



HAL
open science

Analyses intégratives de biomarqueurs immunologiques dans les études épidémiologiques. Applications à trois études cliniques

Marie-Quitterie Picat

► **To cite this version:**

Marie-Quitterie Picat. Analyses intégratives de biomarqueurs immunologiques dans les études épidémiologiques. Applications à trois études cliniques. Santé publique et épidémiologie. Université de Bordeaux, 2015. Français. NNT : 2015BORD0144 . tel-01664948

HAL Id: tel-01664948

<https://theses.hal.science/tel-01664948>

Submitted on 15 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE PRESENTEE POUR OBTENIR LE GRADE DE

**DOCTEUR DE
L'UNIVERSITE DE BORDEAUX**

École doctorale Sociétés, Politique, Santé Publique
SPECIALITE SANTE PUBLIQUE, OPTION EPIDEMIOLOGIE

Par Marie-Quitterie PICAT

**Analyses intégratives de biomarqueurs immunologiques
dans les études épidémiologiques.
Applications à trois études cliniques.**

Sous la direction du Professeur Rodolphe THIEBAUT

Soutenue le 26 octobre 2015

Membres du jury

Mme Geneviève CHENE, PU-PH, INSERM U897, Bordeaux	Président
Mme Dominique COSTAGLIOLA, DR, INSERM - UMR S 1136, Paris	Rapporteur
Mme Laurence MEYER, PU-PH, INSERM - SC10-US019, Paris	Rapporteur
Mme Sylvie CHEVRET, PU-PH, INSERM - UMR S 717, Paris	Examineur
Mr Jean-Luc PELLEGRIN, PU-PH, INSERM U897, Bordeaux	Examineur
Mr Rodolphe THIEBAUT, PU-PH, INSERM U897, Bordeaux	Directeur

TITRE

Analyses intégratives de biomarqueurs immunologiques dans les études épidémiologiques. Applications à trois études cliniques.

RESUME

Les processus biologiques sont nombreux et leurs interactions complexes. Les mesures de ces phénomènes génèrent des biomarqueurs multiples. Ainsi, l'épidémiologie doit évoluer dans ce contexte de données complexes et de nature multidimensionnelle. Les maladies du système immunitaire et les troubles immunologiques qui leur sont associés constituent un bon exemple de pathologies où les questions clinico-épidémiologiques sont de plus en plus complexes, nécessitant des méthodes biostatistiques et épidémiologiques adaptées. Dans cette thèse d'Université, des méthodes permettant de prendre en compte les difficultés méthodologiques générées par les données d'immunologie sont présentées autour de trois applications cliniques. Notre approche consiste en l'utilisation de méthodes intégratives qui prennent en compte l'ensemble des mesures concernant une pathologie donnée. Nous montrons l'intérêt de l'analyse en composantes principales et de la classification hiérarchique ascendante pour résumer et extraire l'information de données multiples de cytométrie en flux et celui des modèles d'équations structurelles pour l'étude de relations complexes entre variables lors de processus multifactoriels. Enfin, via l'exemple d'un modèle de reconstitution immunitaire asymptotique utilisant une fonction exponentielle, nous illustrons l'importance de s'appuyer sur la structure même des données et sur la compréhension des mécanismes biologiques qui sous-tendent la variabilité de ces données dans la réflexion qui concourt au choix d'un modèle statistique. Les méthodes et la réflexion proposées dans cette thèse sont transposables à d'autres domaines d'application avec des données multiples complexes.

MOTS CLES

Données multidimensionnelles, analyse en composantes principales, classification hiérarchique ascendante, modèles d'équations structurelles, analyses intégratives, fonction exponentielle, biomarqueurs, immunologie

TITLE

Integrative analyses of immunological biomarkers in epidemiologic studies. Applications to three clinical studies.

ABSTRACT

Numerous biological processes with potentially complex interactions exist. Measurements of these processes allow to produce multiple biomarkers. Thus, there is a need for epidemiology to evolve within the context of complex and multidimensional data. Immune system diseases and associated immune disorders are an example of a field where clinical and epidemiological issues are increasingly complex, requiring appropriate statistical and epidemiological methods. In this thesis, methods taking into account methodological difficulties generated by immunology data are presented through three motivating examples. The general paradigm of our approach is to take into account all measurements on a given pathology using integrative methods. We propose principal component analysis and hierarchical clustering to summarize multidimensional cytometry data and structural equation modelling for dealing with complex relationships between variables in multifactorial processes. Then, through the example of an asymptotic model of immune reconstitution using an exponential function, we illustrate the importance about the data's structure and the biological mechanisms underlying its variability when building a mathematical model. The methods and the thinking advocated in this thesis are transposable to other research domains with complex data.

KEYWORDS

Multidimensional data, principal component analysis, hierarchical clustering, structural equation modelling, integrative analyses, exponential function, biomarkers, immunology

LABORATOIRE

Centre de Recherche INSERM U897 Epidémiologie et Biostatistiques
ISPED, Institut de Santé Publique, d'Epidémiologie et de Développement
Université de Bordeaux, case 11
146, rue Léo Saignat,
33076 Bordeaux Cedex, FRANCE

REMERCIEMENTS

A Madame le Professeur Geneviève CHENE

Vous me faites l'honneur de présider ce jury de thèse. Je vous remercie de la confiance que vous m'avez accordée en m'accueillant au sein de votre équipe et pour les connaissances et la rigueur scientifique que vous m'avez transmises toutes ces années.

A Madame le Professeur Dominique COSTAGLIOLA

C'est un honneur que vous me faites en acceptant d'être rapporteur de cette thèse et je vous remercie beaucoup de la spontanéité avec laquelle vous avez accepté cette fonction. Soyez assurée de mon plus grand respect.

A Madame le Professeur Laurence MEYER

Vous avez accepté d'être rapporteur de ce travail de thèse. Soyez assurée de tous mes remerciements pour l'intérêt que vous portez à mon sujet et le temps précieux que vous y avez consacré.

A Madame le Professeur Sylvie CHEVRET

Je vous remercie sincèrement d'avoir accepté de juger ce travail et de la confiance que vous m'accordez en m'accueillant très prochainement au sein de votre équipe. Soyez assurée de toute ma reconnaissance.

A Monsieur le Professeur Jean-Luc PELLEGRIN

Vous m'avez transmis l'intérêt pour la Médecine Interne et les Maladies Infectieuses et je suis très heureuse que cette thèse trouve son application dans ces champs de recherche. Vous m'avez soutenue toutes ces années. Votre présence dans ce jury est pour moi une évidence.

A Monsieur le Professeur Rodolphe THIEBAUT

Je te remercie infiniment de m'avoir accordé ta confiance, de ton encadrement toujours présent, de tes précieux enseignements et de ton écoute. Travailler et apprendre à tes côtés a été très enrichissant. Tes conseils, tes encouragements et ton soutien m'ont permis de mener à bien ce travail. Sois assuré de toute ma gratitude.

A mes parents, à ma sœur et à mes grands-parents,

A tous mes amis,

A tous les enseignants ayant contribué à ma formation en épidémiologie clinique, et en particulier au Professeur Roger SALAMON, pour m'avoir soutenue et encouragée toutes ces années,

A l'ensemble de mes collègues,

Aux cliniciens et immunologistes du CHU Bordeaux,

A Nigel KLEIN, Robin CALLARD, Sarah WALKER et Joanna LEWIS, pour m'avoir accueillie à l'Institute of Child Health de Londres,

Aux patients, petits et grands. Je vous dédie ce travail de recherche.

LISTE DES ABREVIATIONS

CD	Clusters de différenciation
CMV	Cytomégalovirus
DICV	Déficit immunitaire commun variable
Ig	Immunoglobulines
OMS	Organisation mondiale de la santé
SEM	Structural equation modelling pour modèles d'équations structurelles
SIDA	Syndrome de l'immunodéficience acquise
VIH	Virus de l'immunodéficience humaine

PRODUCTIONS SCIENTIFIQUES LIEES A LA THESE

Articles originaux publiés dans des revues à comité de lecture

Picat MQ, Thiébaud R, Lifermann F, Delbrel X, Adoue D, Wittkop L, Fauchais AL, Rispal P, Moreau JF, Viallard JF. T-cell activation discriminates subclasses of symptomatic primary humoral immunodeficiency diseases in adults. *BMC Immunol* 2014 Mar 12;15(1):13.

Picat MQ, Lewis J, Musiime V, Prendergast A, Nathoo K, Kekitiinwa A, Ntege PN, Gibb DM, Thiébaud R, Walker AS, Klein N, Callard R and the ARROW trial team. Predicting patterns of long-term CD4 reconstitution in HIV-infected children starting antiretroviral therapy in sub-saharan Africa: a cohort-based modelling study. *PLoS Med* 2013;10(10):e1001542.

Article original soumis dans une revue à comité de lecture

Picat MQ, Pellegrin I, Bitard J, Wittkop L, Proust-Lima C, Liquet B, Moreau JF, Bonnet F, Blanco P, Thiébaud R and the ANRS CO3 Aquitaine Cohort. Integrative analysis of immunological data to explore chronic immune activation in successfully treated HIV patients. Soumis à *AIDS*.

Communication orale dans un congrès international avec comité de sélection

Picat MQ, Pellegrin I, Bitard J, Wittkop L, Proust-Lima C, Liquet B, Moreau JF, Thiébaud R. A structural equation modelling approach to explore the role of interferon α on chronic immune activation in successfully treated HIV-infected patients. 35th Annual Conference of the International Society for Clinical Biostatistics, Vienna, Austria, 24-28 August 2014.

Communication orale dans un congrès national avec comité de sélection

Picat MQ, Pellegrin I, Bitard J, Wittkop L, Proust-Lima C, Liquet B, Moreau JF, Thiébaud R. Rôle de l'interféron- α dans l'activation immune chronique des patients VIH avec une charge virale indétectable sous traitement : une approche par équations structurelles. 8^{ème} Conférence Francophone d'Epidémiologie Clinique, Bordeaux, France, 14-16 Mai 2014.

Communication affichée dans un congrès international avec comité de sélection

Picat MQ, Lewis J, Walker AS, Nathoo K, Musiime V, Kekitiinwa A, Ntege PN, Prendergast A, Callard R, Klein N and the ARROW Trial Team. Long-term profiles of CD4+ reconstitution in HIV-infected children initiating antiretroviral therapy in Uganda and Zimbabwe. 19th Conference on Retroviruses and Opportunistic Infections, Seattle, Etats-Unis, 5-8 Mars 2012.

Financement et prix liés à la thèse

Young Investigator Award (bourse “Jeune chercheur”) pour assister à la Conférence on Retroviruses and Opportunistic Infections, Seattle, Etats-Unis, 5-8 Mars 2012 et présenter la communication affichée « Long-term profiles of CD4+ reconstitution in HIV-infected children initiating antiretroviral therapy in Uganda and Zimbabwe ».

Prix attribué par le Comité Scientifique de la 8^{ème} Conférence Francophone d’Epidémiologie Clinique, Bordeaux, France, 14-16 Mai 2014 pour communication orale « Rôle de l’interféron- α dans l’activation immune chronique des patients VIH avec une charge virale indétectable sous traitement : une approche par équations structurelles. »

« Le hasard ne favorise l'invention que pour des esprits préparés aux découvertes
par de patientes études et de persévérants efforts. »

Œuvres de Louis Pasteur

TABLE DES MATIERES

PRODUCTIONS SCIENTIFIQUES LIEES A LA THESE	7
1. INTRODUCTION	12
1.1. Le système immunitaire	12
1.2. Troubles immunologiques des maladies du système immunitaire.....	13
1.3. Recherche translationnelle et immunologie	14
1.4. Avènement des technologies en immunologie.....	15
1.5. Problèmes méthodologiques liés à l'analyse des biomarqueurs	17
1.5.1. Rationnel du travail de thèse	18
1.6. Equipes d'accueil et contexte de réalisation de la thèse.....	19
2. OBJECTIF ET PLAN DU TRAVAIL DE THESE.....	20
3. ROLE DE L'ACTIVATION IMMUNITAIRE CHRONIQUE AU COURS DES DEFICITS IMMUNITAIRES HUMORAUX PRIMITIFS SYMPTOMATIQUES DE L'ADULTE.....	22
3.1. Problématique.....	22
3.2. Limites des méthodes classiques de régression pour répondre à la problématique	23
3.3. Apport des méthodes exploratoires pour analyse de données multidimensionnelles	25
3.3.1. Analyse en composantes principales.....	25
3.3.2. Classification hiérarchique ascendante	27
3.4. Application	29
3.4.1. Article publié dans BMC Immunology.....	30
3.5. Discussion	54
3.6. Conclusion et perspectives	56
4. DETERMINANTS DE L'ACTIVATION IMMUNITAIRE CHRONIQUE CHEZ LES PATIENTS ADULTES ATTEINTS DE VIH TRAITES EFFICACEMENT .	58
4.1. Problématique.....	58
4.2. Travaux antérieurs utilisant une méthode de régression linéaire multivariable.....	59
4.3. Apport des modèles d'équations structurelles pour étudier des relations complexes entre variables	62
4.3.1. Variables observées et variables latentes	62
4.3.2. Diagramme de chemin	63
4.3.3. Modèle d'équations structurelles complet.....	63
4.3.4. Exemple d'un diagramme de chemin et équations structurelles associées	64
4.3.5. Construction d'un modèle SEM et sélection du modèle	65
4.4. Application	65
4.4.1. Article soumis à AIDS	66
4.5. Discussion	98
4.6. Conclusion et perspectives	100

5. ETUDE DE LA RECONSTITUTION IMMUNITAIRE A LONG TERME, APRES MISE SOUS TRAITEMENTS ANTIRETROVIRAUX, CHEZ LES ENFANTS INFECTES PAR LE VIH DANS LES PAYS A RESSOURCES LIMITEES	102
5.1. Problématique.....	102
5.2. Modélisation longitudinale de la reconstitution immunitaire dans le VIH : exemple de l'adulte.....	103
5.3. Hypothèse d'une reconstitution immunitaire asymptotique chez l'enfant.....	104
5.4. Application	108
5.4.1. Article publié dans Plos Medicine	109
5.5. Discussion	127
5.6. Conclusion et perspectives	129
6. DISCUSSION.....	130
7. CONCLUSION	135
REFERENCES.....	136
ANNEXES.....	155
Annexe 1. Projets pris en charge dans le champ de la médecine interne, des maladies infectieuses et de l'immunologie.....	155
Annexe 2. Représentation des atteintes organiques retrouvées dans le déficit immunitaire commun variable de l'adulte.	157
Annexe 3. Analyse en composantes principales : valeurs propres, variances expliquées par les composantes principales et contribution factorielle des marqueurs immunologiques. Etude ALTADIH.....	158
Annexe 4. Principes de l'index de reclassification.	161

1. INTRODUCTION

1.1. Le système immunitaire

Une brève présentation du système immunitaire et de ses composants est proposée ci-après, afin d'introduire la terminologie immunologique qui sera employée dans ce travail de thèse et d'en faciliter la lecture.

Une des finalités premières du système immunitaire est le maintien de l'intégrité de l'individu ou intégrité du « soi », plaçant l'Immunologie comme une discipline transversale par excellence dans le champ de la Médecine et de la Biologie.

Le système immunitaire est un système complexe de défense de l'organisme contre les agressions infectieuses par les bactéries, les virus et les parasites [Chatenoud et Bach, 2012]. Il joue également un rôle majeur dans la défense de l'organisme contre les tumeurs, à la base des réflexions actuelles sur l'immunothérapie dans le champ de l'oncologie [De Aquino et al., 2015 ; Trinchieri, 2015].

Pour répondre à une agression, l'hôte met en jeu deux lignes de défenses. L'immunité innée correspond à la première ligne de défense de l'organisme. Les acteurs de la réponse immunitaire innée sont les monocytes, les macrophages, les polynucléaires basophiles, éosinophiles, neutrophiles, les cellules natural killer et les cellules dendritiques. Ils déterminent une réponse immunitaire précoce et participent au déclenchement d'une réponse immunitaire secondaire. Les acteurs de cette réponse immunitaire secondaire dite adaptative sont les lymphocytes. Les lymphocytes B sont responsables de la réponse humorale produisant les anticorps ou immunoglobulines (Ig). Les lymphocytes T sont responsables de la réponse cellulaire. Parmi eux, on distingue notamment les lymphocytes T effecteurs auxiliaires CD4⁺ et cytotoxiques CD8⁺, les lymphocytes T régulateurs, impliqués dans les mécanismes de tolérance et les lymphocytes T $\gamma\delta$, qui joueraient un rôle de défense de première ligne face à des agents infectieux. Les cellules dendritiques constituent des cellules clés du système immunitaire, par leur fonction de cellules présentatrices d'antigènes aux lymphocytes. La reconnaissance des antigènes par les lymphocytes naïfs, c'est-à-dire n'ayant pas encore rencontré l'antigène, va permettre leur activation et ainsi leur différenciation en lymphocytes effecteurs, capables d'éliminer l'agent infectieux et en lymphocytes mémoires, cellules à la base des concepts de la vaccination. L'activation et l'expansion des lymphocytes sont ainsi étroitement liées.

La figure 1 ci-après présente de manière synthétique et simplifiée les différents acteurs du système immunitaire.

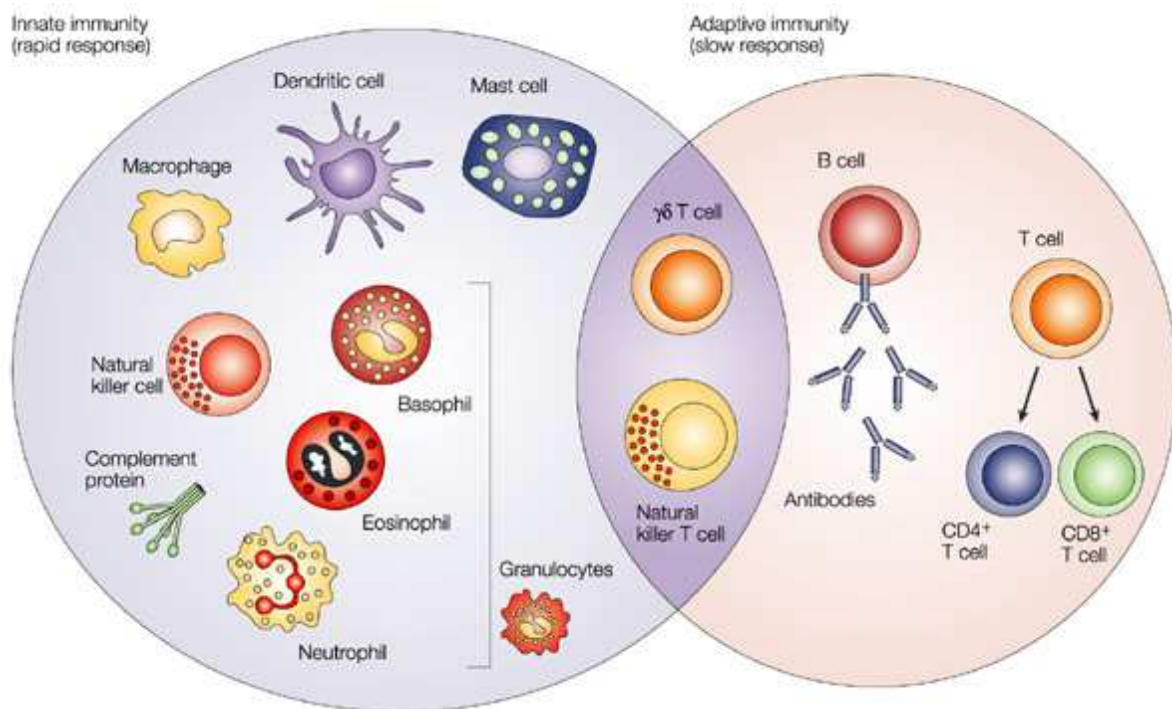


Figure 1. Acteurs des réponses immunitaires innée et adaptative.

D'après Dranoff G. Nat Rev Cancer 2004 Jan;4(1):11-22.

1.2. Troubles immunologiques des maladies du système immunitaire

Une dérégulation du système immunitaire est un des éléments clés de la pathogenèse de nombreuses affections [Chatenoud et Bach, 2012]. Schématiquement, il existe deux situations différentes. La première situation est celle qui est caractérisée par un défaul de fonctionnement du système immunitaire, ce qui se traduit par des déficits immunitaires, primaires [Al-Herz et al., 2014] ou secondaires [Chinen et Shearer, 2010], dont l'exemple le plus connu de déficit immunitaire secondaire ou acquis est lié au Virus de l'Immunodéficience Humaine (VIH). La deuxième situation est celle caractérisée par un fonctionnement par excès du système immunitaire, qui peut se traduire par un état d'activation chronique du système immunitaire, responsable de nombreuses altérations, et par des maladies auto-immunes [Regent et al., 2014 ; Wahren-Herlenius et Dorner, 2013], le lupus érythémateux systémique en étant l'archétype.

Les troubles immunologiques des maladies du système immunitaire recouvrent donc un ensemble de pathologies variées, incluant notamment des pathologies répondant à la définition de maladies rares pour lesquelles la recherche est promue par le Ministère chargé

de la Santé (seuil admis en Europe d'une personne atteinte sur 2000, soit pour la France moins de 30 000 personnes pour une maladie) [Ministère de la Santé, 2015].

