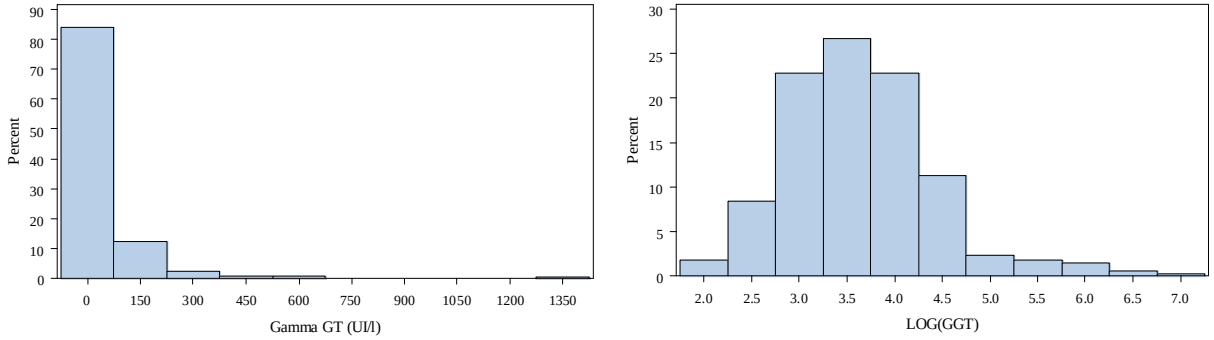


Figure 10 : Histograms of GGT and ln(GGT) distributions



A2. Observed information

Table 9 : Observed correlations (PE=Pearson, PC=Polychoric, PS=Polyserial).

	Age	*Sex*	*Race1*	*lsCD14*	*lCD4*	*lCD8*
Age	*1.000*					
Sex	*-0.226 (PS)*	*1.000*				
Race1	*-0.350 (PS)*	*0.655 (PC)*	*1.00*			
lsCD14	*0.260 (PE)*	*-0.043 (PS)*	*-0.230 (PS)*	*1.00*		
lCD4	*0.047 (PE)*	*-0.042 (PS)*	*-0.228 (PS)*	*-0.145 (PE)*	*1.00*	
lCD8	*0.108 (PE)*	*-0.264 (PS)*	*-0.164 (PS)*	*0.092 (PE)*	*0.237 (PE)*	*1.00*
lB2microgl	*0.124 (PE)*	*-0.222 (PS)*	*-0.264 (PS)*	*0.353 (PE)*	*-0.247 (PE)*	*0.247 (PE)*
lViralLoad	*-0.195 (PE)*	*-0.130 (PS)*	*-0.055 (PS)*	*-0.023 (PE)*	*-0.246 (PE)*	*0.212 (PE)*
lGGT	*0.205 (PE)*	*-0.271 (PS)*	*-0.083 (PS)*	*0.156 (PE)*	*0.021 (PE)*	*0.048 (PE)*
lCRP	*0.169 (PE)*	*-0.061 (PS)*	*-0.075 (PS)*	*0.241 (PE)*	*0.074 (PE)*	*0.179 (PE)*
HDL	*-0.115 (PE)*	*0.568 (PS)*	*0.322 (PS)*	*0.074 (PE)*	*-0.111 (PE)*	*-0.240 (PE)*
BMI	*0.127 (PE)*	*0.301 (PS)*	*0.212 (PS)*	*-0.131 (PE)*	*0.227 (PE)*	*0.113 (PE)*
TimeTreat	*0.388 (PE)*	*-0.091 (PS)*	*-0.224 (PS)*	*0.039 (PE)*	*0.173 (PE)*	*0.045 (PE)*
TimeDiag	*0.461 (PE)*	*-0.119 (PS)*	*-0.352 (PS)*	*0.089 (PE)*	*0.160 (PE)*	*0.118 (PE)*
Treat	*0.077 (PS)*	*0.110 (PC)*	*0.048 (PC)*	*0.018 (PS)*	*-0.115 (PS)*	*-0.031 (PS)*
ViralFail	*0.189 (PS)*	*-0.164 (PC)*	*-0.148 (PC)*	*0.135 (PS)*	*-0.017 (PS)*	*0.194 (PS)*
NadirCD4	*-0.150 (PE)*	*-0.081 (PS)*	*-0.111 (PS)*	*-0.133 (PE)*	*0.463 (PE)*	*0.149 (PE)*