En dépit d'une grande hétérogénéité, les maladies du système immunitaire présentent des caractéristiques communes. Elles sont très souvent chroniques et évolutives. Leur physiopathologie reste encore mal connue et elles ont des conséquences substantielles en termes de morbidité, qualité de vie et de mortalité chez les patients qui en sont atteints. La physiopathologie de ces troubles met en jeu les multiples acteurs de l'immunité innée et de l'immunité adaptative (voir Figure 1). La rupture de la tolérance immunitaire envers des antigènes du « soi » peut être due soit à une inactivation des systèmes de régulation ou de tolérance active soit à une « hyper » activation du système immunitaire. Ces troubles recouvrent des mécanismes complexes, les déficits immunitaires et l'auto-immunité ayant récemment été rapportés non pas comme deux entités distinctes mais comme deux processus interconnectés [Grammatikos et Tsokos, 2012 ; Xiao et al., 2014].

1.3. Recherche translationnelle et immunologie

Le concept le plus connu en recherche translationnelle est celui du laboratoire au lit du malade (« bench to bedside »), même si des concepts plus larges sous-tendent la définition de la recherche translationnelle [Fiscella et al., 2008 ; Woolf, 2008]. La recherche translationnelle est souvent présentée comme le chaînon manquant entre la recherche fondamentale et la recherche clinique. Située à l'interface de ces deux disciplines, elle se nourrit des progrès de la connaissance scientifique issus de la recherche fondamentale et utilise la rigueur des outils de la recherche clinique pour traduire en données quantifiables l'observation du malade, tout en respectant sa complexité. La recherche translationnelle a ainsi pour vocation d'assurer un double transfert, des connaissances fondamentales vers des applications cliniques, mais également de l'observation clinique vers la recherche fondamentale.

Ce concept de recherche translationnelle est apparu ces vingt dernières années. L'expression d'un besoin essentiel d'accélérer le transfert des connaissances a notamment donné lieu, pour exemple, à la mise en place, en 2013, d'un Programme de Recherche Translationnelle en Santé, proposé par la Direction Générale de l'Offre de Soins et l'Agence Nationale de la Recherche [PRTS DGOS, 2015].

L'Immunologie a fourni à la Médecine et à la Biologie des concepts à l'origine de solutions prophylactiques et thérapeutiques comme la vaccination, la transplantation et des outils, les anticorps, ayant des applications dans toutes les spécialités. En raison de la place transversale

qu'elle occupe en Médecine et en Biologie, elle s'inscrit logiquement dans les domaines où la recherche doit connaître un essor particulier.

Améliorer les connaissances dans le domaine de l'Immunologie est ainsi un enjeu majeur de la recherche translationnelle [Peakman, 2012]. En particulier, la compréhension des troubles immunologiques des maladies du système immunitaire fait appel à deux disciplines majeures de la Santé Publique, que sont la Biostatistique et l'Epidémiologie, la Biostatistique étant la méthode de raisonnement permettant de prendre en compte des données dont le caractère essentiel est la variabilité et l'Epidémiologie étant la science des problèmes de santé à l'échelle d'une population. Ces deux disciplines ont une place prépondérante dans la recherche translationnelle [Khoury et al., 2010], l'interdisciplinarité étant la clé pour générer des idées originales de recherche et la méthodologie étant un élément majeur dès la conception d'un projet. Le terme d'immuno-épidémiologie a notamment été introduit dans le domaine de la parasitologie en 2001, soulignant la nécessité de cette approche interdisciplinaire incorporant l'Immunologie, la Biostatistique et l'Epidémiologie pour créer de nouvelles perspectives [Hellriegel, 2001].

1.4. Avènement des technologies en immunologie

L'étude des troubles immunologiques des maladies du système immunitaire repose largement sur les examens biologiques.

La cytométrie en flux est une technique incontournable en immunologie. Elle est utilisée pour la numération des très nombreux sous-types cellulaires et permet la mesure de l'état d'activation, de maturation et de prolifération des cellules. Les cellules immunitaires sont comptées grâce à la reconnaissance de marqueurs qu'ils expriment à leur surface, appelés Clusters de Différenciation (CD) (par exemple, le marqueur CD4+ pour les lymphocytes T auxiliaires, qui sont les cellules de prédilection infectées par le VIH ou le marqueur d'activation HLA-DR+ qui sera présenté pour les problématiques d'activation lymphocytaire à l'étude dans ce travail de thèse). Cette technique rapide présente l'intérêt majeur de caractériser les cellules et de les quantifier. Elle trouve ainsi une application indispensable pour le diagnostic, le pronostic et le suivi thérapeutique des maladies du système immunitaire en fournissant un monitoring objectif des cellules de l'immunité innée et adaptative et des troubles immunologiques correspondants. La technique de cytométrie de flux a connu des avancées majeures en ce qui concerne le développement technologique de matériel et de réactifs, permettant ainsi une mesure fine jusqu'à 20 paramètres indépendants sur chaque

cellule, à des dizaines de milliers de cellules par seconde [Chattopadhyay et Roederer, 2012 ; Lugli et al., 2010]. En conséquence, cette technique génère désormais un grand spectre de marqueurs immunologiques pour un même individu donné.

Parallèlement, l'essor des technologies en immunologie s'est accompagné d'avancées majeures concernant les données « omiques » (par exemple, génomique, protéomique, transcriptomique) permettant aux scientifiques de collecter des données multiples inédites sur les gènes, les protéines et le transcriptome [Mason et al., 2014]. En particulier, le débit de production des données de séquençage du génome est de plus en plus rapide, pour un coût de moins en moins élevé [Thiébaud et al., 2014].

L'exploitation de ces données en Biologie et en Médecine suscite de nombreux espoirs tout en engendrant des questions méthodologiques complexes quant à l'intégration adaptée de ces informations multidimensionnelles (voir figure 2), pour produire de nouvelles connaissances.

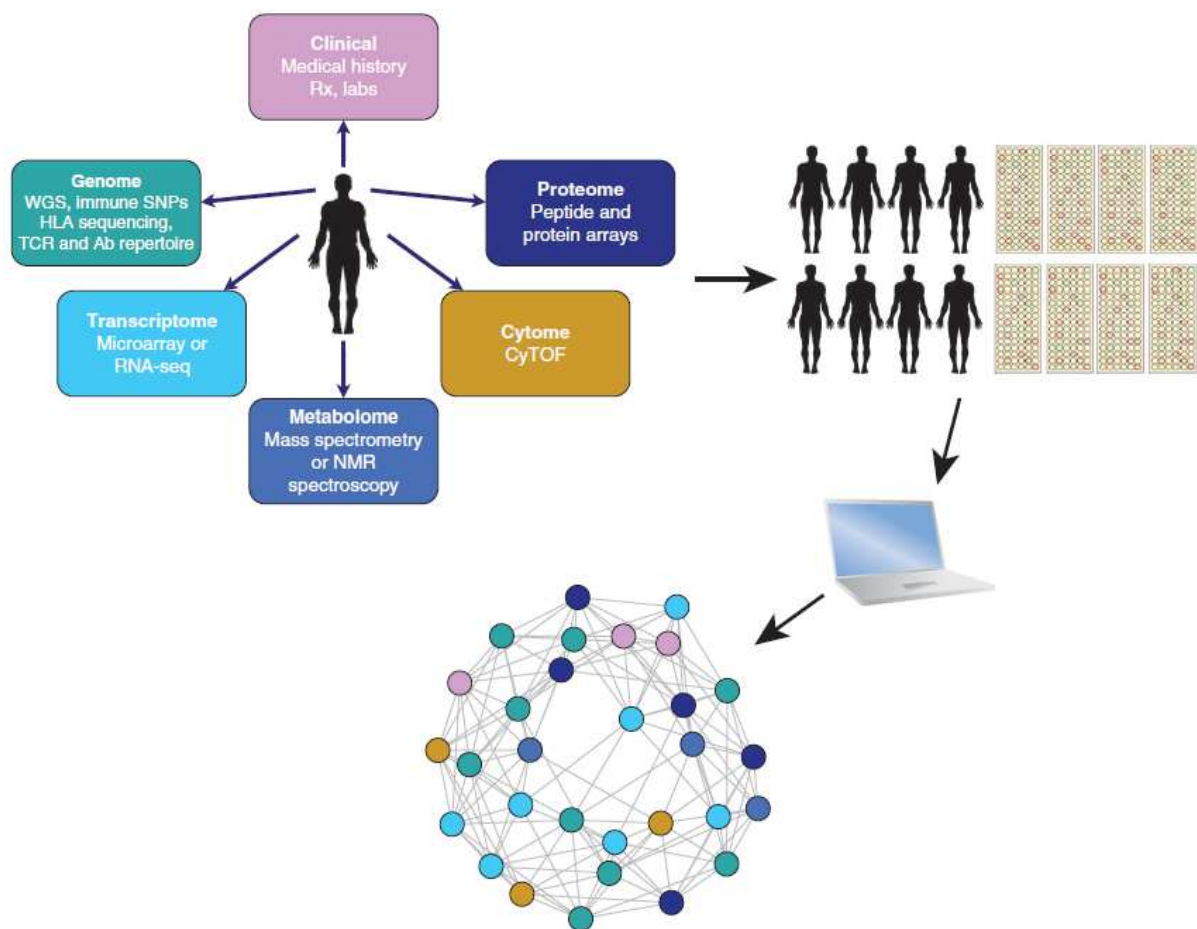


Figure 2. Avènement des technologies en immunologie.

D'après Brian AK et al. Nature Immunology 2014;15:118-127.

1.5. Problèmes méthodologiques liés à l'analyse des biomarqueurs

L'étude des maladies liées à un dysfonctionnement du système immunitaire est une problématique difficile, en raison de l'hétérogénéité et la chronicité de ces maladies. Pour les comprendre, il est nécessaire de prendre en compte à la fois les données cliniques et les données immunologiques pour les aborder sous différents angles.

Le volume et la complexité des données générées par les technologies en immunologie donnent une vue sans précédent de l'ensemble du système biologique, mais ils constituent un défi, en raison des difficultés méthodologiques présentées en termes d'analyse, de modélisation et d'inférence. Tout d'abord, si l'avènement des technologies permet de disposer de plus de données, la multidimensionnalité des données implique des données de nature différentes, en grand nombre et des corrélations entre les différents biomarqueurs mesurés, rendant difficile la compréhension du système à l'étude. Ainsi, dans le contexte actuel de l'épidémiologie clinique (plus de marqueurs et/ou plus de patients), il est particulièrement compliqué d'extraire l'information biologiquement pertinente de données multiples. En parallèle, dans le traitement de telles données, la multiplicité des variables à tester engendre un problème de multiplicité des tests statistiques réalisés, augmentant ainsi la probabilité de trouver un test statistiquement significatif alors qu'il n'y a pas de réelle association dans la population [Thiébaud et al., 2014 ; Sinha et al., 2009]. Dans ce contexte, la « false discovery rate » est définie par la proportion attendue de résultats positifs parmi toutes les hypothèses nulles rejetées. De plus, l'analyse d'un grand nombre de biomarqueurs chez des effectifs potentiellement limités de patients, selon la rareté des maladies à l'étude, peut avoir comme conséquence des difficultés de modélisation statistique dans des modèles classiques de régression (surajustement ou « overfitting », colinéarité des marqueurs) [Yoo et al., 2014]. Se pose également la question de la pertinence de tester séparément ces marqueurs multiples pour les introduire un à un dans des modèles classiques de régression et celle de la meilleure manière de sélectionner un marqueur pertinent parmi ceux disponibles pour apporter un gain par rapport aux données existantes dans les pathologies d'intérêt.

Un autre point est la disponibilité de plusieurs marqueurs immunologiques pour mesurer un même processus immunologique, par exemple le test Quantiféron et le test Elispot pour mesurer la réponse immunitaire induite par le cytomégalovirus. Ce point pose alors la question d'une mesure potentiellement imparfaite du processus immunologique avec les marqueurs séparés ou celle de la pertinence de réaliser des analyses séparées avec chacun des marqueurs disponibles.

Une autre difficulté rencontrée par les épidémiologistes est l'interprétation statistique des associations retrouvées entre ces marqueurs immunologiques et un évènement d'intérêt clinique, qui reste difficile à interpréter en termes de causalité. La méthode scientifique basée sur le raisonnement expérimental implique que les seuls moyens d'accéder à la causalité sont les essais cliniques randomisés contrôlés. Néanmoins, la recherche épidémiologique, incluant celle sur les troubles immunologiques des maladies du système immunitaire, est largement basée sur des études observationnelles. Ces études observationnelles sont nécessaires à la compréhension de certains phénomènes, notamment car certains facteurs ne peuvent pas être manipulés, à la fois pour des raisons éthiques et pratiques [Rutter, 2009]. Au-delà des critères connus de Bradford et Hill pour discuter l'interprétation causale d'une association observée [Hill, 1965], avec des données multiples, les questions d'inférence causale sont devenues plus complexes, l'étude de maladies hétérogènes et multifactorielles impliquant la prise en compte de plusieurs facteurs pouvant interagir entre eux et de déterminants qui s'organisent sur plusieurs niveaux [VanderWeele, 2009]. Dans ce contexte, des approches d'inférence causale apparaissent nécessaires pour mieux comprendre les relations complexes entre variables [Dumas et al., 2014 ; Geneletti et al., 2011].

1.5.1. Rationnel du travail de thèse

L'augmentation de la taille des jeux de données est une problématique croissante en épidémiologie clinique. Cette problématique est particulièrement développée dans le champ de l'Immunologie. Les méthodes statistiques usuelles présentent des limites dans un tel contexte. L'extraction et la compréhension des informations portées par les données multidimensionnelles immunologiques nécessitent de considérer des approches biostatistiques et épidémiologiques autres, plus adaptées ou complémentaires des méthodes classiques. En particulier, notre approche consiste en l'utilisation de méthodes intégratives qui permettent une approche globale du système immunitaire et la prise en compte de l'ensemble des mesures disponibles pour une pathologie donnée.

1.6. Equipes d'accueil et contexte de réalisation de la thèse

Ce travail de thèse a été réalisé à l'interface de deux équipes de recherche :

- l'équipe « Infections par les Virus de l'Immunodéficience Humaine et maladies associées », Centre de Recherche INSERM U897 « Epidémiologie et Biostatistiques » (responsables : Pr Geneviève CHENE / Dr Linda WITTKOP),
- l'équipe « Statistics In System biology and Translational Medicine », INRIA (responsable : Pr Rodolphe THIEBAUT).

Le travail issu d'une collaboration outre-manche, avec mobilité de 6 mois à l'Institute of Child Health, University College London, dans une Unité de Microbiologie et Maladies Infectieuses (Pr Nigel KLEIN, Pr Robin CALLARD), en collaboration avec la Medical Research Council Clinical Trials Unit de Londres (Pr Sarah WALKER), s'est également inscrit dans la thématique de cette thèse.

Cette activité de recherche pour la thèse a été réalisée dans le cadre d'un poste d'assistante hospitalo-universitaire (novembre 2012 – octobre 2015), rattaché à l'Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (responsable : Pr Rodolphe THIEBAUT), du Service d'Information Médicale du CHU de Bordeaux (responsable : Pr Geneviève CHENE). La compréhension des troubles immunologiques des maladies du système immunitaire a été renforcée par les multiples interactions avec cliniciens et immunologistes via la prise en charge méthodologique de plusieurs projets dans le champ de la médecine interne, des maladies infectieuses et de l'immunologie (voir annexe 1).

2. OBJECTIF ET PLAN DU TRAVAIL DE THESE

L'objectif du travail de thèse présenté ici était de proposer et appliquer les méthodes épidémiologiques et biostatistiques les plus adaptées pour répondre à des questions épidémiologiques concrètes, rencontrées dans les troubles immunologiques des maladies du système immunitaire.

La description des travaux réalisés est présentée ci-après.

Des méthodes permettant de prendre en compte les difficultés méthodologiques, liées notamment à la multiplicité des marqueurs, sont présentées autour de trois exemples concrets. Parmi le large spectre des troubles immunologiques des maladies du système immunitaire, les travaux de cette thèse concernent en particulier des recherches appliquées aux champs des déficits immunitaires humoraux primitifs et celui du VIH. Parmi les nombreuses données générées par les technologies en immunologie, les travaux de cette thèse concernent principalement l'analyse des données de cytométrie en flux et incluent également l'analyse de données d'expression génique.

Dans le premier travail (section 3), nous proposons des méthodes exploratoires multidimensionnelles, pour prendre en compte l'ensemble des marqueurs impliqués dans la physiopathologie des déficits immunitaires humoraux primitifs de l'adulte, afin d'évaluer l'importance relative de ces marqueurs pour la classification de ces patients.

Dans le deuxième travail (section 4), nous proposons des modèles d'équations structurelles, basés sur l'utilisation de diagrammes causaux et de variables latentes, pour mieux comprendre les déterminants qui concourent à une activation immunitaire chronique chez des patients adultes atteints de VIH traités efficacement avec une charge virale indétectable (pays du Nord).

Ces deux premiers travaux ont été réalisés à partir de jeux de données, existant au CHU de Bordeaux, issus de Programmes Hospitaliers de Recherche Clinique ou Appels d'Offres Internes concernant des maladies du système immunitaire. Compte tenu du nombre important de marqueurs immunologiques mesurés dans les deux travaux étudiés, ces bases de données ont constitué une opportunité de mettre en application les méthodes proposées.

Enfin, dans le troisième travail (section 5), nous proposons un modèle de reconstitution immunitaire à long terme des lymphocytes T CD4+ après mise sous traitement antirétroviral chez des enfants atteints de VIH dans les pays à ressources limitées. Ce travail a été réalisé

dans le cadre de la collaboration avec mobilité de 6 mois à l'Institute of Child Health, University College London.

Une discussion générale se trouve en section 6 et une conclusion du travail de thèse se trouve en section 7.

3. ROLE DE L'ACTIVATION IMMUNITAIRE CHRONIQUE AU COURS DES DEFICITS IMMUNITAIRES HUMORAUX PRIMITIFS SYMPTOMATIQUES DE L'ADULTE

3.1. Problématique

Les déficits immunitaires humoraux primitifs symptomatiques de l'adulte constituent un groupe très hétérogène de maladies, caractérisées par une hypogammaglobulinémie (déficit en lymphocytes B), entraînant un risque accru d'infections récurrentes ou graves [Al-Herz et al., 2014]. Chez l'adulte, après exclusion des causes secondaires d'hypogammaglobulinémie, trois déficits immunitaires humoraux primitifs symptomatiques prédominent : le déficit immunitaire commun variable [Gathmann et al., 2014 ; Oksenhendler et al., 2008], le déficit en sous-classes d'immunoglobulines G [Preud'homme et Hanson, 1990] et le syndrome de Good (hypogammaglobulinémie associée à un thymome) [Kelesedis et Yang, 2010 ; Malphettes et al., 2015]. Si les manifestations cliniques de ces déficits sont dominées par les infections, d'autres évènements cliniques non infectieux peuvent survenir, plus particulièrement référencés dans le déficit immunitaire commun variable, comme des affections auto-immunes, granulomateuses, des hyperplasies lymphoïdes, des splénomégalies et/ou des entéropathies chroniques qui compliquent ces maladies [Chapel et al., 2008]. Il n'existe pas de traitement curatif de ces déficits. La prise en charge thérapeutique des patients est symptomatique. Si l'administration d'immunoglobulines par voie intraveineuse ou sous-cutanée à doses substitutives permet de diminuer l'incidence et la gravité des infections, elle ne résout pas les complications non infectieuses, qui sont des causes majeures de morbi-mortalité et pour lesquelles la nécessité de traitements immunosuppresseurs est très problématique chez des patients dont l'immunité est altérée [Cunningham-Rundles, 2010]. Enfin, la multiplicité des atteintes organiques de ces déficits (voir annexe 2) nécessite une prise en charge multidisciplinaire. L'information et l'éducation des patients sur ces maladies rares, complexes et non curables, les prédisposant aux infections, restent primordiales.

La nécessité de mettre en commun les compétences pour fédérer l'organisation de la prise en charge des patients et la qualité de la recherche épidémiologique a donné lieu à la création, en 2005, du centre national français de référence des déficits immunitaires dans le cadre du Plan Maladies Rares et, en 2010, à la qualification d'un registre national des déficits immunitaires par le comité national des registres maladies rares [CEREDIH, 2010].

La physiopathologie de ces déficits reste à ce jour encore mal élucidée. Au sein même du déficit immunitaire commun variable, l'hétérogénéité des formes cliniques est responsable d'un retard de diagnostic considérable et suggère une physiopathologie complexe, recouvrant probablement plusieurs maladies ayant des mécanismes différents [Yong et al., 2011] mais qui aboutissent au phénotype commun d'hypogammaglobulinémie [Cunningham-Rundles, 2012]. En parallèle, certains patients atteints de déficit en sous-classes d'immunoglobulines G pourraient développer des complications non infectieuses, similaires à celles observées dans le déficit immunitaire commun variable, ce qui suggère que ces deux entités pourraient avoir des déterminants communs [Barton et al., 2014].