	lB2microgl	*lViralLoad*	*lGGT*	*lCRP*	*HDL*	*BMI*
lB2microgl	*1.00*					
lViralLoad	*0.381 (PE)*	*1.00*				
lGGT	*0.093 (PE)*	*-0.115 (PE)*	*1.00*			
lCRP	*0.101 (PE)*	*-0.088 (PE)*	*0.228 (PE)*	*1.00*		
HDL	*-0.209 (PE)*	*-0.186 (PE)*	*-0.074 (PE)*	*-0.029 (PE)*	*1.00*	
BMI	*-0.142 (PE)*	*-0.083 (PE)*	*0.075 (PE)*	*0.346 (PE)*	*-0.097 (PE)*	*1.00*
TimeTreat	*-0.165 (PE)*	*-0.110 (PE)*	*0.038 (PE)*	*-0.050 (PE)*	*-0.081 (PE)*	*-0.100 (PE)*
TimeDiag	*-0.085 (PE)*	*-0.191 (PE)*	*0.022 (PE)*	*-0.082 (PE)*	*-0.097 (PE)*	*-0.070 (PE)*
Treat	*0.105 (PS)*	*-0.168 (PS)*	*-0.204 (PS)*	*-0.047 (PS)*	*-0.034 (PS)*	*-0.125 (PS)*
ViralFail	*0.150 (PS)*	*-0.059 (PS)*	*0.047 (PS)*	*0.130 (PS)*	*-0.074 (PS)*	*-0.158 (PS)*
NadirCD4	*-0.038 (PE)*	*0.246 (PE)*	*-0.067 (PE)*	*-0.023 (PE)*	*-0.116 (PE)*	*0.159 (PE)*

	TimeTreat	*TimeDiag*	*Treat*	*ViralFail*	*NadirCD4*
TimeTreat	*1.00*				
TimeDiag	*0.846 (PE)*	*1.00*			
Treat	*-0.084 (PS)*	*-0.062 (PS)*	*1.00*		
ViralFail	*0.364 (PS)*	*0.367 (PS)*	*0.349 (PC)*	*1.00*	
NadirCD4	*-0.035 (PE)*	*-0.001 (PE)*	*-0.270 (PS)*	*-0.181 (PS)*	*1.00*

Race1:Black(1) or not(0), lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

Table 10 : Effective sample sizes (univariate in diagonal and pairwise bivariate off diagonal).

	Age	Sex	Race1	lsCD14	lCD4	lCD8
Age	443					
Sex	443	443				
Race1	422	422	422			
lsCD14	443	443	422	443		
lCD4	372	372	353	372	372	
lCD8	372	372	353	372	372	372
lB2microgl	230	230	230	230	203	203
lViralLoad	416	416	396	416	363	363
lGGT	357	357	343	357	316	316
lCRP	340	340	326	340	295	295
HDL	237	237	221	237	211	211
BMI	200	200	189	200	172	172
TimeTreat	357	357	340	357	308	308
TimeDiag	337	337	321	337	293	293
Treat	432	432	411	432	366	366
ViralFail	291	291	280	291	247	247
NadirCD4	422	422	403	422	361	361

	lB2microgl	lViralLoad	lGGT	lCRP	HDL	BMI
lB2microgl	230					
lViralLoad	219	416				
lGGT	207	343	357			
lCRP	208	328	307	340		
HDL	143	231	214	200	237	
BMI	93	187	143	139	91	200
TimeTreat	189	335	298	279	194	148
TimeDiag	176	320	281	265	178	148
Treat	226	407	347	330	230	196
ViralFail	155	272	242	229	161	110
NadirCD4	219	398	339	321	223	188

	TimeTreat	TimeDiag	Treat	ViralFail	NadirCD4
TimeTreat	357				
TimeDiag	325	337			
Treat	357	337	432		
ViralFail	270	246	291	291	
NadirCD4	354	335	422	289	422

Race1:Black(1) or not(0), lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

Table 11 : Observed means and standard deviations (SD).