Les déficits immunitaires humoraux primitifs symptomatiques de l'adulte sont abordés dans la littérature par des anomalies qui concerneraient une sous-population du système immunitaire. Les anomalies les mieux répertoriées concernent les lymphocytes B. Elles ont permis d'établir plusieurs classifications, qui se proposent de caractériser des groupes de malades selon le déficit en lymphocytes B [Piqueras et al., 2003 ; Warnatz et al., 2002]. Parmi les nombreuses tentatives de classification, la plus consensuelle est la classification EUROclass existant chez les malades atteints de déficit immunitaire commun variable [Wehr et al., 2008]. Par ailleurs, des associations ont été décrites avec d'autres acteurs du système immunitaire, impliquant la différenciation des lymphocytes T [Giovannetti et al., 2007], l'activation des lymphocytes T [Rezaei et al., 2010 ; Viallard et al., 2006], la régulation des lymphocytes T [Fevang et al., 2007 ; Horn et al., 2009] et les cellules de l'immunité innée [Fulcher et al., 2009 ; Scott-Taylor et al., 2006]. Néanmoins, la classification des déficits immunitaires humoraux primitifs symptomatiques de l'adulte reste basée sur les anomalies des lymphocytes B et les autres composants du système immunitaire n'ont pas été suffisamment pris en compte.

3.2. Limites des méthodes classiques de régression pour répondre à la problématique

La physiopathologie des troubles immunologiques des déficits immunitaires humoraux primitifs symptomatiques de l'adulte met en jeu les acteurs multiples de l'immunité innée et de l'immunité adaptative. De plus, les hypothèses formulées dans la littérature quant à la genèse des complications non infectieuses de ces déficits sont d'ordre immunologique.

L'étude de ces troubles s'intègre donc dans un contexte de marqueurs multiples, proposant l'alternative de plusieurs possibilités de modélisation via les méthodes classiques [Thiébaud et Alioum, 2012] :

- régressions linéaires multivariées pour expliquer les variations de chaque marqueur immunologique (variable à expliquer quantitative) en fonction d'autres marqueurs immunologiques et d'autres variables explicatives,
- régression logistique multivariée, pour expliquer la survenue de complications non infectieuses (variable à expliquer binaire) en fonction d'autres marqueurs immunologiques et d'autres variables explicatives.

Dans ce contexte est clairement illustrée la problématique de multiplicité des tests [Sinha et al., 2009] qu'engendreraient de telles analyses, en particulier des analyses de régressions linéaires multiples pour évaluer chaque anomalie répertoriée (analyses séparées pour évaluer les anomalies des lymphocytes B puis la variation des marqueurs de différenciation des lymphocytes T, de l'activation des lymphocytes T, de la régulation des lymphocytes T et des cellules de l'immunité innée). La réalisation de telles analyses entraînerait une augmentation du nombre de tests à réaliser simultanément et par conséquent la probabilité de détecter des associations à tort. De plus, dans ce contexte d'hypothèses physiopathologiques basées plus spécifiquement sur des biomarqueurs, la réalisation de telles analyses pourrait engendrer un problème de colinéarité entre les variables explicatives quantitatives multiples, lié à l'existence d'une liaison linéaire entre ces variables [Yoo et al., 2014 ; Wittkop et al., 2008]. La colinéarité peut entraîner une instabilité dans l'estimation des coefficients et des variances et, par conséquent, donner des estimations erronées des paramètres du modèle [Elmstahl et Gullberg, 1997].

Au-delà de ces considérations statistiques classiques, une telle approche réalisant de multiples régressions engendre de multiples résultats dont la synthèse peut être particulièrement difficile. Dans le contexte de la problématique des déficits immunitaires humoraux primitifs de l'adulte, les méthodes classiques de régression présentent ainsi plusieurs limites. Des approches alternatives plus adaptées apparaissent ainsi nécessaires pour traiter ces données dans leur ensemble et aborder ainsi ces déficits, non plus par des anomalies qui concerneraient une sous-population du système immunitaire, mais dans leur globalité, pour fournir des informations utiles pour la prise en charge des patients, mieux expliquer la survenue de complications et améliorer les schémas de classification [Park et al., 2008].

Face aux données de grande dimension, deux approches sont souvent proposées [Reinhardt et al., 2010] : soit l'utilisation de méthodes de régularisation (comme les méthodes pénalisées par le Least Absolute Shrinkage and Selection Operator (LASSO)) [Walter et Tiemeier, 2009 ; Tibshirani, 1996 ; Bunea et al., 2011 ; Avalos et al., 2015], soit une réduction de dimension dans une première étape suivie d'une régression plus classique. C'est cette dernière approche que nous proposons pour répondre à la problématique des déficits immunitaires humoraux primitifs symptomatiques de l'adulte, ainsi qu'une méthode de classification, pour prendre en compte la problématique de classification rencontrée dans ces déficits.

3.3. Apport des méthodes exploratoires pour analyse de données multidimensionnelles

Les données de cytométrie en flux, en plus d'être multidimensionnelles, présentent la difficulté d'absence de seuil évident pour définir une positivité. Les données traitées sont donc majoritairement de nature quantitative. Sont présentées, dans cette section, deux méthodes exploratoires d'analyse de données multidimensionnelles, permettant d'intégrer l'ensemble des données immunologiques disponibles : d'une part, l'analyse en composantes principales, qui est une méthode de réduction de dimension adéquate pour traiter des données de nature quantitative et, d'autre part, la méthode de classification hiérarchique ascendante, qui est une méthode d'analyse en clusters.

Ces deux méthodes sont des méthodes non supervisées, qui supposent l'absence d'hypothèses à priori au moment de l'analyse des données [Janes et Yaffe, 2006 ; Van helden, 2013], la finalité étant de mettre en évidence les relations entre variables et d'identifier l'importance relative de certaines parmi l'ensemble des variables considérées. Ces deux méthodes ont été appliquées pour prendre en compte les données de cytométrie en flux multidimensionnelles impliquées dans la physiopathologie des déficits immunitaires humoraux primitifs symptomatiques de l'adulte (section 3.4). Une discussion de ces travaux est proposée en section 3.5.

3.3.1. Analyse en composantes principales

L'analyse en composantes principales est une méthode de réduction de dimension qui permet de décrire et résumer l'information contenue dans des données multidimensionnelles, grâce à des composantes principales, qui sont des variables résumées des variables d'origine [Lebart et al., 2000 ; Ringner, 2008].

Considérons un tableau X de données multidimensionnelles, composé de n individus et de p variables quantitatives.

$$X = \begin{pmatrix} X_{11} & \cdots & \cdots & \cdots & X_{1p} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ X_{n1} & \cdots & \cdots & \cdots & X_{np} \end{pmatrix}$$

Pour réduire la dimension des données tout en conservant les informations du tableau X initial, l'objectif de l'analyse en composantes principales est de rechercher une représentation des n individus dans un sous-espace de l'espace initial. De nouvelles variables k sont ainsi définies, combinaisons linéaires des p variables de l'espace initial. Ces variables sont appelées composantes principales C_k et les axes qu'elles déterminent axes principaux.

Les variables quantitatives d'origine présentant généralement des ordres de grandeur différents, les données considérées sont centrées réduites dans une étape préliminaire à la réalisation de l'analyse en composantes principales. Ainsi, le tableau de données s'écrit désormais \tilde{X} :

$$\tilde{X} = \begin{pmatrix} \frac{X_{11} - \bar{X}_1}{\sigma(X_1)} & \cdots & \cdots & \cdots & \frac{X_{1p} - \bar{X}_p}{\sigma(X_p)} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \frac{X_{n1} - \bar{X}_1}{\sigma(X_1)} & \cdots & \cdots & \cdots & \frac{X_{np} - \bar{X}_p}{\sigma(X_p)} \end{pmatrix}$$

L'analyse en composantes principales est ensuite réalisée à partir de la matrice de corrélation des variables centrées réduites (matrice de corrélation de Pearson pour des variables quantitatives avec une distribution normale, matrice de corrélation de Spearman pour des variables quantitatives avec une distribution non normale). Les composantes principales C_k générées sont ainsi des combinaisons linéaires des variables d'origine.

$$C_k = \beta_{1k} \tilde{X}_1 + \beta_{2k} \tilde{X}_2 + \cdots + \beta_{pk} \tilde{X}_p$$

Ces composantes principales sont non corrélées (les axes déterminés par les composantes principales sont ainsi orthogonaux), de variance maximale afin d'expliquer au mieux la variabilité des données et d'importance décroissante. Les valeurs propres de la matrice de corrélation permettent d'évaluer la part de variance expliquée par chaque composante principale (voir tableau 1 et figure 1 de l'annexe 3 pour exemple). L'interprétation des composantes principales est basée sur les variables apportant la contribution factorielle la plus élevée pour définir chaque composante principale (voir tableau 2 de l'annexe 3 pour exemple). En parallèle, l'orthogonalité des composantes principales présente l'avantage de

pouvoir représenter graphiquement les données, généralement dans un plan à deux dimensions ou encore dans un espace à trois dimensions, définis par les axes principaux des premières composantes principales, résumant le plus l'information (voir figure 1 de l'article section 3.4.1 pour exemple). Le cercle des corrélations correspond à la projection du nuage des variables sur le plan formé par les axes principaux des composantes principales (cercle de rayon 1 du fait de la normalisation des données). Graphiquement, i) plus la variable est proche du cercle, meilleure est sa qualité de représentation, ii) deux variables proches sont corrélées positivement, iii) deux variables qui s'opposent sont corrélées négativement, iiiii) deux variables orthogonales sont non corrélées.

La représentation des individus se déroule selon les mêmes principes, à l'exception de la disparition du cercle unité, la normalisation ne concernant que les variables et pas les individus (voir figure 3 de l'article section 3.4.1 pour exemple).

3.3.2. Classification hiérarchique ascendante

La classification hiérarchique ascendante s'intègre dans les méthodes d'analyse en clusters, qui permettent d'identifier des groupes homogènes partageant des caractéristiques communes (classes ou clusters) [Everitt, 1995]. Le principe des méthodes d'analyse en clusters repose sur la notion de « similarité ». Ces analyses permettent d'identifier des groupes homogènes de sujets aussi bien que de variables. L'analyse en clusters par classification hiérarchique ascendante est une méthode non supervisée qui consiste à rassembler des groupes homogènes partageant des caractéristiques communes en classes, de manière ascendante, jusqu'à l'obtention d'une seule classe. La classification est ascendante car elle part des observations individuelles ; elle est hiérarchique car elle produit des classes de plus en plus vastes, incluant des sous-classes en leur sein. Cette classification est non supervisée car les classes sont définies, non pas au préalable (par une variable à expliquer), mais à partir de la structure même des données [Lebart et al., 2000].

Pour mettre en œuvre une méthode de classification hiérarchique ascendante, différents paramètres nécessitent d'être définis. De la même manière qu'en analyse en composantes principales, les données sont généralement centrées réduites pour prendre en compte des variables quantitatives avec des ordres de grandeur différents. Un autre point est la nécessité d'identifier une distance entre individus, généralement définie par la distance euclidienne (petite si deux individus sont semblables, grande s'ils sont différents). Enfin, il est nécessaire de choisir le critère d'agrégation entre deux classes. Différentes stratégies d'agrégation des classes existent, basées sur des critères de proximité définis par des mesures de distance entre

clusters (méthodes basées sur la distance moyenne, sur les centres de gravité ou méthode de Ward). La méthode de Ward est généralement préférée, car basée sur la variance, et propose de choisir, à chaque étape, le regroupement de classes tel que l'augmentation de l'inertie intra-classes soit minimale [Ward, 1963]. Ainsi, avec cette méthode, en prenant l'exemple d'une classification de patients, deux patients ayant des profils les plus semblables sont regroupés en une classe et, ainsi de suite, jusqu'à ce que les patients soient tous regroupés en une seule classe. L'agrégation des classes cherche toujours à minimiser la variance intra-classes (similarité entre individus d'une même classe maximale) et maximiser la variance inter-classes (similarité entre individus de classes différentes minimale), afin d'obtenir des groupes bien différenciés. La procédure est itérée jusqu'à ce que le nuage soit réduit à une seule classe. La variance des données est ainsi décomposée selon la hiérarchie des classes.

Pour illustration de la méthode de Ward, la décomposition de l'inertie totale est présentée ci-après [Casin, 1999]. Considérons une partition de n individus en K classes, chacune de ces classes comportant n_k individus. La distance euclidienne est notée d . Le centre du nuage des n points est noté O et le centre du groupe k est noté O_k .

L'inertie totale I_T correspond à la moyenne des carrés des distances des n individus au centre du nuage et à la somme de l'inertie inter-classes I_{INTER} et de l'inertie intra-classes I_{INTRA} .

$$I_T = \frac{1}{n} \sum_{i=1}^n d^2(i, O) = I_{INTER} + I_{INTRA}$$

L'inertie inter-classes I_{INTER} correspond à la moyenne pondérée des carrés des distances des centres de chacun des groupes au centre du nuage.

$$I_{INTER} = \frac{1}{n} \sum_{k=1}^K n_k d^2(O_k, O)$$

L'inertie intra-classes I_{INTRA} correspond à la moyenne, pour l'ensemble des individus, des carrés des distances des individus au centre de leur groupe.

La notation $\sum_{i \in k}$ signifie la somme pour tous les individus du groupe k .

$$I_{INTRA} = \frac{1}{n} \sum_{k=1}^K \sum_{i \in k} d^2(i, O_k)$$

L'interprétation de l'analyse est ensuite basée sur la représentation graphique des résultats par un arbre de classification, appelé dendrogramme, qui permet d'identifier le regroupement des patients et le nombre de clusters selon le niveau de hiérarchisation des classes (voir figure 2 de l'article section 3.4.1 pour exemple). En parallèle, il existe des critères objectifs calculés, conçus pour présenter un maximum local au voisinage du nombre optimal de clusters, tel que le « cluster cubic criterion » par exemple, mais le choix final peut aussi se baser sur la mise en cohérence de ces critères objectifs avec la pertinence clinique du nombre de clusters identifiés sur le dendrogramme.

3.4. Application

Dans le cadre d'une étude financée par Programme Hospitalier de Recherche Clinique interrégional (cohorte ALTADIIH du Pr Jean-François Viallard, Service de Maladies Infectieuses et Médecine Interne, CHU de Bordeaux), nous avons à disposition un immunophénotypage de l'ensemble des marqueurs impliqués dans les déficits immunitaires humoraux primitifs symptomatiques de l'adulte. Nous avons appliqué les deux méthodes exploratoires d'analyse de données multidimensionnelles présentées ci-dessus afin d'évaluer, à partir d'une cohorte de 81 patients adultes symptomatiques atteints de déficits immunitaires humoraux primitifs, l'importance relative de 23 marqueurs immunologiques, la finalité étant de sélectionner les marqueurs pertinents qui pourraient être utiles pour le diagnostic et le pronostic. Au moment de notre étude, seules les données d'inclusion de la cohorte étaient disponibles. Nos analyses ont donc consisté en une analyse transversale des données d'inclusion de la cohorte ALTADIIH. L'analyse en composantes principales a été réalisée avec une matrice de corrélation de Spearman. La classification hiérarchique ascendante a été réalisée avec la méthode de Ward. Les données ont été centrées réduites préalablement à la réalisation de ces analyses.

L'idée sous-tendue par notre objectif était d'aborder ces pathologies dans leur globalité, en prenant en compte l'ensemble des acteurs du système immunitaire répertoriés dans la littérature au moment de l'étude ALTADIIH et disponibles. Les 23 marqueurs immunologiques d'intérêt, retenus pour réaliser l'analyse en composantes principales et la classification hiérarchique ascendante, sont présentés dans le tableau 1 ci-après, pour faciliter la lisibilité de l'article présenté en section 3.4.1 et, en particulier, celle de la figure 1 de cet article.

Tableau 1. 23 marqueurs immunologiques sélectionnés. Etude ALTADIH.

Populations immunitaires	Notation pour figure 1 de l'article section 3.4.1
Lymphocytes B	CD19+
Différenciation des CD19+	CD27-IgD+, CD27-IgD- CD27+IgD-, CD27+IgD+
Lymphocytes T	CD3+, CD4+, CD8+
Activation HLA-DR+	CD3+HLA-DR+, CD4+HLA-DR+, CD8+HLA-DR+
Différenciation des CD8+	RA+CCR7+, RA-CCR7+, RA-CCR7-, RA+CCR7-
Marqueur d'immunosénescence	CD8+CD57+
Lymphocytes T régulateurs	Reg T
Lymphocytes T $\gamma\delta$ (Gamma Delta)	GD, GD2
Cellules Natural Killer	NK
Cellules dendritiques	Myéloïdes (mDC), plasmacytoïdes (pDC)
Ratio CD4+/CD8+	CD4+/CD8+ ratio

3.4.1. Article publié dans BMC Immunology

Ce travail a donné lieu à un article original, publié dans BMC Immunology et présenté ci-après.

RESEARCH ARTICLE

Open Access

T-cell activation discriminates subclasses of symptomatic primary humoral immunodeficiency diseases in adults

Marie-Quitterie Picat^{1,2,3,4}, Rodolphe Thiébaud^{1,2,3,4*}, François Lifermann⁵, Xavier Delbrel⁶, Daniel Adoue⁷, Linda Wittkop^{1,2,3,4}, Anne-Laure Fauchais⁸, Patrick Rispal⁹, Jean-François Moreau^{10,11,12} and Jean-François Viallard^{13,14*}

Abstract

Background: Symptomatic Primary Humoral Immunodeficiency Diseases (PHID) constitute a highly heterogeneous group of diseases characterized by a shared hypogammaglobulinemia, resulting in increased risk of recurrent or severe infections. Associations have been described with a variety of immunological abnormalities involving B and T-cell differentiation, T-cell activation and innate immunity. However, PHID discrimination remains based on B-lymphocyte abnormalities and other components of the immune system have not been sufficiently taken into account. We carried out unsupervised and supervised methods for classification in a cohort of 81 symptomatic PHID patients to evaluate the relative importance of 23 immunological parameters and to select relevant markers that may be useful for diagnosis and prognosis.

Results: We identified five groups of patients, among which the percentage of PHID complications varied substantially. Combining the set of markers involved in PHID supported the existence of two distinct mechanisms associated with complications. Switched memory B-cell attrition and CD8+ HLA-DR + activated T-cell increase were the prominent abnormalities observed in PHID complications. Furthermore, in a subgroup of 57 patients with common variable immunodeficiency, the classification that added CD8+ HLA-DR + to the consensual EUROclass classification was better than the EUROclass model in predicting complications.

Conclusion: These results highlight the importance of T-cell activation that may improve discrimination of PHID patients in specific subgroups and help to identify patients with different clinical outcomes.

Keywords: Symptomatic primary humoral immunodeficiency, T-cell activation, HLA-DR marker, Hierarchical clustering, Principal component analysis, Common variable immunodeficiency, IgG subclass deficiency, Good's syndrome

Background

Primary Humoral Immunodeficiency Diseases (PHID) are a heterogeneous group of diseases characterized by a deficit in immunoglobulin (Ig) production, resulting in increased risk of recurrent or severe infections [1]. In adults, after elimination of secondary causes, in

particular lymphoid hemopathies, drug-related causes or renal/digestive leakage of Ig, three major symptomatic PHID have to be entertained: Common Variable Immunodeficiency [2] (CVID), IgG SubClass Immunodeficiency [3] (IgG SD) or Good's syndrome [4]. Although recurrent respiratory tract infections is common to nearly all patients with symptomatic PHID, distinct clinical complications have been described [5-7] according to the occurrence of autoimmune manifestations, lymphoid hyperplasia, chronic enteropathy, splenomegaly and/or granulomatous disease. It has been shown that both outcome and prognosis vary within these

* Correspondence: rodolphe.thiebaud@isped.u-bordeaux2.fr; jean-francois.viallard@chu-bordeaux.fr

¹INSERM, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, Bordeaux F-33076, France

¹³Department of Internal Medicine and Infectious Diseases, Bordeaux University Hospital, Pessac F-33604, France

Full list of author information is available at the end of the article

subgroups of phenotypes in PHID [4,6]. Immunoglobulin substitution reduces the incidence of acute infections but does not solve the appearance of complications that are the major cause of morbidity and death among patients with symptomatic PHID [4,8].

The various attempts of PHID classifications have focused on CVID because of their frequency [2]. Different classification proposals have attempted to define CVID using flow cytometry techniques based on B-cell differential phenotyping [9,10]. Recently, data from the EURO-class group including over 300 European CVID patients, unified previous findings with a classification based on the percentage of B cells among circulating lymphocytes and the percentage of switched memory B cells (smB) among B cells [5]. However, this updated classification does not include other components of immune dysregulation, either causal or consequential, that have been reported in CVID patients. Giovanetti et al. [11] performed an in-depth analysis of the T-cell compartment in CVID patients and demonstrated multiple T-cell abnormalities. T-cell activation is an important process underlined [12,13] that need to be explored, together with T-cell regulation [14,15]. Indeed, gammadelta T cells [16] and innate cell abnormalities reported in Natural Killer cells (NK) [17] or myeloid dendritic (mDC) and plasmacytoid dendritic cells (pDC) [18] may further contribute to the heterogeneous presentation of CVID. Literature on IgG subclass deficiency is poor and immunological characteristics of these patients have not been explored as they have been in CVID. In parallel, Good's syndrome is characterized by hypogammaglobulinemia and thymoma, low or absent B cells, variable defects in cell-mediated immunity with a CD4 T lymphopenia, an inverted CD4/CD8+ T-cell ratio and reduced T-cell mitogen proliferative responses [4].

Because of its intrinsic complexity, distinguishing patients with symptomatic PHID in practice can be challenging. Particularly, the median delay of diagnosis reported in CVID is 7 years [2]. Some patients with IgG subclass deficiency might develop non infectious complications as those observed in CVID [19], suggesting that these two entities could have common determinants. Moreover, IgG subclass deficiency may progress to have typical common variable immunodeficiency [3]. The immunophenotypic evaluation of symptomatic PHID could provide diagnostic clues as well as information useful to manage patients, to predict clinical outcomes [20,21] and to improve classification schemes. The purpose of our study was therefore to characterize subgroups of PHID patients that could be defined by differentiation, activation and regulation markers of B and T-lymphocytes expressing $\alpha\beta$ or $\gamma\delta$ TCR, natural killer cells and dendritic cells using supervised and unsupervised methods of classification appropriate for the high-dimensional data [22] generated by flow cytometry [23,24].

Methods

Design and data collection

The ALTADIH Cohort is a prospective hospital-based cohort of symptomatic PHID patients [1], initiated at the Bordeaux University Hospital and five other public hospitals (Agen, Dax, Limoges, Pau and Toulouse) in Aquitaine, South-Western France. First patients were enrolled in 2007. Patients included were over eighteen years old with primary hypogammaglobulinemia (serum IgG level < 5 g/L for CVID, or IgG subclass deficiency or Good's syndrome). They were already followed for symptomatic or newly diagnosed PHID (Ig level determined before substitution therapy). Differential diagnoses of hypogammaglobulinemia have been excluded and the reduction in the level of the seric gammaglobulins was confirmed at least 3 times at 3-month intervals for each patient, thus excluding transient hypogammaglobulinemia. A standardized questionnaire was filled by physicians at each hospital contact, according to routine clinical management procedures, generally every 6 months. Clinical data collected focused on infections and PHID complications while biological data focused on innate and adaptive markers.

The ALTADIH Cohort was approved by the Bordeaux University Institutional Review Board on December 20th, 2006. Each patient gave informed written consent before participating in the study. Some PHID patients were also enrolled in the French national prospective Cohort DEFI [7].

In this paper, the study is cross-sectional. We only deal with inclusion measurements.

Flow cytometric analysis of peripheral blood lymphocytes

Peripheral blood samples were collected at each visit for systematic lymphocyte phenotyping of PHID patients under or not current immunoglobulin substitution. Blood samples from each contributing center were analyzed centrally at the Bordeaux University Hospital Laboratory of Immunology, by flow cytometry using a FC500 flow cytometer from Beckman-Coulter.

All blood samples were withdrawn on EDTA anti-coagulant in Vacutainers 5 ml tubes (BD, Biosciences, Mountain View, CA) and kept at room temperature until processed. All samples were analyzed fresh following withdrawal within a working day. Following manufacturer recommendations, labeling were carried out on whole blood and red blood cells lysed at room temperature with a Versalyse (Beckman-Coulter, France, ref = A0937) lysing solution added with Iotest fixative solution (Beckman-Coulter, France, ref = A07800). Events were acquired with the dedicated CXP-1 software. In these settings, we did not use any viability staining.

Unless specified, all monoclonal antibodies were purchased from Beckman-Coulter (France) whose catalog

references are reported below and in Additional file 1: Table S1. We used the following panel of monoclonal antibodies (Mab):

- B cells subsets: anti- CD19 (ref: A07766), -CD27 (ref: 6607107), -CD21 (ref: PN IMU473U) and -sIgD (ref: 736000)
- CD8+ T-cell differentiation: anti-CD8 (ref: A07756), -CD45RA (ref: PN IM 271U), -CCR7 from R&D (Minneapolis, MN, USA ref: FAB197A)
- HLA-DR T-lymphocyte activation markers, we used the following panel of Mab: anti-CD45, -3, -4 and -8 (ref: 6607013) -HLA-DR (ref: PNA40579).
- CD25 T-lymphocyte activation marker: anti-CD3, -4 and -8 (ref: 6607013) -CD25 (ref: IM2646)
- CD28 T-lymphocyte activation marker: anti-CD3, -4 and -8 (ref: 6607013) -CD28 (ref: 6607108)
- CD38 T-lymphocyte activation marker: anti-CD3, -4 and -8 (ref: 6607013) -CD38 (ref: A07780)
- regulatory T-cell: anti-CD4 (ref: A07750), -CD25 (ref: IM2646) -CD127 (ref: PN IM1980U)
- NK and B cells: anti-CD45, -56, -19 and -3 (ref: 6607073) -CD16 (ref: A07766)
- myeloid and plasmacytoid dendritic cells, we used the Beckman Coulter kit (ref: A23413 for mDC and A23416 for pDC).
- T cells expressing the gamma/delta TCR: anti-Vdelta2 (ref: PN IM1464) and anti-pan-delta (ref: PN IM1418U).

T and B-cell subpopulation counts were obtained using the flow count beads kit from Beckman-Coulter Flow Count (ref: 7547053) following a lyse and no wash procedure according to the manufacturer's recommendations. Markers of CD4+ T-cell differentiation were not performed in this study.

Results are expressed as percentages of total circulating lymphocytes or of CD19+, CD3+, CD4+ or CD8+ lymphocytes as appropriate, and/or absolute counts in cells/mm³ as appropriate. An exception are dendritic cells, whose numbers are expressed per milliliter of blood, obtained by calculations from the leukocyte count.

Concerning the gating strategies, as examples, we choose to show in Additional file 2: Figure S1, the gating strategy used to assessed CD3+, CD3 + CD4+, CD3 + CD8+, and the proportion of cells expressing HLA-DR among these subsets: in Panel A, applied to a normal healthy donor whereas, in Panel B, applied to a CVID patient. Panel C depicts the B/NK cells phenotyping of the CVID patient.

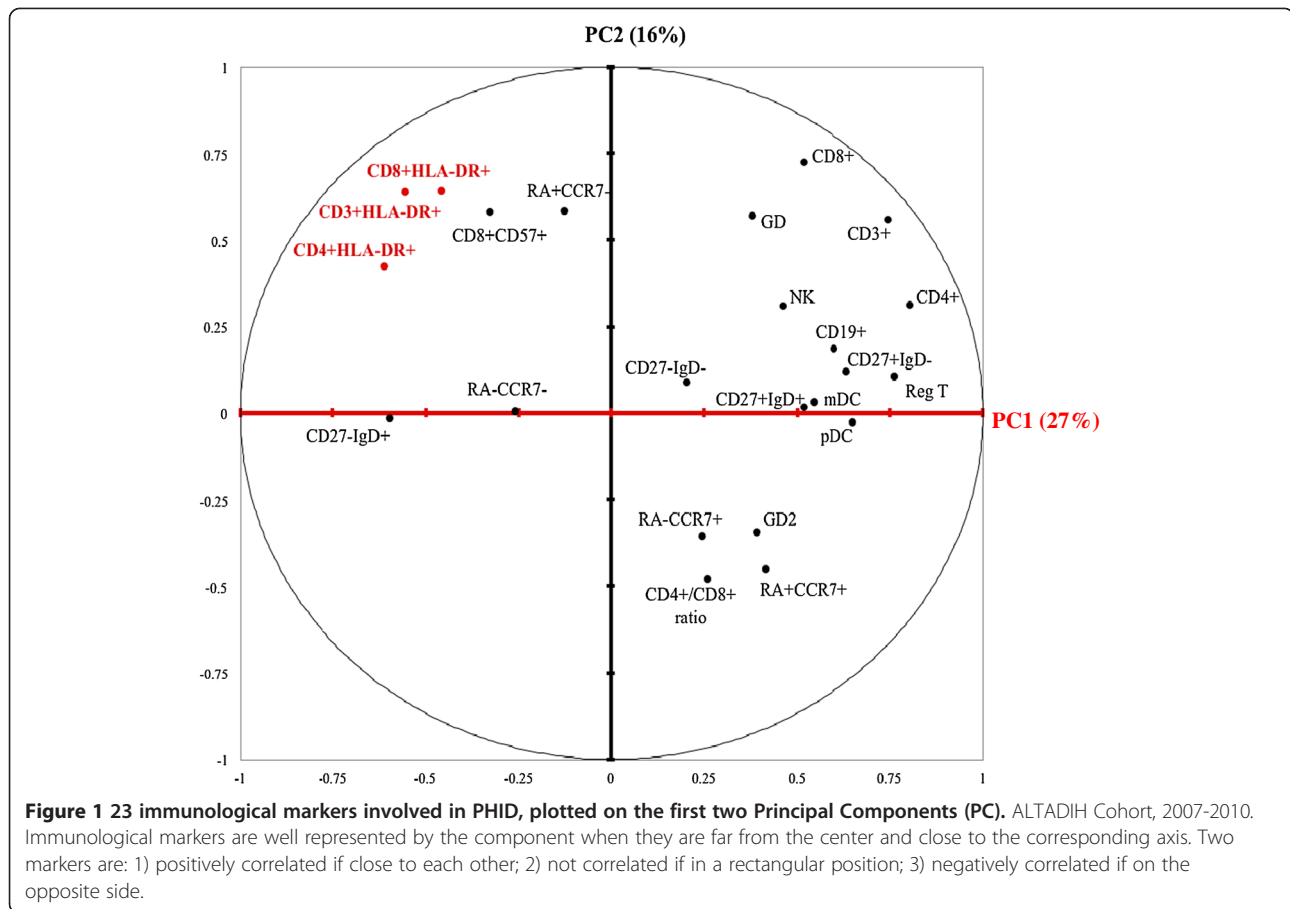
Statistical analyses

Analyses were performed using SAS® 9.1 (SAS Institute, Inc., Cary, NC) and XLSTAT® 2010 software (Addinsoft,

Paris, France). Clinical and therapeutic data were described as numbers and percentages. Immunoglobulin serotypes and immunological markers were described as medians and InterQuartile Range values (IQR). We conducted multidimensional exploratory analyses (principal component analysis and cluster analysis) to study the links between immunological markers and clinical characteristics. To run the multidimensional exploratory analyses, CD19+ B cells, CD3+, CD4+, CD8+ T cells, regulatory T cells, NK cells and gamma/delta T-cells have been used in absolute counts (cells/mm³). CD3 + HLA-DR + activated T cells have been used in percentages of CD3+, CD4 + HLA-DR + in percentages of CD4+, and CD8 + HLA-DR + in percentages of CD8+. Differentiation markers of T-cells and CD8 + CD57+ have been used in percentages of CD8+ and differentiation markers of B-cells have been used in percentages of CD19+. Numbers of dendritic cells were expressed per milliliter of blood and gamma/delta2 T-cells in percentages of gamma/delta T-cells. The use as absolute numbers or percentage has been defined before performing any statistical analyses (see Tables with each immunological marker for a summary of the units used to run the multidimensional exploratory analyses). Robustness analyses have been performed using absolute count in place of percentage and vice versa and results were not substantially different.

Principal component analysis

Principal Component Analysis (PCA) is an unsupervised dimension-reduction method that generates Principal Components (PC), that are linear combinations of the original variables [25] (in our case, represented by differentiation, activation and regulation markers of B and T-lymphocytes, natural killer cells, dendritic cells and T-cells expressing a gamma/delta TCR). Each component explains a part of the variability of the data. We used 2D plots to project data on the plane spanned by the first two components to display 23 immunological markers involved in PHID (missing data excluded). As data run in the principal component analysis have been normalized as referred to the sample (minus sample mean and divided by standard deviation), data are presented on a unit correlation circle of radius 1 (see Figure 1). Because each component is a weighted linear combination of the original variables, a component may have a meaning according to the contribution of each immunological marker. We chose to present the plan spanned by the two major principal components because they explained the most, variability of data (> 10%). On this plan, the contribution of markers is unequal and depends on the contribution of each marker on each axis. A given marker may contribute on the third axis and therefore would not be well represented on the first two axes. PCA was also performed in CVID



patients. A PCA on markers of 12 healthy donors is given in Additional file 3: Figure S2.

Cluster analysis

Hierarchical clustering is a common method used to determine clusters of similar data points [26]. Ward's hierarchical clustering method was used here [27]. This method aggregates consecutively the patients presenting the closest immunological characteristics, by maximizing inter-class variance using Ward's criterion. This process leads to the construction of a classification tree (patients with missing immunological data excluded), which allowed us to identify some distinct subgroups of PHID patients. The projection of the identified clusters onto the first two principal components then helped us characterize them according to immunological markers. A cluster analysis was also performed in CVID patients.

PHID complication analyses

PHID complications were described in each subgroup of patients identified by the cluster analysis in order to provide clinical-immunological links. These complications are those defined by H Chapel et al. [6] to distinguish different clinical phenotypes (autoimmunity, lymphoid hyperplasia

and chronic enteropathy) and granulomatous disease and splenomegaly in addition [12]. Univariable logistic regression was performed to study the association between PHID complications and the 23 immunological markers (variables with a p-value < 0.25 were further tested). A forward-selection procedure was used to build a final multivariable model. The first marker included in the model was the most significant in univariable analysis. The next variable included was the one with most significant association with complications in the remaining set and so on until any additional marker became non significant. As the immunological markers in our study were expressed in different units, we reported the Odds Ratio (OR) per 1 standard deviation for a unit free representation.

CVID substudy

The most consensual classification in symptomatic PHID concerns CVID. EUROclass [5] is the classification commonly used in CVID, segregating patients with nearly absent B cells (less than 1%), severely reduced switched memory B cells (less than 2%), and expansion of CD21^{low} B cells (more than 10%). To evaluate if T-cell markers could improve this classification [28], we evaluated the improvement in predictive value for CVID complications by a net reclassification

approach [29]. To do so, we estimated predicted probabilities of CVID complications using two models: one with EUROclass, the other with EUROclass and CD8+ HLA-DR + (the most significantly associated in our analyses). Based on these two models, we categorized the predicted probabilities into two clinical meaningful categories (probability < 50% or ≥ 50%) and cross-tabulated the classifications. The improvement in reclassification can be quantified as the sum of differences in proportion of individuals moving up minus the proportion moving down for people who developed complications and the proportion of individuals moving down minus the proportion moving up for people who didn't develop complications. Predictive capacity was also evaluated using the area on the receiving operating characteristics curve (c statistics).

Results

Patient characteristics

Sociodemographic and clinical characteristics

Between January 2007 and March 2010, 81 PHID patients were enrolled in the ALTADIH Cohort (among whom 33 were also included in the DEFI Cohort), including 57 CVID patients, 21 IgG subclass deficit patients and 3 patients with

Good's syndrome. At the time of evaluation, none of the PHID patients had evidence of an acute infection. The majority of patients were followed at the Bordeaux University Hospital (66 patients). Baseline epidemiological and clinical characteristics are reported in Table 1. Of the 81 patients, 59% were women. The median age at PHID diagnosis was 41 years old (IQR: 35–54). The most frequent symptoms in disease histories were respiratory tract infections. Altogether, 38% of the patients developed one or more complications: splenomegaly, autoimmune manifestations, granulomatous disease, lymphoid hyperplasia and/or chronic enteropathy (Table 1). More than 80% of the patients were given subcutaneous or intravenous immunoglobulin substitution regimen.

Biological characteristics

Immunoglobulin serum levels before substitution for IgG, IgA and IgM in PHID patients were significantly reduced: baseline levels IgG (median: 4.3, IQR: 3.1–5.3 g/L; normal range: 6.6–12.8 g/L), IgA (median: 0.5, IQR: 0.05–1.2 g/L; normal range: 0.7–3.4 g/L), IgM (median: 0.4, IQR: 0.2–0.8 g/L; normal range: 0.5–2.1 g/L). Specifically, in CVID patients, median baseline IgG was 3.8 g/L (IQR: 1.7–4.4 g/L),

Table 1 Baseline characteristics of 81 PHID patients, ALTADIH Cohort, 2007-2010

Characteristics	Distribution	
Sociodemographic data		
Female, n (%)	48	(59)
Median age at diagnosis in years (interquartile range)		41 (35–54)
Median age at enrolment in years (interquartile range)		46 (38–58)
Clinical data, n (%)		
Respiratory tract infections		
Otitis, sinusitis, nasopharyngitis, polyposis	64	(79)
Bronchitis, pneumonia	65	(80)
Bronchiectasis	21	(26)
Gastrointestinal tract infection		
Chronic diarrhea	16	(20)
Acute diarrhea	13	(16)
Lambliasis	11	(14)
PHID complications: at least one of the following	31	(38)
Lymphoid hyperplasia	16	(20)
Splenomegaly	14	(17)
Autoimmune manifestations*	14	(17)
Granulomatous disease [†]	9	(11)
Villous atrophy	5	(6)
Chronic inflammatory intestinal disease	4	(5)
Therapeutic data, n (%)		
Subcutaneous or intravenous immunoglobulin substitution	66	(81)

*Autoimmune anemia, autoimmune neutropenia, autoimmune thrombocytopenia, Hashimoto thyroiditis, Goujerot-Sjögren syndrome.

[†]Location of biopsy proven granuloma: lymph node, digestive tract, spleen, bone marrow.

median baseline IgA was 0.3 g/L (IQR: 0.05-1.00 g/l) and median baseline IgM was 0.3 g/L (IQR: 0.1-0.6 g/L). Baseline immunological marker values in PHID patients are reported in Table 2 (values of normal patients are reported in Additional file 4: Table S2). Some patients were included in ALTADIH at diagnosis, before the introduction of IgG replacement therapy, and had a longitudinal follow-up of their immunological parameters. The replacement therapy did not change the values of these parameters on two years of monitoring.

Principal component analysis of differentiation, activation and regulation markers of B and T-lymphocytes, natural killer cells, dendritic cells and T-cells expressing a gamma/delta TCR

For studying the association between the various immunological markers reported in the pathogenesis of

PHID, statistical analyses would have been hampered by the high number of markers studied, some of them being highly correlated (e.g. activation markers). In order to circumvent these biases, we used PCA to reduce the multidimensionality of the dataset and capture immunologically relevant information. The first two principal components accounted for 43% of the variability among immunological markers (Figure 1). The first principal component, accounting for 27% of the marker variability, separated regulatory T-cells from activated cells (HLA-DR + on CD3+, CD4+ and CD8+ grouped together) and CD19 + CD27 + IgD- switched memory B cells from CD19 + CD27-IgD + naïve B cells. The second principal component, which accounted for 16% of variability, mainly represented T-cell differentiation, opposing CD45RA + CCR7+ naïve CD8+ T cells and CD45RA + CCR7- terminal effector CD8+ T cells. Activated CD3+,

Table 2 Baseline immunological marker values of 81 PHID patients, ALTADIH Cohort 2007-2010

		Unit	Median	IQR
Total circulating lymphocytes (TCL)		cells/mm ³	1384	(1017, 2000)
B cells	CD19+	cells/mm ³	128	(62, 193)
		% TCL	9.37	(6.21, 12.91)
Naïve	CD27-IgD+	% CD19+	74.01	(56.42, 83.35)
Switched	CD27-IgD-	% CD19+	2.62	(1.86, 5.16)
Marginal zone	CD27 + IgD+	% CD19+	9.42	(4.70, 15.37)
Switched memory	CD27 + IgD-	% CD19+	8.07	(3.63, 19.45)
T cells	CD3+	cells/mm ³	1078	(746, 1585)
	CD4+	cells/mm ³	657	(422, 941)
	CD8+	cells/mm ³	329	(228, 524)
	CD4+/CD8+ ratio		1.99	(1.30, 2.70)
CD8+ T cells				
Naïve	CD45RA + CCR7+	% CD8+	18.01	(6.79, 36.07)
Central memory	CD45RA-CCR7+	% CD8+	1.25	(0.59, 2.48)
Effector memory	CD45RA-CCR7-	% CD8+	34.39	(24.47, 46.90)
Terminal effector	CD45RA + CCR7-	% CD8+	33.48	(23.31, 48.26)
Immunosenescent	CD8 + CD57+	% CD8+	15.97	(8.04, 31.75)
Activated T cells				
HLA-DR	CD3 + HLA-DR+	% CD3+	12.87	(7.99, 20.37)
	CD4 + HLA-DR+	% CD4+	7.96	(5.14, 13.79)
	CD8 + HLA-DR+	% CD8+	24.49	(14.31, 41.11)
Regulatory T cells	CD4 + CD25 + CD127-	cells/mm ³ % TCL	32.11	(16, 63) (0.51, 1.76)
Natural Killer cells	CD3-CD16 + CD56+	cells/mm ³	126	(75, 194)
Dendritic cells				
Myeloid	mDC	/ml	9381	(6597, 15611)
Plasmacytoid	pDC	/ml	4671	(2810, 7427)
Gammadelta cells				
Gammadelta 2 cells		cells/mm ³	33	(15, 72)
		% GD	56.32	(22.19, 77.17)

IQR, interquartile range.

CD4+ and CD8+ T cells expressing HLA-DR+ were positively correlated to each other in PCA (Spearman rank correlation coefficient $r=0.88$ between CD3+ HLA-DR+ and CD4+ HLA-DR+; $r=0.90$ between CD3+ HLA-DR+ and CD8+ HLA-DR+; $r=0.70$ between CD4+ HLA-DR+ and CD8+ HLA-DR+). HLA-DR+ activation markers were also correlated to CD45RA+CCR7-terminal effector CD8+ T cells and to CD8+CD57+, known to be a marker of senescence [30]. Hence, this first analysis and an additional analysis extended to 38 markers available on 72 patients including CD21+ B cells (see Additional file 5: Figure S3) led us to conclude that all activation markers were highly correlated and one of them could be chosen to reflect alone the T-cell activation process. In addition, the first two components defined an upper-left corner grouping HLA-DR+ activation markers with CD45RA+CCR7-terminal effector CD8+ T cells and the CD8+CD57+ senescence marker.