	Mean	*SD*
Age	*42.508*	*11.921*
Sex	*0.000*	*1.000*
Race1	*0.000*	*1.000*
lsCD14	*7.400*	*0.277*
lCD4	*6.228*	*0.591*
lCD8	*6.699*	*0.474*
lB2microgl	*0.725*	*0.333*
lViralLoad	*4.979*	*3.123*
lGGT	*3.607*	*0.826*
lCRP	*0.218*	*1.125*
HDL	*0.560*	*0.180*
BMI	*25.646*	*4.743*
TimeTreat	*82.899*	*66.366*
TimeDiag	*104.448*	*80.854*
Treat	*0.000*	*1.000*
ViralFail	*0.000*	*1.000*
NadirCD4	*279.634*	*194.361*

Race1:Black(1) or not(0), lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

A3. Estimation of the structural equations model

From initial model, we look at all the parameter's estimations in order to remove one by one non significant relationships. We use a succession of small steps because the removing of one parameter my change completely the other's estimation and if I go too fast, I remove too much parameters. Significance is given by the t-value=estimation/standard error, and we consider that a |t| ratio larger than 2 represents a statically significant departure from 0 (corresponding to α=0.05)

Table 12 : Estimation of the structural equations: final model constitution's steps

Step	*Parameter*	*Path*	*Estimate*	*s.e.*	*t-value (=Estimate/s.e)*
1	γ_{27}	TimeDiag→lCD4	-0.00002	0.0010	-0.0156
2	γ_{63}	Race1→lCRP	-0.0092	0.0080	-0.1167
3	β_{17}	lGGT→lsCD14	-O.0144	0.0315	-0.4570
4	β_{56}	lCRP→lViralLoad	-0.1083	0.2040	-0.5310
5	γ_{39}	ViralFail→lCD8	0.0199	0.0282	0.7057
6	γ_{34}	HDL→lCD8	-0.1645	0.1983	-0.8299
7	γ_{11}	Age→lsCD14	0.0023	0.0025	0.9125
8	β_{71}	lsCD14→lGGT	0.1413	0.1504	0.9399
9	γ_{52}	Sex→lViralLoad	0.2609	0.2008	1.2994
10	γ_{31}	Age→lCD8	-0.0049	0.0032	-1.5506
11	γ_{71}	Age→lGGT	0.0057	0.0035	1.6242
12	β_{25}	lViralLoad→lCD4	0.0471	0.0342	1.3767
13	β_{12}	CD4→lsCD14	-0.1474	0.1081	-1.3637
14	γ_{16}	NadirCD4→lsCD14	0.0002	0.0001	1.2295
15	γ_{61}	Age→lCRP	0.0108	0.0057	0.9131
16	γ_{42}	Sex→lB2microgl	0.0378	0.0198	1.9133
17	β_{21}	lsCD14→lCD4	0.2427	0.2263	1.0728

Race1:Black(1) or not(0), lsCD14=ln(sCD14 level), lCD4=ln(absolute CD4), lCD8=ln(absolute CD8), lB2microgl=ln(β_2-microglobulin), lViralLoad=ln(viral load), lCRP=ln(CRP), lGGT=ln(GGT), HDL: HDL choslesterol, TimeDiag: Time since first diagnosis, TimeTreat: Time since first treatment, ViralFail : History of viral response failure indicator (0=no, 1=yes), Treat : Patient treated with protease inhibitor? (0=no, 1=yes).