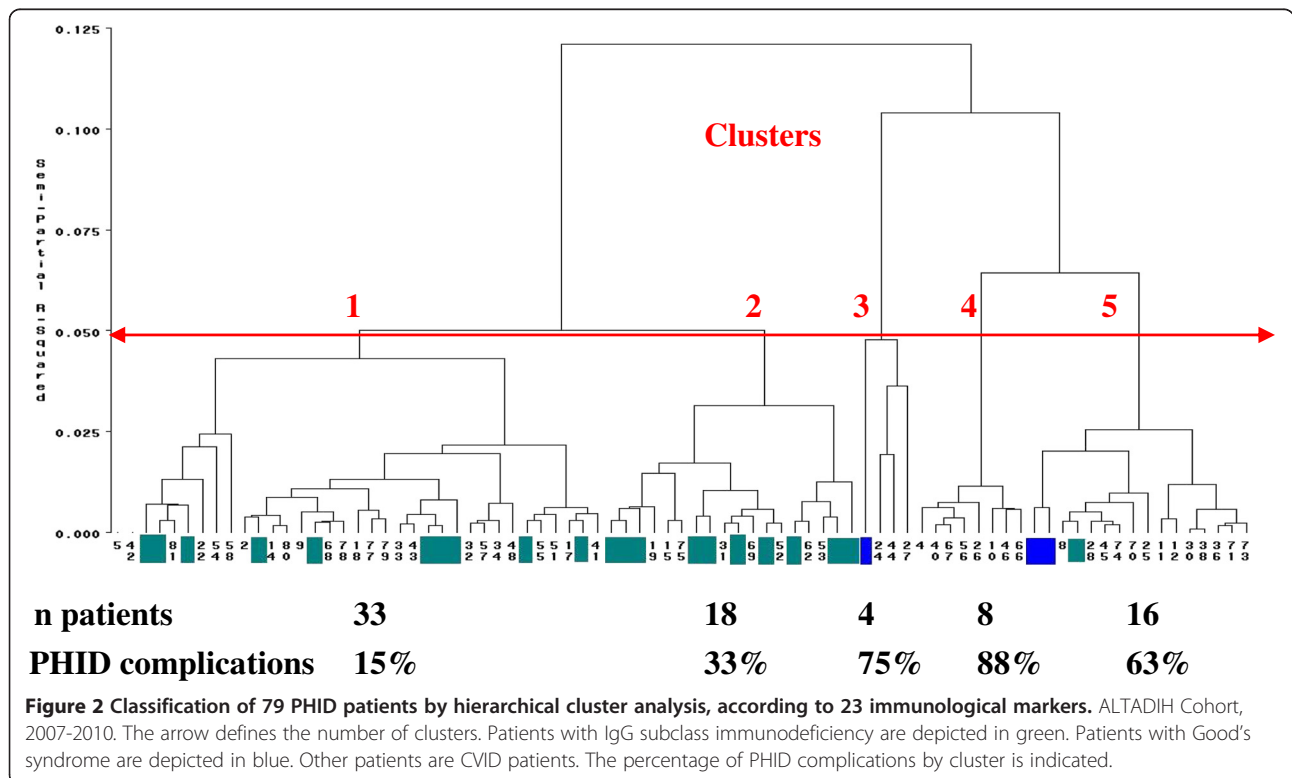
PCA using 23 immunological markers of CVID patients only (excluding IgG SubClass Immuno deficient and Good's patients) gave consistent results as those observed with the whole PHID population.

Classification of PHID patients according to cluster analysis

Next, an unsupervised classification was performed to define groups of patients that would be more homogeneous

according to immunological markers. Hierarchical cluster analysis could be performed for 79 PHID patients of the 81 in the ALTADIH cohort (2 CVID patients not classified). Based on cubic clustering criterion, we would have chosen two clusters but the identification of five populations of PHID patients, as revealed by the five branches of the tree down stream (Figure 2), was chosen as these five clusters were clinically relevant. Clusters 1 and 2, made up by patients with CVID or IgG subclass immunodeficiency, were grouped together, separated from clusters 3, 4 and 5, mainly made up by patients with CVID or either Good's syndrome. In the cluster analysis in CVID diagnoses (excluding patients with Good's syndrome and IgG subclass immunodeficiency), the patients were grouped in the same way than in the previous cluster analysis performed on the whole population (identification numbers of CVID patients given for detailed parallel between Figure 2 and Additional file 6: Figure S4).

The projection of PHID patients, classified in the above 5 defined clusters, onto the plan spanned by the first two principal components led to define their immunological characteristics (Figure 3 that needs to be read in conjunction with Figure 1; Table 3). For example, patients plotted in the upper-left corner of Figure 3, mainly patients of cluster 4, are characterized by variables plotted in the same quadrant in Figure 1, in particular HLA-DR+ markers. Thus, the 33 patients making up cluster 1 (23 CVID patients and 10 IgG SD



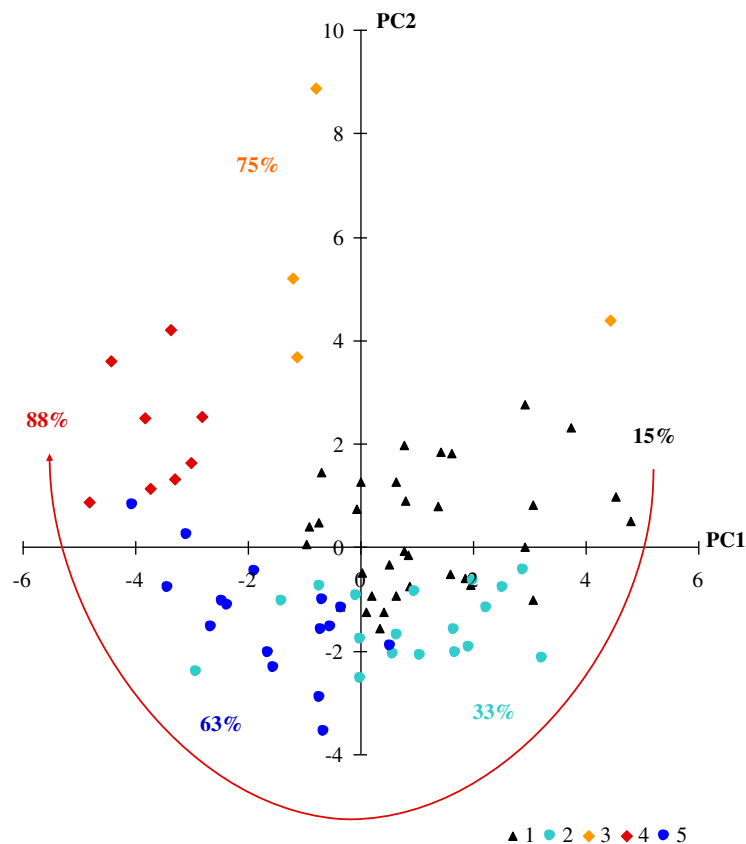


Figure 3 Immunological interpretation of 5 clusters of CVID plotted on the first two principal components. ALTADIH Cohort, 2007-2010. The percentage of CVID complications, noted next to the clusters, increased in a clockwise manner as phenotypes approach the projection of activated HLA-DR + markers plotted in Figure 1 (red arrow).

patients) were characterized by normal innate and adaptive immunity. They were represented on the upper-right corner that grouped innate cells, CD19+ B-cells, CD19 + CD27 + IgD- switched memory B-cells, CD3+, CD4+, CD8+ T-cells, and regulatory T-cells (Figure 1). Cluster 2 (lower-right corner, Figure 3) represented 18 patients (8 CVID patients and 10 IgG SD patients) characterized by increased naïve CD8+ T-cells. Cluster 1 and cluster 2 looked similar and were combined in an upstream clustering. Cluster 3 (1 patient with Good's syndrome and 3 CVID patients) and cluster 4 (8 CVID patients) were characterized by an increase of HLA-DR + activation, terminal effector CD8+ T cells and immunosenescence marker (Table 3). As expected, these patients were represented in the upper-left corner near CD3+, CD4+ and CD8+ T cells expressing HLA-DR+, CD45RA + CCR7-terminal effector CD8+ T cells, and the CD8 + CD57+ marker (Figure 1). In cluster 3, one CVID patient had been splenectomized before enrolment (in 2001). In this patient, splenectomy deeply affects the absolute numbers but not grossly the percentages of the lymphocyte subpopulations in the blood. In parallel, the patient with Good's

syndrome in cluster 3 was differently plotted, i.e. in the upper-right corner near CD8+, reflecting the high blood CD8+ T-cell absolute numbers known in Good's syndrome. Cluster 5 (lower-left corner, Figure 3) consisted of 16 patients with B-cell deficiency (13 CVID, 1 IgG SD and 2 Good's syndrome): the lower-left corner where no immunological parameters are plotted has to be interpreted as the symmetrical and diagonal opposite of the upper-right corner. Thus, the meaning of cluster 5 (lower-left corner, Figure 3) is a cluster gathering B-cell deficiencies because the patients are plotted on the opposite corner of the plot of CD19+, CD27 + IgD+, CD27 + IgD-, CD27-IgD- which all are markers plotted in the upper-right corner. Table 3 confirms the given interpretation of the clusters.

PHID complications according to clusters and T-cell activation

Interestingly, the proportion of patients with PHID complications (information not used in the previous analysis) varied significantly between the clusters: 15%, 33%, 75%, 88% and 63% in clusters 1 to 5 respectively (Figure 3, Table 4).

Table 3 Baseline immunological characteristics of 5 clusters of 79 PHID patients, ALTADIH Cohort, 2007-2010

		Cluster 1 n = 33	Cluster 2 n = 18	Cluster 3 n = 4	Cluster 4 n = 8	Cluster 5 n = 16
IgG	g/L	4.37 (3.80, 5.32)	5.73 (4.7, 6.3)	4.68 (2.69, 5.29)	1.44 (0.30, 2.71)	3.69 (1.65, 4.22)
IgA	g/L	0.63 (0.14, 1.62)	1.02 (0.07, 1.55)	0.60 (0.04, 2.09)	0.15 (0.01, 0.43)	0.19 (0.05, 0.79)
IgM	g/L	0.55 (0.31, 1.09)	0.48 (0.24, 0.78)	0.54 (0.34, 0.62)	0.16 (0.05, 0.42)	0.27 (0.20, 0.47)
B cells						
CD19+	cells/mm ³	157 (128, 244)	128 (102, 211)	49 (20, 87)	122 (94, 176)	51 (29, 119)
CD27-IgD+	% CD19+	68.85 (52.03, 74.47)	68.89 (57.94, 77.24)	49.52 (10.73, 80.43)	90.89 (82.94, 93.31)	84.03 (70.93, 90.82)
CD27-IgD-	% CD19+	2.80 (2.18, 4.38)	3.90 (2.24, 6.42)	6.13 (2.64, 23.40)	2.40 (1.67, 2.58)	2.65 (1.72, 4.22)
CD27 + IgD+	% CD19+	15.01 (10.11, 23.19)	10.53 (4.42, 13.73)	5.67 (2.43, 8.57)	3.57 (2.21, 7.20)	4.92 (1.32, 7.92)
CD27 + IgD-	% CD19+	14.39 (7.08, 20.05)	13.93 (7.26, 24.33)	7.77 (2.43, 19.38)	2.55 (0.94, 3.91)	2.97 (0.36, 6.66)
T cells						
CD3+	cells/mm ³	1363 (1076, 1787)	1045 (883, 1456)	2756 (1472, 4183)	831 (658, 1075)	579 (407, 837)
CD4+	cells/mm ³	835 (571, 1214)	789 (573, 967)	848 (609, 1227)	382 (302, 565)	308 (244, 565)
CD8+	cells/mm ³	477 (358, 604)	270 (159, 383)	1621 (710, 2682)	293 (196, 362)	213 (141, 266)
CD4+/CD8+ ratio		1.89 (1.45, 2.42)	2.77 (2.31, 4.08)	0.60 (0.42, 1.06)	1.26 (0.67, 2.65)	1.89 (1.17, 2.61)
CD8+ T cells						
CD45RA + CCR7+	% CD8+	19.05 (11.11, 34.59)	28.47 (13.32, 43.67)	1.52 (0.56, 2.75)	4.91 (1.53, 53.19)	35.56 (10.29, 53.19)
CD45RA-CCR7+	% CD8+	0.83 (0.46, 1.90)	4.60 (1.96, 7.97)	0.53 (0.21, 0.82)	0.82 (0.30, 1.20)	1.37 (0.85, 1.96)
CD45RA-CCR7-	% CD8+	31.28 (24.26, 36.96)	43.13 (32.60, 57.14)	36.82 (16.56, 57.25)	39.24 (30.69, 59.20)	29.14 (17.47, 42.55)
CD45RA + CCR7-	% CD8+	41.48 (32.61, 52.32)	22.30 (16.87, 27.10)	40.75 (13.96, 59.90)	48.28 (30.47, 63.41)	29.25 (18.76, 39.37)
CD8 + CD57+	% CD8+	17.52 (9.42, 30.97)	7.38 (4.41, 15.97)	27.36 (19.36, 47.30)	39.61 (33.78, 51.16)	13.51 (8.35, 23.44)
Activated T cells						
CD3 + HLA-DR+	% CD3+	12.04 (7.62, 16.95)	9.87 (7.11, 12.83)	52.00 (32.71, 59.56)	46.56 (40.76, 57.25)	13.65 (6.19, 20.33)
CD4 + HLA-DR+	% CD4+	5.88 (4.94, 9.17)	6.75 (4.95, 9.13)	48.30 (26.32, 56.25)	43.87 (30.88, 49.95)	11.54 (6.52, 18.33)
CD8 + HLA-DR+	% CD8+	25.48 (14.86, 31.93)	19.73 (14.18, 26.44)	53.48 (37.05, 65.52)	62.14 (47.01, 76.39)	15.12 (10.65, 28.64)
Regulatory T cells	cells/mm ³	54 (33, 76)	41 (24, 82)	18 (13, 22)	19 (12, 24)	15 (8, 18)
NK cells	cells/mm ³	143 (100, 214)	174 (121, 230)	268 (146, 388)	75 (55, 103)	41 (26, 70)
Dendritic cells						
Myeloid	/ml	11940 (8751, 17805)	11422 (6616, 13457)	9122 (5662, 29989)	7795 (4086, 10548)	5420 (4107, 9683)
Plasmacytoid	/ml	7217 (4801, 11084)	6152 (3599, 9836)	2771 (1766, 8490)	3073 (1351, 3618)	2766 (2194, 4255)
Gammadelta cells	cells/mm ³	44 (20, 74)	27 (11, 54)	396 (178, 579)	19 (6, 92)	16 (9, 26)
Gammadelta 2	% GD	62.96 (40.85, 73.68)	78.71 (75.00, 89.39)	13.40 (8.77, 30.62)	6.59 (4.09, 25.61)	25.00 (12.41, 55.96)

Results are expressed as medians (interquartile range values).

As shown by the dendrogram (Figure 2), clusters 3, 4 and 5 were closer to each other than clusters 1 and 2, in which the rate of complications was the lowest. Among the 8 patients in cluster 4, 7 (88%) presented with a splenomegaly (14 splenomegaly recorded in the cohort) and 4 (50%) with a granulomatous disease (9 granulomatous diseases recorded in the cohort). Patients in cluster 5, characterized by a severe loss of switched memory B cells, also presented a high percentage of complications (63%) (Figure 3, Table 4). In Figure 3, the highest percentages of complications (from clusters 3 and 4) are in the upper-left corner, and the lowest (from cluster 1) on the upper-right. Hence, visually, complication rates increased in a clockwise manner.

To quantify the independent effects of immunological markers, we then performed a supervised analysis based on a logistic regression model. As the activated HLA-DR + CD3+, CD4+ and CD8+ T-cells were correlated positively in the PCA, CD8+ HLA-DR + was chosen for the ascendant multivariable analysis (as it was the most significantly associated with PHID complications in univariable analysis (Table 5). The final model contained the following three markers. Increased CD8+ HLA-DR + T-cells were strongly associated with PHID complications (odds ratio per one standard deviation = 2.55; confidence interval [1.33; 4.89]; p-value = 0.0047). Both decreased plasmacytoid dendritic cells (odds ratio per one standard

Table 4 Type of PHID and PHID complications in 5 clusters of 79 PHID patients, ALTADIH Cohort, 2007-2010

Type of PHID	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	n = 33		n = 18		n = 4		n = 8		n = 16	
	n	%	n	%	n	%	n	%	n	%
CVID	23	(70)	8	(44)	3	(75)	8	(100)	13	(82)
IgG subclass deficiency	10	(30)	10	(56)	0	(0)	0	(0)	1	(6)
Good's syndrome	0	(0)	0	(0)	1	(25)	0	(0)	2	(12)
PHID complications (at least one of the following)	5	(15)	6	(33)	3	(75)	7	(88)	10	(63)
Lymphoid hyperplasia	2	(6)	5	(28)	1	(25)	3	(38)	5	(31)
Splenomegaly	1	(3)	2	(11)	1	(25)	7	(88)	3	(19)
Autoimmune manifestations	3	(9)	1	(5)	1	(25)	2	(25)	7	(44)
Granulomatous disease	0	(0)	1	(5)	2	(50)	4	(50)	2	(13)
Villous atrophy	0	(0)	2	(11)	0	(0)	0	(0)	3	(19)
Chronic inflammatory intestinal disease	0	(0)	2	(11)	1	(25)	1	(13)	0	(0)

deviation = 0.49; confidence interval [0.25; 0.96]; p value = 0.0363) and decreased switched memory B cells (odds ratio per one standard deviation = 0.51; confidence interval [0.28; 0.94]; p value = 0.0312) were also associated with PHID complications.

Classification of CVID patients according to EUROclass and CD8+ HLA-DR + marker

In symptomatic PHID, the only disease for which a consensual classification exists is CVID. Among the 57 CVID patients in the ALTADIH cohort, EUROclass distinguishes 1 patient with equal or less than 1% of B cells of lymphocytes (B- group) from the 56 patients with a higher percentage (B + group). B + patients were divided into 15 patients with severe deficiency of class-switched memory B cells ($\leq 2\%$ of B cells, smB- group) and 40 patients with more than 2% of class-switched memory B cells (smB + group). The smB status of one patient could not be evaluated. EUROclass also discriminates between patients according to the expansion of CD21^{low} B cells: above or below 10% within B cells (CD21^{low} versus CD21^{norm}). In our study, among 14 PHID patients with splenomegaly, 11 (79%) were classified in the CD21^{low} group.

As we found a strong association between PHID complications and increased CD8+ HLA-DR + T-cells (expressed as percentages of CD8+ lymphocytes) and because CD8+ HLA-DR + has been involved in CVID [12], we examined whether the inclusion of CD8+ HLA-DR + in addition to EUROclass criteria, would improve the prediction of CVID complications (Table 6). Among 19 patients with CVID complications and classified with less than 50% risk using EUROclass, 10 were reclassified with a risk greater than 50% when adding the CD8+ HLA-DR + activation marker. Moreover, the addition of CD8+ HLA-DR + to the model increased its predictive value (c statistic 0.57 without

CD8+ HLA-DR + versus 0.77 with CD8+ HLA-DR+, a value close to 1 indicating a good prediction). Specifically, adding CD8 + HLA-DR + to EUROclass model leads to a substantial increase in sensitivity (from 24% to 60%) and some loss of specificity (from 90% to 80%) (Additional file 7: Tables S3 and Additional file 8: Table S4).

Discussion

In the present study, we focused on three PHID in adults with symptomatic implications: CVID, IgG SD and Good's syndrome (referred in the 2011 update of Primary Immunodeficiency by the International Union of Immunological Societies [1]). We used unsupervised and supervised statistical methods on a cohort of 81 PHID patients in an attempt to select relevant markers that may be useful for diagnosis and prognosis of these patients.

Our rationale to pool patients diagnosed with CVID, IgG SubClass Immuno deficient and Good's patients, and thus, to include in the ALTADIH cohort the active file of PHID patients with symptoms in Aquitaine, South-Western France, was based on the following arguments: 1) Because of its complexity, distinguishing patients with symptomatic PHID can be challenging for the diagnosis and the treatment. That is why we did not focus only on CVID patients, and we wanted to address the symptomatic PHID population seen in clinical routine practice by physicians, including IgG SubClass Immuno deficient patients as well as Good's patients; 2) Literature on IgG subclass deficiency is poor and immunological characteristics of these patients have not been explored as they have been in CVID. Indeed, IgG subclass deficiency may progress to have typical CVID [3]. Thus, the study of both CVID and IgG subclass deficiency in a combined analysis was a novel aspect of our

Table 5 Immunological markers association with CVID complications

	Unit	Odds ratio	95% confidence interval	p-value
B cells				
CD19+	cells/mm ³	0.65	[0.36; 1.15]	0.1373
CD27-IgD+	% CD19+	1.49	[0.89; 2.48]	0.1310
CD27-IgD-	% CD19+	0.54	[0.24; 1.23]	0.1401
CD27 + IgD+	% CD19+	0.62	[0.34; 1.12]	0.1124
CD27 + IgD-	% CD19+	0.93	[0.29; 0.86]	0.0119
T cells				
CD3+	cells/mm ³	0.68	[0.40; 1.17]	0.1646
CD4+	cells/mm ³	0.45	[0.26; 0.78]	0.0046
CD8+	cells/mm ³	1.04	[0.66; 1.62]	0.8818
CD4+/CD8+ ratio		0.99	[0.63; 1.56]	0.9649
CD8+ T cells				
CD45RA + CCR7+	% CD8+	0.69	[0.43; 1.13]	0.1385
CD45RA-CCR7+	% CD8+	0.75	[0.46; 1.25]	0.2710
CD45RA-CCR7-	% CD8+	1.38	[0.86; 2.20]	0.1800
CD45RA + CCR7-	% CD8+	0.77	[0.48; 1.23]	0.2666
CD8 + CD57+	% CD8+	1.30	[0.83; 2.06]	0.2567
Activated T cells				
CD3 + HLA-DR+	% CD3+	2.99	[1.56; 5.73]	0.0010
CD4 + HLA-DR+	% CD4+	2.70	[1.40; 5.23]	0.0032
CD8 + HLA-DR+	% CD8+	2.75	[1.54; 4.90]	0.0006
Regulatory T cells				
	cells/mm ³	0.41	[0.22; 0.78]	0.0065
Natural Killer cells				
	cells/mm ³	0.75	[0.46; 1.21]	0.2369
Dendritic cells				
Myeloid	/ml	0.65	[0.36; 1.17]	0.1532
Plasmacytoid	/ml	0.37	[0.18; 0.73]	0.0046
Gammadelta cells				
	cells/mm ³	1.30	[0.79; 2.13]	0.3028
Gammadelta 2 cells	% GD	0.52	[0.32; 0.85]	0.0087

Univariable logistic regression analysis (Odds ratio per 1 standard deviation). ALTADIH Cohort, 2007-2010.

work; 3) Good syndrome (thymoma with immunodeficiency) was listed as a predominantly antibody deficiency in the 2009 update of Primary Immunodeficiency by the International Union of Immunological Societies [1]. Based on our view to include eligible patients with symptomatic PHID seen in routine clinical practice and this pivotal article, we do not have any arguments to exclude it from eligibility criteria.

First, PCA gave some insight to the association between the 23 immunological markers measured. An increase in the proportion of HLA-DR + T-cell as an indication for activation, positively correlated to an increase in CD45RA + CCR7- terminal effector CD8+ T-cells and to the CD8 + CD57+ immunosenescence marker, defined one group of markers. In parallel, a defect in B-cell differentiation (increased CD19 + CD27-

IgD + naïve B cells and by symmetry, decreased CD19 + CD27 + IgD- switched memory B cells) might identify another distinct pathway. The PCA gave the opportunity to explore and to observe graphically the links between immunological markers and we made the choice to present the plane spanned by the two first principal components because they explained the most the data variability (43%). Next, we performed a hierarchical cluster analysis on the basis of the cell profile under investigation. As it was more relevant biologically, cluster analysis was performed with immunological markers rather than with principal components. Thus, the variance left on other principal components was not an obstacle for the cluster analysis.

The cluster analysis separated five major groups in our PHID population with distinct rates of complications.

Table 6 Reclassification of 55 CVID patients smB+/-, separated according to experience of complications

Model with EUROclass Predicted probabilities (%)	Model with EUROclass and CD8+ HLA-DR+		Total
	< 50%	≥ 50%	
Patients who experienced CVID complications			
< 50%	9	10	19
≥ 50%	1	5	6
Total	10	15	25
Patients who didn't experience CVID complications			
< 50%	22	5	27
≥ 50%	2	1	3
Total	24	6	30

ALTADIH Cohort, 2007-2010.

The agreements between the two models for classifying the patients are depicted in gray. Adding CD8 + HLA-DR + to EUROclass model leads to better determine CVID complications for 10 patients who experienced CVID complications (upward movement in category for patients who experienced CVID complications) and to better determine the absence of CVID complications for 2 patients who did not experience CVID complications (downward movement in category for patients who did not experience CVID complications). In parallel, it leads to worsened classification for 1 patient who experienced CVID complications (downward movement in category for patients who did experience CVID complications) and 5 patients who did not experience CVID complications (upward movement in category for patients who did not experience CVID complications).

The positive net reclassification improvement (0.26) revealed that CD8+ HLA-DR + offers improved prediction of CVID complications.

Particularly, patients in cluster 5 (13 with CVID, 1 with IgG SD and 2 with Good's syndrome), 63% of whom suffering PHID complications, were characterized by a severe loss of CD19 + CD27 + IgD- switched memory B-cells whereas patients in clusters 3 and 4, mainly with CVID (75% and 88% of whom having PHID complications, respectively) were characterized by increased HLA-DR + activation of T-cells. Regarding available data for activated T cells or B cell subsets for 12 normal healthy donors (Additional file 4: Table S2) and data from PHID patients in Table 3, activated T cells were higher in PHID patients from clusters 3 and 4, and switched memory B cells were particularly lower in PHID patients from clusters 4 and 5. Furthermore, the strong association we found between PHID complications and increased CD8+ HLA-DR + T-cells reinforced the importance of the link between T-cell activation and clinical complications. More importantly, in a substudy with CVID patients, the classification that added CD8+ HLA-DR + to EUROclass was superior to the EUROclass model in predicting CVID complications. The advantage of adding CD8+ HLA-DR + to EUROclass resulted in a substantial increase of sensitivity and some loss of specificity to determine CVID complications. As a consensus does not exist for the cut-off of probability, we explored a threshold of 50% (Additional file 7: Tables S3, Additional file 8: Table S4).

Our results on switched memory B-cell defect, defining PHID cluster 5 in particular, are consistent with previous findings on B-cell phenotype in PHID [4,5,9,10]. Our study also highlights the importance of considering T-cell activation [12,13] in PHID (clusters 3 and 4 with increased T-cell activation with the highest percentage of complications, comprising mainly CVID patients and

the three patients with Good's syndrome identified in our cohort as PHID types). Moreover, the occurrence of two complications considered in addition to those defined by H Chapel et al. [6] (ie: splenomegaly and granulomatous disease) were found with the highest frequency in the clusters 3 and 4 with increased T-cell activation, suggesting that the HLA-DR + activation marker could help to discriminate them. However, this cross-sectional analysis of data from the ALTADIH cohort does not allow us to determine whether increased T-cell activation is a causal factor of PHID progression or if it is induced by PHID complication. Only longitudinal analyses of activation markers would help in deciphering the role of T-cell activation, determining whether or not patients with normal phenotype will progress to an abnormal phenotype and will develop complications. Adding T-cell parameters for discriminating PHID patients is important to better dissect these heterogeneous syndromes. This has been previously suggested in CVID [31] by several authors especially the Rome group [11] (Giovannetti et al.) as well as the Czech group [32] (Vlková et al.) and the French DEFI group [7] (Mouillot et al.). These authors suggested CD4 naive T cell percentage as additional marker in CVID. Markers of CD4+ T-cell differentiation were not performed in our study. At the time of research protocol writing, set up in 2005, the importance of CD4+ T-cell differentiation has not been reported in PHID literature as it would be few years later. Thus, markers of CD4+ T-cell differentiation have not been planned in the ALTADIH protocol and the prospective design of the study did not allow for this B-cell marker to be assessed afterwards. This is a clear limitation of the study. However, previous authors also highlighted that a decrease of naive CD4 cells was also

associated to T-cell activation and it would not be unreasonable to consider T-cell activation as an indirect marker of decreased naive CD4 cells in our study. In addition, the signification of the memory status in the CD4+ compartment has been a subject of debates until recently as exemplified by Su LE, et al. [33].

Conclusion

To date, PHID diseases have only been examined through the analysis of separated compartments of the immune system. On this ground, our approach of combining a wide set of immunological markers using appropriate statistical methods for the high-dimensional is a novel aspect in PHID. The meaning of the identified clusters would need to be validated with a larger number of patients, for example the French national prospective Cohort DEFI [7]. It would be the opportunity to include markers of CD4+ T-cell differentiation in the analyses and consider potential other markers, as mucosal associated invariant T cells [34,35], recently reported in CVID [36].

In conclusion, by combining a wide set of immunological markers, this study reveals that T-cell activation could be of additional value to discriminate patients with specific PHID features. Such information might have implications for diagnosis and clinical management. A systematic recording of the HLA-DR + phenotype in every patient with PHID seems relevant.

Additional files

Additional file 1: Table S1. References of the panels of monoclonal antibodies used for identifying cells by flow cytometry using a FC500 flow cytometer from Beckman-Coulter.

Additional file 2: Figure S1. Gating strategies for flow cytometric assessment of activated T-lymphocytes and other major lymphocyte sub-populations in the blood on a CVID patient.

Additional file 3: Figure S2. Immunological markers of 12 healthy donors, plotted on the first two principal components.

Additional file 4: Table S2. Immunological marker values of normal patients.

Additional file 5: Figure S3. Thirty-eight immunological markers involved in PHID on the two first principal components. ALTADIH Cohort, 2007-2010.

Additional file 6: Figure S4. Classification of 55/57 CVID patients by hierarchical cluster analysis, according to 23 immunological markers involved in the disease. ALTADIH Cohort, 2007-2010.

Additional file 7: Table S3. Sensibility [95% confidence interval] and specificity [95% confidence interval] of the model EUROclass to determine CVID complications with a probability $\geq 50\%$ or $< 50\%$.

Additional file 8: Table S4. Sensibility [95% confidence interval] and specificity [95% confidence interval] of the model EUROclass and CD8+HLA-DR+ to determine CVID complications with a probability $\geq 50\%$ or $< 50\%$.

Abbreviations

CVID: Common Variable ImmunoDeficiency; GD: GammaDelta T cells; Ig: Immunoglobulin; IgG SD: IgG SubClass ImmunoDeficiency; IQR: InterQuartile Range; mab: Monoclonal antibody; mDC: Myeloid dendritic

cells; NK: Natural Killer cells; OR: Odds Ratio; PCA: Principal Component Analysis; PC: Principal Component; PC1: First Principal Component; PC2: Second Principal Component; pDC: Plasmacytoid Dendritic Cells; PHID: Primary Humoral Immunodeficiency Diseases; SD: Standard Deviation; smB: Switched memory B cells; T reg: Regulatory T cells.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DA, XD, A-LF, FL, PR, and J-FV enrolled PHID patients in the cohort and provided clinical data. J-FM performed the flowcytometric analysis. M-QP was involved in data collection. M-QP and RT participated in the study design and performed the statistical analysis of the data. M-QP wrote the manuscript. J-FM, RT, LW and J-FV critically read the manuscript. All authors read and approved the final manuscript.

ALTADIH study group members

Scientific committee: Dr F. Lifermann, Pr J-F. Moreau, R. Thiébaud, Pr J-F. Viallard Participating Hospital Departments (participating physicians): Bordeaux University Hospital (Pr J-F. Viallard); Agen Hospital (Dr P. Rispal); Dax Hospital (Dr F. Lifermann); Limoges Hospital (Dr A-L. Faucher); Pau Hospital (Dr X. Delbrel); Toulouse Hospital (Pr D. Adoue); Bordeaux Immunology Department: Pr P. Blanco, J-C. Carron, K. Eschechel, M. Garcia, Pr J-F. Moreau, Dr I. Pellegrin
Methodology: R. Thiébaud, M-Q. Picat, L. Wittkop
Data collection: D. Jacquemart, M-Q. Picat
Computing and data management: G. Dupouy, C. Maldonado, I. Perrot
Statistical analysis: M-Q. Picat, R. Thiébaud

Acknowledgments

The authors thank all patients for their participation. The authors thank Myriam Garcia, Kamal Eschechel, Jean-Claude Carron and Céline Cognet for the skillful collaboration in carrying out the immunophenotyping. They are grateful to Joanna Lewis, PhD in Biostatistics (University College London, United Kingdom), for expert reviewing of the English. This study was supported by a grant from the French Ministry of Health (French inter-regional Clinical Research Hospital Program: PHRC 2006-A00576-45). Marie-Quitterie Picat was supported by funding from the French Foundation for Medical Research (FRM).

Author details

¹INSERM, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, Bordeaux F-33076, France. ²University of Bordeaux, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, Bordeaux F-33076, France. ³Department of Medical Information, Bordeaux University Hospital, Bordeaux F-33000, France. ⁴INRIA SISTM, F-33405 Talence, France. ⁵Department of Internal Medicine, Dax Hospital, Dax F-40107, France. ⁶Department of Internal Medicine, Pau Hospital, Pau F-64046, France. ⁷Department of Internal Medicine, Purpan Hospital, Toulouse F-31059, France. ⁸Department of Internal Medicine, Dupuytren Hospital, Limoges F-87042, France. ⁹Department of Internal Medicine, Agen Hospital, Agen F-47923, France. ¹⁰University of Bordeaux, Bordeaux F-33076, France. ¹¹CNRS, UMR 5164, Bordeaux F-33076, France. ¹²Bordeaux University Hospital, Laboratory of Immunology, Bordeaux F-33076, France. ¹³Department of Internal Medicine and Infectious Diseases, Bordeaux University Hospital, Pessac F-33604, France. ¹⁴Hôpital Haut-Lévêque - Service de médecine interne, 5, avenue de Magellan, Pessac 33604, France.

Received: 31 May 2013 Accepted: 28 February 2014

Published: 12 March 2014

References

1. Al-Herz W, Bousfiha A, Casanova JL, Chapel H, Conley ME, Cunningham-Rundles C, Etzioni A, Fischer A, Franco JL, Geha RS, Hammarström L, Nonoyama S, Notarangelo LD, Ochs HD, Puck JM, Roifman CM, Seger R, Tang ML: Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol* 2011, **2**:54.
2. Oksenhendler E, Gerard L, Fieschi C, Malphettes M, Mouillot G, Jaussaud R, Viallard JF, Gardembas M, Galicier L, Schleinitz N, Suarez F, Soulas-Sprauel P,

- Hachulla E, Jaccard A, Gardeur A, Théodorou I, Rabian C, Debré P, Viallard JF, Gardembas M, Galicier L, Schleinitz N, Suarez F, Soulas-Sprauel P, Hachulla E, Jaccard A, Gardeur A, Théodorou I, Rabian C, DEFI Study Group, et al: **Infections in 252 patients with common variable immunodeficiency.** *Clin Infect Dis* 2008, **46**(10):1547–1554.
3. Preud'homme JL, Hanson LA: **IgG subclass deficiency.** *Immunodef Rev* 1990, **2**(2):129–149.
 4. Kelesidis T, Yang O: **Good's syndrome remains a mystery after 55 years: a systematic review of the scientific evidence.** *Clin Immunol* 2010, **135**:347–363.
 5. Wehr C, Kivioja T, Schmitt C, Ferry B, Witte T, Eren E, Vlkova M, Hernandez M, Detkova D, Bos PR, Poerksen G, von Bernuth H, Baumann U, Gutenberger S, Schlesier M, Bergeron-van der Cruyssen F, Le Garff M, Debré P, Jacobs R, Jones J, Bateman E, Litzman J, van Hagen PM, Plebani A, Schmidt RE, Thon V, Quinti I, Espanol T, Webster AD, Chapel H, et al: **The EUROclass trial: defining subgroups in common variable immunodeficiency.** *Blood* 2008, **111**(1):77–85.
 6. Chapel H, Lucas M, Lee M, Bjorkander J, Webster D, Grimbacher B, Fieschi C, Thon V, Abedi MR, Hammarstrom L: **Common variable immunodeficiency disorders: division into distinct clinical phenotypes.** *Blood* 2008, **112**(2):277–286.
 7. Mouillot G, Carmagnat M, Gerard L, Garnier JL, Fieschi C, Vince N, Karlin L, Viallard JF, Jaussaud R, Boileau J, Donadieu J, Gardembas M, Schleinitz N, Suarez F, Hachulla E, Delavigne K, Morisset M, Jacquot S, Just N, Galicier L, Charron D, Debré P, Oksenhendler E, Rabian C, DEFI Study Group: **B-cell and T-cell phenotypes in CVID patients correlate with the clinical phenotype of the disease.** *J Clin Immunol* 2010, **30**(5):746–755.
 8. Cunningham-Rundles C: **How I treat common variable immune deficiency.** *Blood* 2010, **116**(1):7–15.
 9. Piqueras B, Lavenu-Bombled C, Bergeron-van der Cruyssen F, Mouthon L, Chevret S, Debré P, Schmitt C, Oksenhendler E: **Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects.** *J Clin Immunol* 2003, **23**(5):385–400.
 10. Warnatz K, Denz A, Drager R, Braun M, Groth C, Wolff-Vorbeck G, Eibel H, Schlesier M, Peter HH: **Severe deficiency of switched memory B cells (CD27+)IgM(-)IgD(-) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease.** *Blood* 2002, **99**(5):1544–1551.
 11. Giovannetti A, Pierdominici M, Mazzetta F, Marziali M, Renzi C, Mileo AM, De Felice M, Mora B, Esposito A, Carello R, Pizzuti A, Paggi MG, Paganelli R, Malorni W, Aiuti F: **Unravelling the complexity of T cell abnormalities in common variable immunodeficiency.** *J Immunol* 2007, **178**(6):3932–3943.
 12. Viallard JF, Blanco P, Andre M, Etienne G, Liferman F, Neau D, Vidal E, Moreau JF, Pellegrin JL: **CD8 + HLA-DR + T lymphocytes are increased in common variable immunodeficiency patients with impaired memory B-cell differentiation.** *Clin Immunol* 2006, **119**(1):51–58.
 13. Rezaei N, Wing JB, Aghamohammadi A, Carlring J, Lees A, Asgarian-Omran H, Pourpak Z, Sarrafnejad A, Kardar GA, Shahrestani T, Masoumi F, Zare A, Saghaei S, Sarrafzadeh S, Foster RA, Heath AW, Read RC: **B-cell-T-cell activation and interaction in common variable immunodeficiency.** *Hum Immunol* 2010, **71**(4):355–362.
 14. Fevang B, Yndestad A, Sandberg WJ, Holm AM, Muller F, Aukrust P, Frøland SS: **Low numbers of regulatory T cells in common variable immunodeficiency: association with chronic inflammation in vivo.** *Clin Exp Immunol* 2007, **147**(3):521–525.
 15. Horn J, Manguiat A, Berglund LJ, Knerr V, Tahami F, Grimbacher B, Fulcher DA: **Decrease in phenotypic regulatory T cells in subsets of patients with common variable immunodeficiency.** *Clin Exp Immunol* 2009, **156**(3):446–454.
 16. Viallard JF, Bloch-Michel C, Caubet O, Parrons M, Texier-Maugein J, Neau-Cransac M, Taupin JL, Moreau JF, Pellegrin JL: **Gammadelta T lymphocytosis associated with granulomatous disease in a patient with common variable immunodeficiency.** *Clin Infect Dis* 2002, **35**(12):134–137.
 17. Fulcher DA, Avery DT, Fewings NL: **Invariant natural killer (iNK) T cell deficiency in patients with common variable immunodeficiency.** *Clin Exp Immunol* 2009, **157**(3):365–369.
 18. Scott-Taylor TH, Green MR, Raieszadeh M, Workman S, Webster AD: **Defective maturation of dendritic cells in common variable immunodeficiency.** *Clin Exp Immunol* 2006, **145**(3):420–427.
 19. Garg V, Lipka S, Rizvon K, Singh J, Rashid S, Mustacchia P: **Diffuse nodular lymphoid hyperplasia of intestine in selective IgG 2 subclass deficiency, autoimmune thyroiditis, and autoimmune hemolytic anemia: case report and literature review.** *J Gastrointestinal Liver Dis* 2012, **21**(4):431–434.
 20. Park MA, Li JT, Hagan JB, Maddox DE, Abraham RS: **Common variable immunodeficiency: a new look at an old disease.** *Lancet* 2008, **372**(9637):489–502.
 21. Aghamohammadi A, Parvaneh N, Rezaei N: **Common variable immunodeficiency: a heterogeneous group needs further subclassification.** *Expert Rev Clin Immunol* 2009, **5**(6):629–631.
 22. O'Gorman MR, Zollett J, Bensen N: **Flow cytometry assays in primary immunodeficiency diseases.** *Methods Mol Biol* 2011, **699**:317–335.
 23. Lugli E, Roederer M, Cossarizza A: **Data analysis in flow cytometry: the future just started.** *Cytometry A* 2010, **77**(7):705–713.
 24. Lugli E, Pinti M, Nasi M, Troiano L, Ferraresi R, Mussi C, Salvioli G, Pateskin V, Robinson JP, Durante C, Cocchi M, Cossarizza A: **Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data.** *Cytometry A* 2007, **71**(5):334–344.
 25. Ringner M: **What is principal component analysis?** *Nat Biotechnol* 2008, **26**(3):303–304.
 26. Everitt B: **Cluster analysis is a generic term for a wide range of numerical methods for examining data.** *Stat Methods Med Res* 2004, **13**(5):343–345.
 27. Ward J: **Hierarchical grouping to optimize an objective function.** *J Am Stat Assoc* 1963, **58**(301):236–244.
 28. Giovannetti A, Pierdominici M, Aiuti F: **T-cell homeostasis: the dark(ened) side of common variable immunodeficiency.** *Blood* 2008, **112**(2):446.
 29. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS: **Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond.** *Stat Med* 2008, **27**(2):157–172.
 30. Focosi D, Bestagno M, Burrone O, Petrini M: **CD57+ T lymphocytes and functional immune deficiency.** *J Leukoc Biol* 2010, **87**(1):107–116.
 31. Yong PF, Thaventhiran JE, Grimbacher B: **"A rose is a rose is a rose," but CVID is not CVID common variable immune deficiency (CVID), what do we know in 2011?** *Adv Immunol* 2011, **111**:47–107.
 32. Vlková M, Thon V, Sárkyová M, Bláha L, Svobodník A, Lokaj J, Litzman J: **Age dependency and mutual relations in T and B lymphocyte abnormalities in common variable immunodeficiency patients.** *Clin Exp Immunol* 2006, **143**(2):373–379.
 33. Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM: **Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults.** *Immunity* 2013, **38**(2):373–383.
 34. Guerri L, Dusseaux M, Martin E, Soudais C, Lantz O: **Mucosal-associated invariant T cells: unconventional development and function.** *Trends Immunol* 2011, **32**(5):212–218.
 35. Young MH, U'ren L, Huang S, Malleveay T, Scott-Browne J, Crawford F, Lantz O, Hansen TH, Kappler J, Marrack P, Gapin L: **MAIT cell recognition of MR1 on bacterially infected and uninfected cells.** *PLoS One* 2013, **8**(1):e53789.
 36. Litzman J, Nechvatalova J, Xu J, Ticha O, Vlkova M, Hel Z: **Chronic immune activation in common variable immunodeficiency (CVID) is associated with elevated serum levels of soluble CD14 and CD25 but not endotoxaemia.** *Clin Exp Immunol* 2012, **170**:321–332.