Table 13 : Total, direct and indirect effects

Effect	*Total effect*	*Direct effect*	*Indirect effect*
Age on lsCD14	0.0024	-	0.0024
Sex on lsCD14	0.0527	-	0.0527
Race1 on lsCD14	-0.1105	-0.0976	-0.0129
HDL on lsCD14	0.1748	-	0.1748
BMI on lsCD14	0.0010	-	0.0010
TimeTreat on lsCD14	-0.0008	-	-0.0008
TimeDiag on lsCD14	0.0004	-	0.0004
Treat on lsCD14	0.0101	-	0.0101
ViralFail on lsCD14	0.0064	-	0.0064
NadirCD4 on lsCD14	-0.0002	-	-0.0002
Age on lCD4	-0.0001	-	-0.0001
Sex on lCD4	0.0592	-	0.0592
Race1 on lCD4	-0.1569	-0.1449	-0.1198
HDL on lCD4	-0.0154	-	-0.0154
BMI on lCD4	0.0260	0.0324	-0.0064
TimeTreat on lCD4	0.0013	0.0016	-0.0003
TimeDiag on lCD4	-0.0001	-	-0.0001
Treat on lCD4	0.0014	-	0.0014
ViralFail on lCD4	-0.0036	-	-0.0036
NadirCD4 on lCD4	0.0013	0.0014	-0.0001
Age on lCD8	0.0001	-	0.0001
Sex on lCD8	-0.1541	-0.2521	0.0980
Race1 on lCD8	0.0312	0.2835	-0.2513
HDL on lCD8	0.0402	-	0.0402
BMI on lCD8	0.0167	-	0.0167
TimeTreat on lCD8	0.0007	-	0.0007
TimeDiag on lCD8	0.0003	0.0009	-0.0006
Treat on lCD8	-0.0035	-	-0.0035
ViralFail on lCD8	0.0093	-	0.0093
NadirCD4 on lCD8	0.0002	-0.0006	0.0007
Age on lB2microgl	0.0025	0.0030	-0.0005
Sex on lB2micorgl	-0.0105	-	-0.0105
Race1 on lB2microgl	-0.0775	-0.0621	-0.0154
HDL on lB2microgl	-0.2163	-0.2020	-0.01430
BMI on lB2microgl	-0.0018	-	-0.0018
TimeTreat on lB2microgl	-0.0010	-0.0011	0.00003
TimeDiag on lB2microgl	-0.0002	-	-0.0002
Treat on lB2microgl	-0.0049	-	-0.0049
ViralFail on lB2microgl	0.0048	-	0.0048
NadirCD4 on lB2microgl	-0.0001	-	-0.0001
Age on lViralLoad	-0.0432	-O.0549	0.0117
Sex on lViralLoad	-0.1329	-	-0.1329
Race1 on lViralLoad	-0.0940	-	-0.0940
HDL on lViralLoad	-2.5849	-3.5038	0.9189
BMI on lViralLoad	-0.0315	-	-0.0315
TimeTreat on lViralLoad	0.0121	0.0178	-0.0057
TimeDiag on lViralLoad	-0.0141	-0.0167	0.0025
Treat on lViralLoad	-0.2605	-2.2993	0.0388
ViralFail on lViralLoad	0.0530	-	0.0530
NadirCD4 on lViralLoad	0.0027	0.0066	-0.0039

Table 12 (cont.)

Effect	*Total effect*	*Direct effect*	*Indirect effect*
Age on lCRP	0.0063	-	0.0063
Sex on lCRP	-0.2433	-0.4806	0.2373
Race1 on lCRP	-0.0963	-	0.0963
HDL on lCRP	1.1448	1.220	0.0228
BMI on lCRP	0.1075	0.1331	-0.0256
TimeTreat on lCRP	0.0014	0.0043	-0.0029
TimeDiag on lCRP	-0.0036	-0.0063	0.0027
Treat on lCRP	-0.0714	-0.2063	0.1350
ViralFail on lCRP	0.2258	0.2849	-0.0591
NadirCD4 on lCRP	-0.0004	-	-0.0004
Age on lGGT	0.0026	-	0.0026
Sex on lGGT	-0.2839	-0.1809	-0.1030
Race1 on lGGT	-0.0408	-	-0.0408
HDL on lGGT	0.4845	-	0.4845
BMI on lGGT	0.0455	-	0.0455
TimeTreat on lGGT	0.0006	-	0.0006
TimeDiag on lGGT	-0.0015	-	-0.0015
Treat on lGGT	-0.1564	-0.1262	-0.030
ViralFail on lGGT	0.0956	-	0.0956
NadirCD4 on lGGT	-0.0002	-	-0.0002
lCD4 on lsCD14	-0.0595	-	-0.0595
lsCD14 on lCD4	-0.2855	-	-0.2855
lCD8 on lsCD14	-0.2417	-0.3882	0.1465
lsCD14 on lCD8	0.7433	1.2751	-0.5317
lsCD14 on lB2microgl	0.3875	0.3058	0.082
lB2microgl on lsCD14	0.014	-	0.014
lViralLoad on lsCD14	-0.0433	-0.0489	0.0056
lsCD14 on lViralLoad	4.2343	4.6903	-0.4559
lCRP on lsCD14	0.0225	0.0558	-0.0333
lsCD14 on lCRP	0.1326	0.8751	-0.7425
lGGT on lsCD14	-0.01438	-	-0.0144
lsCD14 on lGGT	0.0561	-	0.0561
lCD4 on lCD8	0.3808	0.6704	-0.2896
lCD8 on lCD4	-0.2166	-0.3841	0.1675
lCD4 on lViralLoad	-1.7345	-2.3950	0.6605
lViralLoad on lCD4	0.0007	-	0.0007
lCD8 on lViralLoad	0.2580	1.5482	-1.2902
lViralLoad on lCD8	-0.0019	0.0667	-0.069
lCD4 on lB2microgl	-0.1245	-0.1134	-0.0111
lB2microgl on lCD4	-0.0058	-	-0.0058
lCD8 on lB2microgl	0.0172	0.1049	-0.0877
lB2microgl on lCD8	0.0154	-	0.0154
lCD4 on lCRP	0.0615	-	0.0615
lCRP on lCD4	-0.0125	-	-0.0125
lCD4 on lGGT	0.0260	-	0.0260
lGGT on lCD4	0.0080	-	0.0080
lCD8 on lCRP	-0.1822	-	-0.1822
lCRP on lCD8	0.0327	-	0.0327
lCD8 on lGGT	-0.0771	-	-0.0771
lGGT on lCD8	-0.0209	-	-0.0209
lCRP on lViralLoad	0.1860	-	0.1860
lViralLoad on lCRP	-0.0941	-0.1084	0.0143

A.3. SAS PROC CALIS Program

```
/* corr matrix from lisrel*/
data corr(type=corr);
 infile cards missover;
 input _type_ $ _Name_ $ Age Sex Race1 lsCD14 lCD4 lCD8 lB2micro lViralLo lGGT lCRP HDL BMI Timetrea Timediag Treatmen VirFail NCD4;
 datalines;
    mean  .      42.508 0.000 0.000 7.400 6.228 6.699 0.725 4.979 3.607 0.218 0.560 25.646 82.899 104.448 0.000 0.000 279.634
    std   .      11.921 1.000 1.000 0.277 0.591 0.474 0.333 3.123 0.826 1.125 0.180 4.743  66.366 80.854  1.000 1.000 194.361
    corr  Age     1.000
    corr  Sex    -0.226    1.000
    corr  Race1   -0.350    0.655    1.000
    corr  lsCD14   0.260   -0.043   -0.230    1.000
    corr  lCD4    0.047   -0.042   -0.228   -0.145    1.000
    corr  lCD8    0.108   -0.264   -0.164    0.092    0.237    1.000
    corr  lB2micro 0.124   -0.222   -0.264    0.353   -0.247    0.247   1.000
    corr  lViralLo -0.195   -0.130   -0.055   -0.023   -0.246    0.212   0.381    1.000
    corr  lGGT    0.205   -0.271   -0.083    0.156    0.021    0.048   0.093   -0.115    1.000
    corr  lCRP    0.169   -0.061   -0.075    0.241    0.074    0.179   0.101   -0.088    0.228    1.000
    corr  HDL    -0.115    0.568    0.322    0.074   -0.111   -0.240  -0.209   -0.186   -0.074   -0.029    1.000
    corr  BMI    0.127    0.301    0.212   -0.131    0.227    0.113  -0.142   -0.083    0.075    0.346   -0.097    1.000
    corr  Timetrea 0.388   -0.091   -0.224    0.039    0.173    0.045  -0.165   -0.110    0.038   -0.050   -0.081   -0.100   1.000
    corr  Timediag 0.461   -0.119   -0.352    0.089    0.160    0.118  -0.085   -0.191    0.022   -0.082   -0.097   -0.070   0.846
1.000
    corr  Treatmen 0.077    0.110    0.048    0.018   -0.115   -0.031   0.105   -0.168   -0.204   -0.047   -0.034   -0.125  -0.084   -
0.062    1.000
    corr  VirFail  0.189   -0.164   -0.148    0.135   -0.017    0.194   0.150   -0.059    0.047    0.130   -0.074   -0.158   0.364    0.367
0.349    1.000
    corr  NCD4    -0.150   -0.081   -0.111   -0.133    0.463    0.149  -0.038    0.246   -0.067   -0.023   -0.116    0.159  -0.035   -
0.001   -0.270   -0.181    1.000
run;

/* Initial model with diagonal error varcov matrix*/
proc calis data=Corr(type=Corr) ucov augment tech=NR edf=442 PALL; * sample size+1, PALL=print all, default method is ML;
lineqs
 lsCD14= alpha1 intercept + beta12 lCD4+ beta13 lCD8+beta15 LViralLo+beta16 lCRP+beta17 lGGT
        +gamma11 Age+gamma13 Race1+gamma16 nCD4 +E1,
 lCD4= alpha2 intercept+beta21 lsCD14+beta23 lCD8+beta25 LViralLo
        +gamma23 Race1+gamma25 BMI+gamma26 nCD4+gamma27 timediag+gamma28 timetrea +E2,
 lCD8= alpha3 intercept + beta31 lsCD14+beta32 lCD4+beta35 lViralLo
        +gamma31 Age+gamma32 Sex+gamma33 Race1+gamma34 HDL+gamma36 nCD4+gamma37 Timediag+gamma39 Virfail+E3,
 lB2micro= alpha4 intercept + beta41 lsCD14+beta42 lCD4+beta43 lCD8+beta45 LViralLo
        +gamma41 Age+gamma42 Sex+gamma43 Race1+gamma44 HDL+gamma48 timetrea+E4,