doi:10.1186/1471-2172-15-13

Cite this article as: Picat et al.: T-cell activation discriminates subclasses of symptomatic primary humoral immunodeficiency diseases in adults. *BMC Immunology* 2014 **15**:13.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

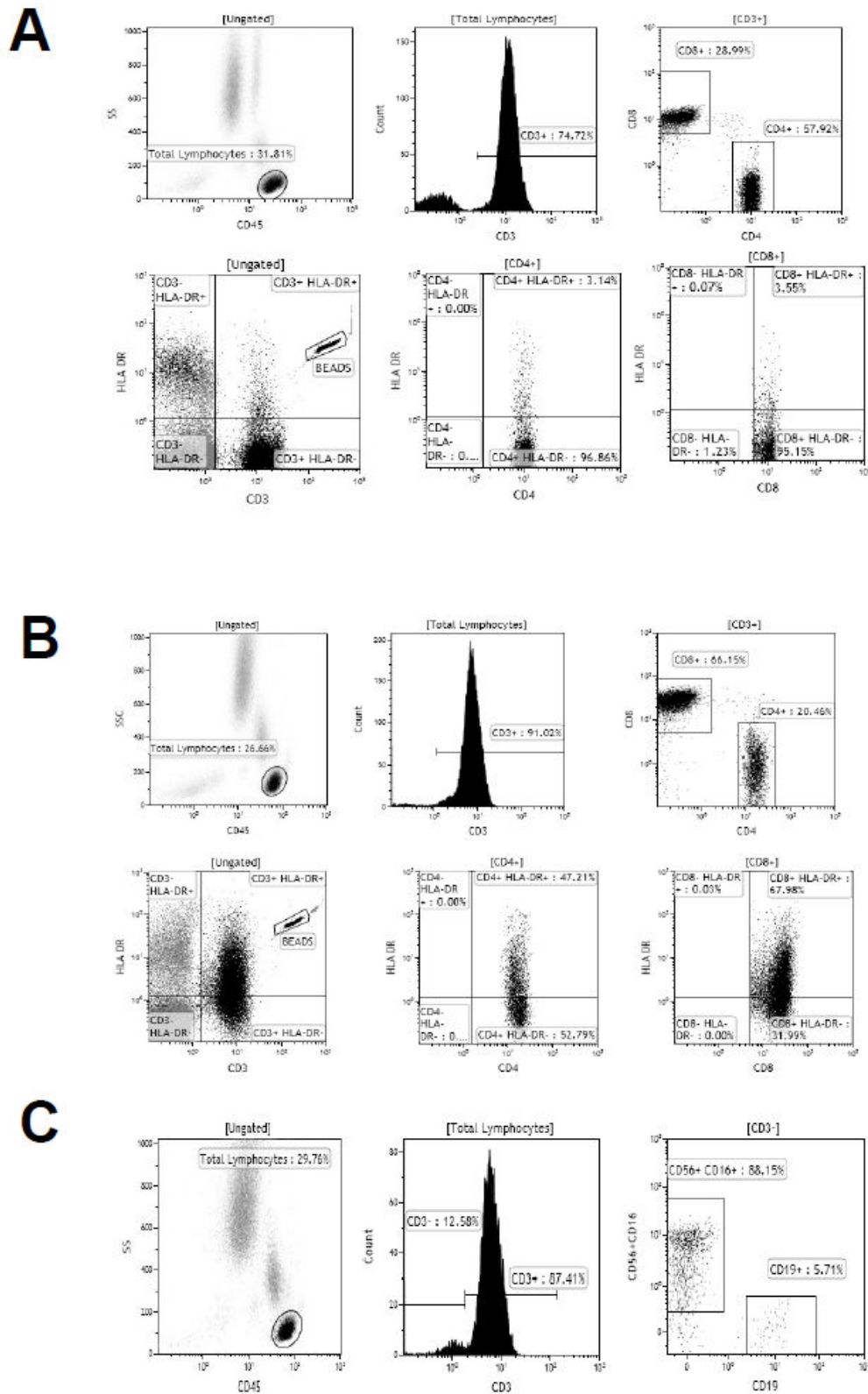
Submit your manuscript at
www.biomedcentral.com/submit



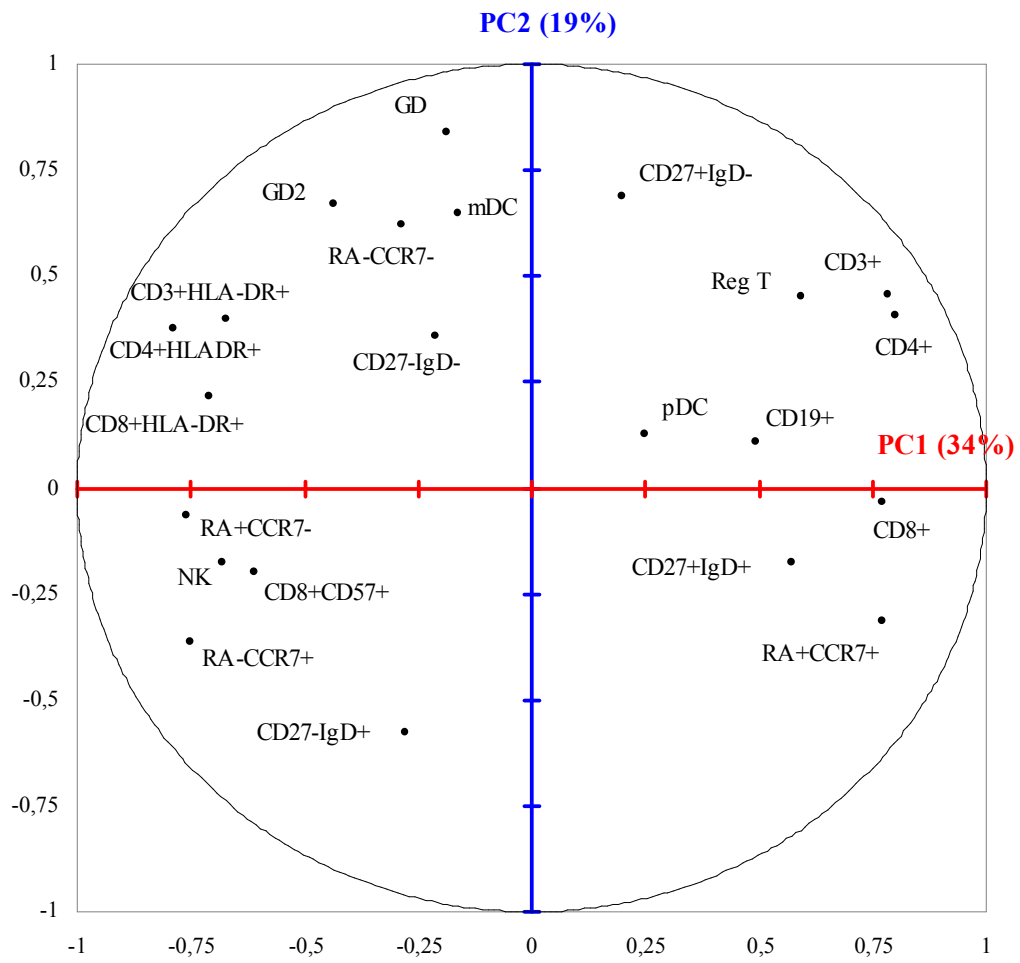
Additional file 1:Table S1. References of the panels of monoclonal antibodies used for identifying cells by flow cytometry using a FC500 flow cytometer from Beckman-Coulter.

Cells labelling*	Monoclonal antibodies	References
B-cell differentiation	anti-CD19	A07766
	anti-CD27	6607107
	anti-CD21	PN IMU473U
	anti-sIgD	736000
T-cell differentiation	anti-CD8	A07756
	anti-CD45RA	PN IM 271U
	anti-CCR7 from R&D (Minneapolis, MN, USA)	FAB197A
HLA-DR T-lymphocyte activation markers	TetraChrome (anti-CD45, -3, -4 and -8)	6607013
	anti-HLA-DR	PNA40579
CD25 T-lymphocyte activation marker	anti-CD3, -4 and -8	6607013
	anti-CD25	IM2646
CD28 T-lymphocyte activation marker	anti-CD3, -4 and -8	6607013
	anti-CD28	6607108
CD38 T-lymphocyte activation marker	anti-CD3, -4 and -8	6607013
	anti-CD38	A07780
Regulatory T-cells	anti-CD4	A07750
	anti-CD25	IM2646
	anti-CD127	PN IM1980U
NK/B cells	anti-CD45, -56, -19 and -3	6607073
	anti-CD16	A07766
Myeloid dendritic cells	Beckman Coulter kit	A23413
Plasmacytoid dendritic cells	Beckman Coulter kit	A23416
T cells expressing the gamma/delta TCR	anti-Vdelta2	PN IM1464
	anti-pan-delta	PN IM1418U

**T and B-cell subpopulation counts were obtained using the flow count beads kit from Beckman-Coulter Flow Count (ref: 7547053) following a lyse (Versalyse ref: A09777) and no wash procedure according to the manufacturer's recommendations.*



Additional file 2: Figure S1. Gating strategies for flow cytometric assessment of activated T-lymphocytes and other major lymphocyte sub-populations in the blood (in Panel A, applied to a normal healthy donor whereas, in Panel B, applied to a COVID patient. Panel C depicts the B/NK cells phenotyping of the COVID patient).



Additional file 3: Figure S2. Immunological markers of 12 healthy donors, plotted on the first two principal components.

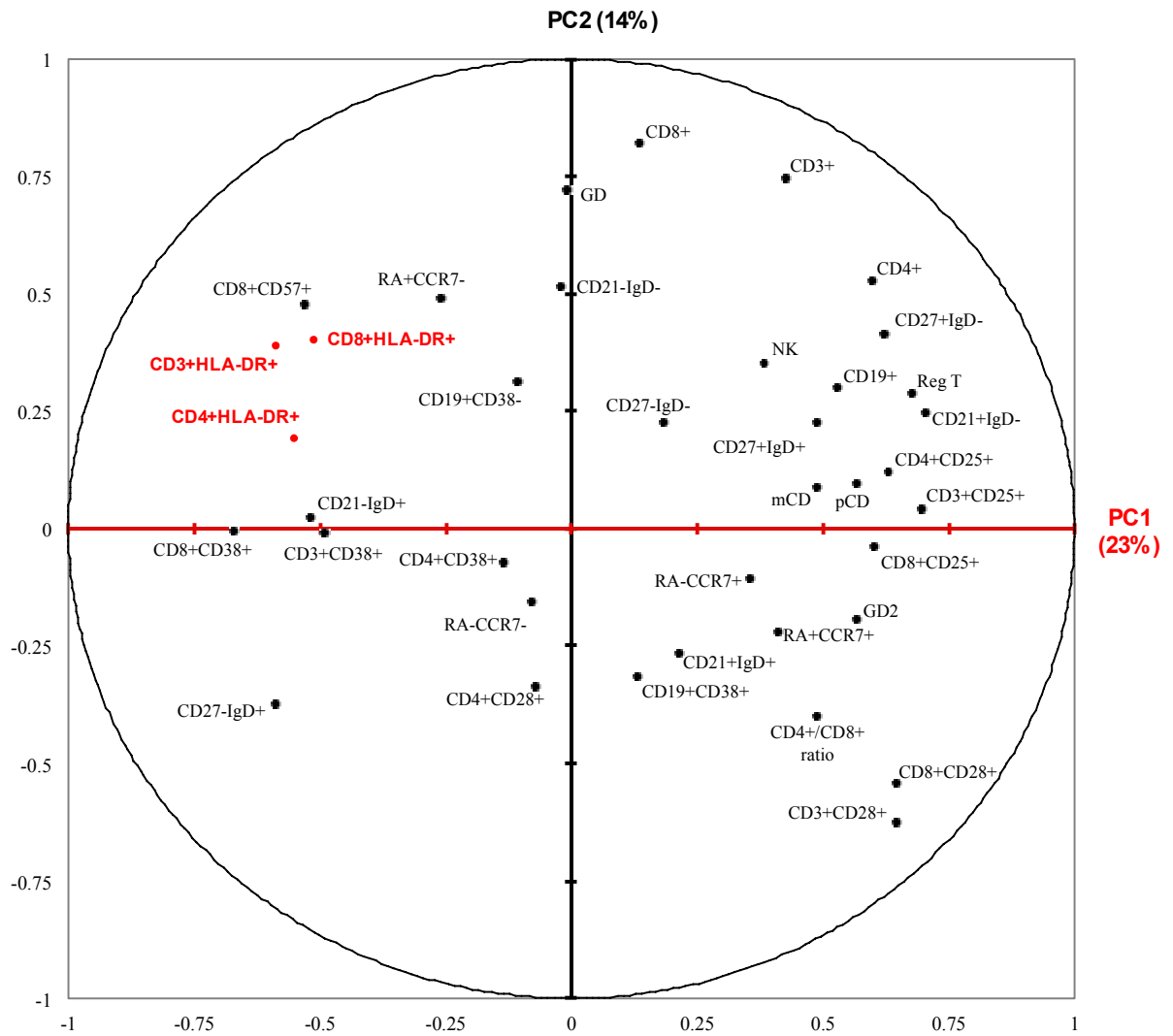
Additional file 4: Table S2. Immunological marker values of normal patients.

		Unit	Records	Median	IQR
Total circulating lymphocytes (TCL)		cells/mm ³	101*	2100	(1600, 2400)
B cells	CD19+	cells/mm ³	101*	300	(200, 400)
		% TCL	101*	32	(28, 39)
Naïve	CD27-IgD+	%	12 [†]	67.45	(55.49,
		CD19+			74.58)
Switched	CD27-IgD-	%	12 [†]	2.75	(1.82, 6.48)
		CD19+			
Marginal zone	CD27+IgD+	%	12 [†]	9.25	(6.07, 15.45)
		CD19+			
Switched memory	CD27+IgD-	%	12 [†]	16.04	(11.88,
		CD19+			20.01)
T cells	CD3+	cells/mm ³	101*	1400	(1100, 1700)
	CD4+	cells/mm ³	101*	800	(700, 1100)
	CD8+	cells/mm ³	101*	700	(500, 900)
	CD4+/CD8+ ratio		101*	1.2	(1.0, 1.5)
Naïve	CD45RA+CCR7+	% CD8+	8 [†]	55.64	(44.57,
					62.19)
Central memory	CD45RA-CCR7+	% CD8+	8 [†]	0.86	(0.57, 1.32)
Effector memory	CD45RA-CCR7-	% CD8+	8 [†]	23.13	(17.89,
					32.09)
Terminal effector	CD45RA+CCR7-	% CD8+	8 [†]	31.87	(13.97,
Immunosenescent	CD8+CD57+	% CD8+	8 [†]	1.11	(0.92, 4.20)
Activated T cells					
HLA-DR	CD3+HLA-DR+	% CD3+	12 [†]	6,36	(5,04, 8.29)
	CD4+HLA-DR+	% CD4+	12 [†]	6,63	(4.25, 7.60)
	CD8+HLA-DR+	% CD8+	12 [†]	14,68	(10.12,
					19.68)
Regulatory T cells	CD4+CD25+CD127-	cells/mm ³	12 [†]	54	(39, 79)
Natural Killer cells	CD3-CD16+CD56+	cells/mm ³		225	(158, 294)

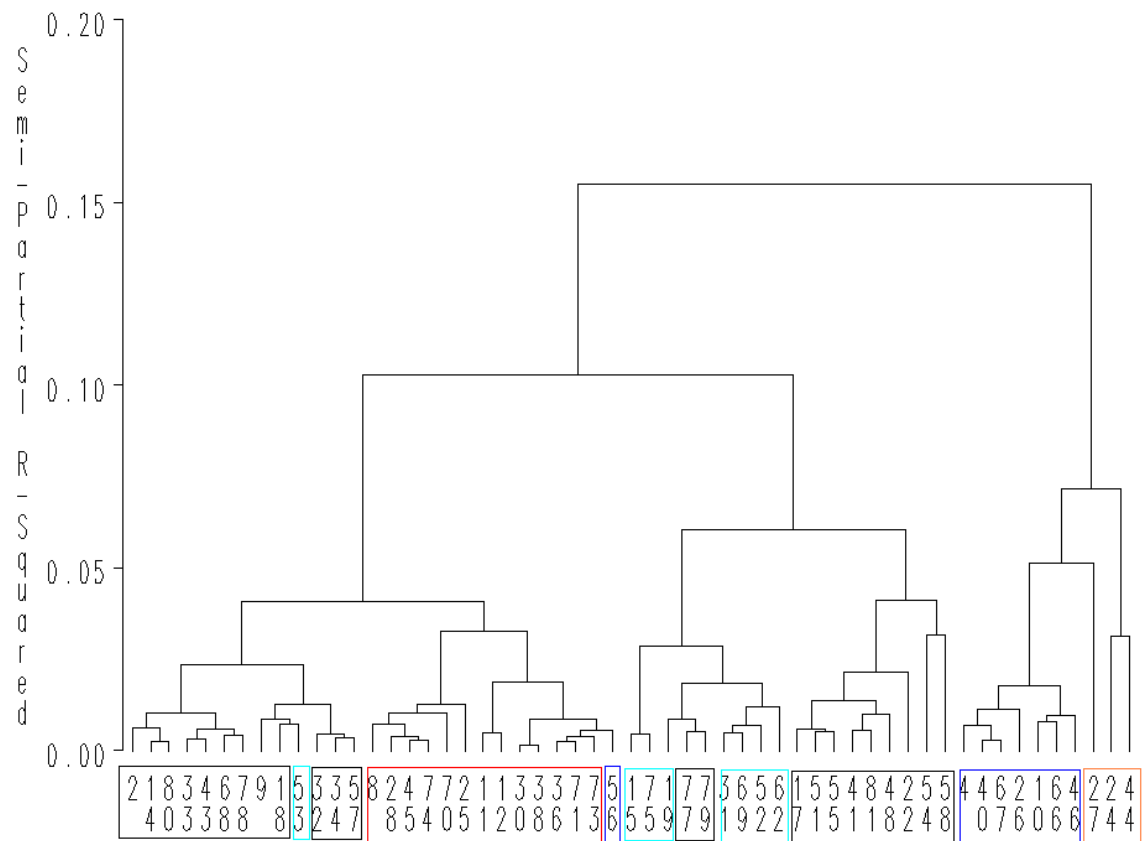
Dendritic cells					
Myeloid	mDC	/ml	12 [†]	14547	(8544, 20648)
Plasmacytoid	pDC	/ml	12 [†]	7233	(5551, 9617)
Gammadelta cells		cells/mm ³	8 [†]	52	(43, 94)
Gammadelta 2 cells		% GD	8 [†]	1.90	(1.43, 4.13)

*Data from 101 normal adults (45 male, 56 female) aged 18 to 70 years, with a median age of 40 years. Erkeller-Yuksel FM et al. J Pediatr 1992 Feb;120:216-22.

[†]Data available from 12 normal adults (4 male, 8 female) aged 23 to 56 years, with a median age of 31 years (missing values for some markers).



Additional file 5: Figure S3. Thirty-eight immunological markers involved in PHID on the two first principal components. ALTADIH Cohort, 2007-2010.



Additional file 6: Figure S4. Classification of 55/57 CVID patients by hierarchical cluster analysis, according to 23 immunological markers involved in the disease. ALTADIH Cohort, 2007-2010.

Additional file 7: Table S3. Sensibility [95% confidence interval] and specificity [95% confidence interval] of the model EUROclass to determine CVID complications with a probability $\geq 50\%$ or $< 50\%$.

Probability of CVID complications predicted by model with EUROclass	Complication = yes	Complication = no	TOTAL
$\geq 50\%$	6	3	9
$< 50\%$	19	27	46
TOTAL	25	30	55

Sensitivity = 24.0 % [12.7; 35.3]

Specificity = 90.0 % [82.1; 97.9]

Additional file 8: Table S4. Sensibility [95% confidence interval] and specificity [95% confidence interval] of the model EUROclass and CD8+HLA-DR+ to determine CVID complications with a probability $\geq 50\%$ or $< 50\%$.