 lViralLo= alpha5 intercept + beta51 lsCD14+beta52 lCD4+beta53 lCD8+beta56 lCRP
        +gamma51 Age+gamma52 Sex+gamma54 HDL+gamma56 nCD4+gamma57 timediag +gamma58 timetrea
        +gamma510 treatmen +E5,
    lCRP= alpha6 intercept + beta61 lsCD14 +beta64 lB2micro+beta65 lViralLo+beta67 lGGT
                + gamma61 Age+gamma62 sex+gamma63 Race1+gamma64 HDL+gamma65 BMI+gamma67 timediag
        +gamma68 Timetrea +gamma69 Virfail +gamma610 treatmen+E6,
              lGGT= alpha7 intercept + beta71 lsCD14+beta76 lCRP
                + gamma71 Age+Gamma72 Sex+gamma710 treatmen+ E7;
 std E1-E7=the1-the7;
 cov E1 E2=0,
   E1 E3=0,
              E1 E4=0,
              E1 E5=0,
              E1 E6=0,
              E1 E7=0,
              E2 E3=0,
              E2 E4=0,
              E2 E5=0,
              E2 E6=0,
              E2 E7=0,
              E3 E4=0,
              E3 E5=0,
              E3 E6=0,
              E3 E7=0,
              E4 E5=0,
              E4 E6=0,
   E4 E7=0,
```

```
            E5 E6=0,
            E5 E7=0,
            E6 E7=0;
run;

/* final model */
ods rtf;
ods graphics on;
proc calis data=Corr(type=Corr) ucov augment tech=NR edf=442 PALL plot=residual; * sample size+1, PALL=print all, default method is ML;
lineqs
 lsCD14= alpha1 intercept + beta13 lCD8+beta15 LViralLo+beta16 lCRP
        +gamma13 Race1 +E1,
 lCD4= alpha2 intercept+beta23 lCD8
        +gamma23 Race1+gamma25 BMI+gamma26 nCD4+gamma28 timetrea +E2,
 lCD8= alpha3 intercept + beta31 lsCD14+beta32 lCD4+beta35 lViralLo
        +gamma32 Sex+gamma33 Race1+gamma36 nCD4+gamma37 Timediag+E3,
 lB2micro= alpha4 intercept + beta41 lsCD14+beta42 lCD4+beta43 lCD8+beta45 LViralLo
        +gamma41 Age+gamma43 Race1+gamma44 HDL+gamma48 timetrea+E4,
 lViralLo= alpha5 intercept + beta51 lsCD14+beta52 lCD4+beta53 lCD8
        +gamma51 Age+gamma54 HDL+gamma56 nCD4+gamma57 timediag +gamma58 timetrea
        +gamma510 treatmen +E5,
    lCRP= alpha6 intercept + beta61 lsCD14 +beta64 lB2micro+beta65 lViralLo+beta67 lGGT
                +gamma62 sex+gamma64 HDL+gamma65 BMI+gamma67 timediag
        +gamma68 Timetrea +gamma69 Virfail +gamma610 treatmen+E6,
             lGGT= alpha7 intercept +beta76 lCRP
                +Gamma72 Sex+gamma710 treatmen+ E7;
 std E1-E7=the1-the7;
 cov E1 E2=0,
   E1 E3=0,
            E1 E4=0,
            E1 E5=0,
            E1 E6=0,
            E1 E7=0,
            E2 E3=0,
            E2 E4=0,
            E2 E5=0,
            E2 E6=0,
            E2 E7=0,
            E3 E4=0,
            E3 E5=0,
            E3 E6=0,
            E3 E7=0,
            E4 E5=0,
            E4 E6=0,
   E4 E7=0,
            E5 E6=0,
            E5 E7=0,
            E6 E7=0;
run;
ods graphics off;
ods rtf close;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Study of the plasma levels of soluble CD14 in HIV infected patients

Richting: **Master of Statistics-Biostatistics**
Jaar: **2012**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

MAES, Nathalie

Datum: **6/09/2012**