Probability of CVID complications predicted by model with EUROclass and CD8+ HLA-DR+	Complication = yes	Complication = no	TOTAL
$\geq 50\%$	15	6	21
$< 50\%$	10	24	34
TOTAL	25	30	55

Sensitivity = 60.0 % [47.1; 72.9]
 Specificity = 80.0 % [67.1; 92.9]

3.5. Discussion

Dans le cadre de ce travail, nous avons identifié cinq groupes de patients, parmi lesquels le pourcentage de complications variait considérablement. La combinaison de l'ensemble des marqueurs impliqués dans les déficits immunitaires humoraux primitifs symptomatiques de l'adulte a rapporté l'existence de deux mécanismes distincts associés à des complications non infectieuses. La diminution des lymphocytes B et l'augmentation des lymphocytes T activés HLA-DR+ étaient les anomalies les plus importantes observées dans les complications non infectieuses. En outre, dans un sous-groupe de 57 patients atteints de déficit immunitaire commun variable, la classification ajoutant le marqueur d'activation des lymphocytes T CD8+HLA-DR+ à la classification consensuelle EUROclass, basée uniquement sur les anomalies des lymphocytes B, permettait de mieux prédire les complications non infectieuses que la classification EUROclass seule. En plus des données connues sur l'implication du déficit en lymphocytes B dans les déficits immunitaires humoraux primitifs symptomatiques de l'adulte, nos résultats ont souligné l'importance de l'activation des lymphocytes T, en particulier celle du marqueur CD8+HLA-DR+ qui pourrait permettre d'améliorer la discrimination des patients et aider à identifier des groupes de patients avec des profils cliniques distincts.

Plusieurs points méritent d'être soulignés quant à la réflexion menée autour de l'application de l'analyse en composantes principales et de la classification hiérarchique ascendante autour de l'exemple concret des déficits immunitaires humoraux primitifs symptomatiques de l'adulte. Tout d'abord, pour prendre en compte des unités de mesure différentes (% , cellules/mm³ ou /mL), les données ont été centrées réduites. Puis, l'analyse en composantes principales a permis de décrire de manière graphique et approfondie les relations étroites en variables. Elle a notamment permis de mettre en évidence la forte corrélation existant entre les marqueurs d'activation HLA-DR+ des lymphocytes T CD3+, CD4+ et CD8+. Un autre point à souligner est notre choix de réaliser la classification hiérarchique ascendante à partir des marqueurs immunologiques et non à partir des composantes principales. Après réduction de dimension de données, les premières composantes principales résumant le mieux l'information peuvent être utilisées dans des modèles classiques de régression ou pour des analyses autres [Song et al., 2009]. Notre choix de réaliser la classification hiérarchique ascendante à partir des marqueurs a été guidé par une meilleure pertinence clinique de l'interprétation des clusters à partir des marqueurs immunologiques. Au terme de la

classification hiérarchique ascendante, les clusters présentant le plus haut pourcentage de complications étaient ceux présentant le plus haut pourcentage d'activation lymphocytaire T et un déficit en lymphocytes B. Suite à des discussions étroites avec les cliniciens, nous avons considéré une définition plus large des complications non infectieuses par rapport aux phénotypes référencés par Chapel H et al (auto-immunité, hyperplasie lymphoïde, entéropathies chroniques) [Chapel et al., 2008], prenant en compte en plus la splénomégalie et les granulomatoses dans la définition des complications non infectieuses pour notre étude [Viillard et al., 2006]. En particulier, les patients atteints de ces deux complications étaient regroupés dans les clusters présentant le plus haut pourcentage d'activation lymphocytaire T, suggérant ainsi que l'activation lymphocytaire T pourrait permettre d'améliorer la discrimination de ces patients.

L'étude préalable exploratoire approfondie des marqueurs ayant permis de mettre en évidence, en plus des anomalies connues sur les lymphocytes B, l'importance de l'augmentation des marqueurs d'activation lymphocytaire T HLA-DR+, nous l'avons complétée par une étude des marqueurs associés à la survenue de complications, via une méthode classique de régression logistique. Cette méthode a confirmé l'association la plus marquée de l'augmentation du marqueur d'activation des lymphocytes T CD8+HLA-DR+ avec les complications non infectieuses et l'association des anomalies connues sur le déficit en lymphocytes B. Enfin, la classification la plus consensuelle existant chez les patients atteints de déficit immunitaire commun variable (EUROclass) [Wehr et al., 2008], l'ensemble de ces résultats nous a conduit à évaluer, chez ces patients, l'apport du marqueur d'activation CD8+HLA-DR+ à cette classification, via la méthode d'index de reclassification proposée par Pencina MJ et al [Pencina et al., 2008] (voir annexe 4 et tableau 6 de l'article section 3.4.1). La classification ajoutant ce marqueur à la classification consensuelle EUROclass, basée uniquement sur les lymphocytes B, permettait de mieux prédire les complications (aire sous la courbe = 0,57 sans le marqueur CD8+HLA-DR+ contre 0,77 avec ce marqueur). Ajouter ce marqueur à la classification EUROclass permettait notamment de mieux reclasser 10 patients ayant eu des complications non infectieuses par rapport au modèle EUROclass seul (patients classés avec une probabilité de complications < 50 % avec le modèle EUROclass seul alors que classés avec une probabilité \geq 50 % avec le modèle EUROclass et CD8+HLA-DR+).

Nos résultats nécessitent d'être validés à plus large échelle avec, dans l'application des méthodes proposées, l'inclusion des marqueurs de différenciation des lymphocytes T CD4+ manquants dans notre étude et d'autres marqueurs plus récemment rapportés dans les déficits immunitaires humoraux primitifs de l'adulte [Litzman et al., 2012]. A ce titre, il existe en

France une large cohorte, nommée DEFI, de patients atteints de déficits immunitaires primitifs [Oksenhendler et al., 2008]. De plus, l'analyse transversale des données d'inclusion de notre cohorte ne nous permettait pas d'établir de relation causale entre l'augmentation des lymphocytes T activés dans le sang des patients atteints de déficits immunitaires primitifs symptomatiques et la survenue de complications.

L'analyse descriptive des premières données longitudinales disponibles de la cohorte ALTADIH a donné lieu à un stage de Master 1 Santé Publique dont l'objectif était de décrire l'évolution des marqueurs d'activation lymphocytaire T (durée de 2 mois, co-encadrement : Pr Rodolphe Thiébaud, Dr Marie-Quitterie Picat). Cette analyse a rapporté une stabilité de l'activation des lymphocytes T dans le temps et sous traitement, témoignant de l'absence d'effet des immunoglobulines substitutives sur l'activation. De plus, dans les clusters 1 et 2 de notre étude qui présentaient le plus faible pourcentage d'activation, l'activation est restée stable au cours du temps et aucune complication non infectieuse n'est survenue pendant la période de suivi, permettant d'émettre l'hypothèse potentielle que l'augmentation de l'activation lymphocytaire T pourrait être plus une cause qu'une conséquence de la survenue des complications.

3.6. Conclusion et perspectives

L'analyse transversale des données d'inclusion de notre cohorte par des méthodes exploratoires multidimensionnelles nous a permis de souligner l'importance de l'activation immunitaire chronique au cours des déficits immunitaires humoraux primitifs symptomatiques de l'adulte. La réflexion menée autour de cet exemple concret montre la complémentarité des méthodes non supervisées pour apporter des informations pertinentes pour guider la mise en œuvre de méthodes supervisées, comme les méthodes classiques de régression [Aguilera et al., 2006 ; Lugli et al., 2007], ou d'autres méthodes, comme la méthode d'index de reclassification utilisée dans cette étude. Ce travail montre également l'intérêt, dans le contexte de pathologies rares avec peu de patients, de la nécessité d'employer des méthodes utilisant l'ensemble des mesures disponibles pour extraire des informations immunologiquement pertinentes (identification spécifique de l'importance du marqueur d'activation lymphocytaire T CD8+HLA-DR+ parmi les 23 marqueurs étudiés).

De plus, la littérature rapportée dans les déficits immunitaires humoraux primitifs symptomatiques de l'adulte impliquant l'ensemble des acteurs de l'immunité, cet exemple concret illustre l'apport des méthodes intégratives pour étudier le système immunitaire dans

sa globalité dans les pathologies où les acteurs de l'immunité innée et de l'immunité adaptative sont mis en jeu.

4. DETERMINANTS DE L'ACTIVATION IMMUNITAIRE CHRONIQUE

CHEZ LES PATIENTS ADULTES ATTEINTS DE VIH

TRAITES EFFICACEMENT

4.1. Problématique

Grâce à l'efficacité et la généralisation des thérapies antirétrovirales hautement actives, l'incidence des cas de SIDA (Syndrome de l'ImmunoDéficiency Acquisée) et la mortalité des personnes infectées par le VIH ont beaucoup diminué, dans les pays où l'accès aux soins est généralisé. Néanmoins, une des caractéristiques de l'infection à VIH chez les patients traités avec une charge virale indétectable est un état d'activation du système immunitaire qui persiste, malgré le traitement. Cette activation immunitaire chronique sur le long terme est un élément clé de la physiopathologie de l'infection par le VIH et un puissant facteur prédictif de progression de la maladie, indépendamment de la charge virale [Deeks, 2001 ; Sooda et Silvestri, 2008]. Elle est responsable non seulement des complications liées au VIH mais également des comorbidités non liées au VIH [Hsue et al., 2012 ; Plaeger et al., 2012 ; Tebas et al., 2008]. Ainsi, les défis auxquels sont confrontés les praticiens pour ces patients se sont étendus de la prise en charge des événements cliniques liés au VIH à celle de comorbidités chroniques, comme les maladies cardiovasculaires, les maladies rénales, le cancer ainsi que des troubles neurologiques et osseux.

L'étude des déterminants qui concourent à une activation immunitaire chronique chez les patients adultes atteints de VIH traités avec une charge virale indétectable constitue un domaine de recherche intense, plusieurs hypothèses de recherche ayant été explorées. L'interféron alpha a un rôle majeur dans l'activation immunitaire chronique et est rapporté comme un facteur prédictif de la progression d'infections persistantes [Tejjaro et al., 2013 ; Wilson et al., 2013 ; Zhou et al., 2014], incluant le VIH [Acchioni et al., 2015 ; Cha et al., 2014 ; Ries et al., 2012]. Le cytomégalo virus (CMV) et la translocation microbienne [Brenchley et al., 2006 ; Jiang et al., 2009, Klatt et al., 2013] ont été rapportés comme des facteurs de genèse de l'activation immunitaire chronique [Appay et Sauce, 2008]. Du fait de la prévalence et de l'intensité de la réponse immunitaire au cytomégalo virus, des interventions contre l'infection chronique à cytomégalo virus ont été testées [Hunt et al., 2011]. La réponse à l'auto-immunité a également été suggérée comme impliquée mais reste très peu explorée dans le VIH [Stratton et al. 2006], hormis les travaux spécifiques de Rawson et al, sur le rôle, dans l'infection chronique à VIH, de certains antigènes issus de produits

apoptotiques, tels que l'actine, la vimentine et la lamine [Rawson et al., 2007]. Ainsi, les déterminants qui concourent à une activation immunitaire chronique sont multifactoriels. En particulier, la contribution respective à l'activation immunitaire chronique de chacun des facteurs proposés reste encore mal élucidée.

4.2. Travaux antérieurs utilisant une méthode de régression linéaire multivariante

Cet état des connaissances a donné lieu à la mise en place, au CHU de Bordeaux, d'une étude financée par un appel d'offre interne et Sidaction : étude transversale ACTHIV nichée au sein de la Cohorte ANRS CO3 Aquitaine de patients infectés par le VIH, Dr Isabelle Pellegrin, Laboratoire d'Immunologie en collaboration avec le Service de Maladies Infectieuses et Médecine Interne, CHU de Bordeaux. L'objectif de cette étude était d'évaluer le rôle de l'infection par le cytomégalovirus, de l'auto-immunité et de la translocation microbienne dans l'activation immunitaire chronique, mesurée par la proportion de lymphocytes T CD8+ exprimant les marqueurs HLA-DR+ et CD38+ chez les patients atteints de VIH traités avec une charge virale indétectable.

Au total, 191 patients ont été inclus dans l'étude ACTHIV. Les travaux préalables de notre équipe, utilisant des méthodes de régression linéaire multivariante, ont rapporté que la réponse immunitaire induite par le cytomégalovirus, mesurée par plusieurs marqueurs, était indépendamment associée à l'activation immunitaire chronique (analyse de la positivité du test Quantiféron ($p=0,0238$), analyse de la positivité du test ELISPOT ($p=0,0141$) et, dans le sous-groupe des patients HLA-A*0201, analyse de la positivité des lymphocytes T CD8+ pp65 spécifiques du cytomégalovirus ($p=0,0538$)). En revanche, aucune association n'était retrouvée entre l'activation immunitaire chronique et la réponse immunitaire induite par l'auto-immunité en analyses ajustées. De même, aucune association n'était retrouvée entre l'activation immunitaire chronique et les marqueurs de translocation microbienne (rADN 16S et marqueur CD14 soluble).

Les résultats de ce travail ont fait l'objet d'une publication courte dans le Journal of Infectious Diseases en 2013 [Wittkop et al., 2013]. Le tableau synthétisant les principaux résultats de cette étude est présenté ci-après.

Table 1. Univariable and Multivariable Linear Regression Analysis of Variables With a Potential Influence on Chronic Immune Activation

Variable	Univariable				Multivariable (n = 167), R ² = 0.16				Multivariable (n = 165), R ² = 0.14				Multivariable (n = 164), R ² = 0.14				Multivariable (n = 70), R ² = 0.20			
	Reg Coeff ^a	SD	Stand Coeff ^b	P	Reg Coeff ^a	SD	Stand Coeff ^b	P	Reg Coeff ^a	SD	Stand Coeff ^b	P	Reg Coeff ^a	SD	Stand Coeff ^b	P	Reg Coeff ^a	SD	Stand Coeff ^b	P
Result of CMV serologic analysis, pos vs neg	0.226	0.070	0.23	.0014	0.05	0.215	0.077	0.2	.006
Result of Quantiferon CMV analysis, pos vs neg	0.130	0.045	0.21	.0042	0.04	0.106	0.047	0.17	.0238
Result of CMV-pp65 ELISpot analysis, pos vs neg	0.191	0.062	0.22	.0025	0.05	0.166	0.067	0.18	.0141
Result of CMV-pp65-specific CD8 ⁺ T-cell analysis, pos vs neg	0.144	0.079	0.20	.0716	0.04	0.176	0.09	0.23	.0538	...
Result of actin-specific CD8 ⁺ T-cell analysis, pos vs neg	0.098	0.064	0.18	.1290	0.03
Age, per 1-y increase	0.005	0.002	0.22	.0026	0.05	0.004	0.002	0.18	.0161	0.004	0.002	0.18	.0194	0.004	0.002	0.18	.0204	0.007	0.002	0.28
CD4 ⁺ T-cell count, per 100-cells/ μ L increase	-0.023	0.006	-0.26	.0003	0.07	-0.015	0.009	-0.17	.0929	-0.015	0.009	-0.17	.1144	-0.016	0.009	-0.18	.0900	-0.020	0.015	-0.24
16S rDNA load, per 1-log ₁₀ -copies/mL increase	-0.081	0.055	-0.11	.1431	0.01	-0.048	0.052	-0.07	.3612	-0.080	0.053	-0.11	.1366	-0.058	0.053	-0.08	.2698	-0.115	0.087	-0.16
Regulatory T-cell count, per 1000-cells/ μ L increase	-2.595	0.793	-0.23	.0013	0.05	-0.627	1.151	-0.06	.5867	-0.742	1.166	-0.07	.5255	-0.550	1.164	-0.05	.6374	0.830	1.919	0.08

Abbreviations: CMV, cytomegalovirus; ELISpot, enzyme-linked immunosorbent spot; neg, negative; pos, positive; rDNA, ribosomal DNA; reg coeff, regression coefficient; stand coeff, standardized regression coefficient. ^aIndicates the difference in the log₁₀ percentage of CD8⁺ T cells that expressed HLA-DR+/CD38+ with respect to variables adjusted for in the model. For example, consider Result of Quantiferon CMV analysis, pos vs neg. In a comparison of 2 patients of the same age and with the same 16S ribosomal DNA load and regulatory T-cell count, a patient with a CD4⁺ T-cell count that is 100 cells/ μ L greater than the count for the other patient has 0.015 log₁₀ lower percentage of CD8⁺ T cells that expressed HLA-DR+/CD38+. ^bComputed by multiplying the regression coefficient by the sample SD of the variable, and dividing the quotient by the SD of the outcome variable. The standardized coefficient is interpretable as the average change in the outcome variable, in SDs, for a 1-SD increase in an explanatory variable, holding constant the other explanatory variables. If there are just 2 variables (ie, the outcome variable and 1 explanatory variable), then the standardized coefficient equals the correlation coefficient. Thus, standardized regression coefficients reveal the relative importance of each of the explanatory variable included in the model to explain the outcome variable.

Deux limites sont à souligner dans cette analyse. Chaque processus immunologique, c'est-à-dire la réponse immunitaire induite par le cytomégalovirus, la réponse immunitaire induite par l'auto-immunité et la translocation microbienne, peut être mesuré par différents marqueurs. Ainsi, chaque analyse séparée, selon les marqueurs disponibles dans l'étude ACTHIV, ne reflète qu'en partie le processus immunologique sous-jacent. De plus, les déterminants qui concourent à une activation immunitaire chronique chez les patients avec une charge virale indétectable sous traitement sont multifactoriels, impliquant également un rôle majeur de l'interféron alpha. La réponse immunitaire induite par l'auto-immunité est notamment étroitement liée à la sécrétion d'interféron alpha dans de nombreuses maladies auto-immunes chroniques [Banchereau et Pascual, 2006 ; Blanco et al., 2008] (« signature interféron » indirectement estimée par l'analyse de l'expression des gènes de l'interféron alpha). A ce jour, elle reste insuffisamment explorée dans le VIH. En particulier, dans l'étude préalable, le rôle de la réponse immunitaire induite par l'auto-immunité a pu être sous-estimé du fait que le dosage semi-quantitatif des anticorps anti-noyaux n'est potentiellement pas assez fin pour évaluer cette réponse dans le VIH et du fait des faibles valeurs de pourcentages des lymphocytes T CD8+ spécifiques des antigènes issus de produits apoptotiques (actine, vimentine, lamine), pris en compte dans des analyses séparées. En parallèle, les relations entre les processus biologiques sont complexes et potentiellement non directes deux à deux, impliquant des processus intermédiaires. De plus, le cytomégalovirus est également capable de stimuler la production d'interféron alpha via l'expression des gènes de l'interféron alpha [Yang et al., 2005].

Ainsi, dans le contexte de l'activation immunitaire chronique chez les patients atteints de VIH traités efficacement, des méthodes alternatives aux méthodes classiques apparaissent nécessaires pour prendre en compte les marqueurs multiples reflétant un processus immunologique et modéliser des relations complexes entre ces processus immunologiques. Les méthodes d'analyse dites intégratives visent à prendre en compte l'ensemble des données disponibles. Parmi ces méthodes, les modèles d'équations structurelles sont des méthodes pour lesquelles un certain nombre d'hypothèses sont faites quant aux associations qu'on peut s'attendre à observer. Les modèles d'équations structurelles permettent de prendre en compte de multiples variables décrivant un ou plusieurs processus et la modélisation explicite des relations entre les processus. Ces modèles apparaissent dès lors d'un intérêt majeur dans l'étude de processus multifactoriels. C'est cette dernière approche d'analyse intégrative par modèle d'équations structurelles que nous avons proposé, comme alternative aux méthodes de

régression, pour mieux comprendre les déterminants qui concourent à l'activation immunitaire chronique chez les patients atteints de VIH traités efficacement.

La description et les concepts généraux des modèles d'équations structurelles sont présentés en section 4.3 et l'application de cette méthode dans l'étude ACTHIV en section 4.4. Une discussion de ces travaux est proposée en section 4.5.

4.3. Apport des modèles d'équations structurelles pour étudier des relations complexes entre variables

Les modèles d'équations structurelles, nommés SEM pour Structural Equation Modelling, ont été développés initialement dans le domaine des sciences sociales et comportementales à partir des années 1970 [Pearl, 2001]. Les modèles d'équations structurelles permettent de modéliser des variables appelées latentes, variables qui existent conceptuellement mais ne sont pas directement observables ou observées [Bollen, 2002]. Ce concept de variable latente trouve son sens notamment en sciences sociales et comportementales où l'on cherche à mesurer des phénomènes non directement observables, tels que la perception d'un phénomène donné, pour exemple la qualité de vie, concept multidimensionnel mesuré par un ensemble d'échelles, expliquant ainsi la popularité des modèles d'équations structurelles dans ces disciplines [Bollen et Noble, 2011].

Les modèles d'équations structurelles correspondent à des modèles de régression représentant les relations supposées entre les données et sont utilisés pour tester la significativité statistique de ces relations. Du fait de leur utilité, étendre leurs applications à des champs autres, tels que la recherche biomédicale, a été souligné ces dernières années [Beran et Violato, 2010], ouvrant ainsi la voie à l'apport de ces méthodes dans le champ de l'épidémiologie analytique [Arlinghaus et al., 2012 ; Bollen et al., 2013 ; Dahly et al., 2009 ; Mansiaux et al., 2015 ; Tu, 2009 ; VanderWeele, 2012].

4.3.1. Variables observées et variables latentes

Les méthodes classiques de régression se basent sur l'utilisation de variables observées uniquement. Les modèles d'équations structurelles se basent sur l'utilisation de deux types de variables : les variables observées et les variables latentes. Comme présenté, les variables latentes sont des variables non observées, pour lesquelles aucune mesure directe n'existe mais dont l'existence peut être révélée par des associations avec des variables observées, supposées être le reflet d'un phénomène latent non observable [Grace et al., 2012]. Ceci sous-entend

ainsi que plusieurs variables observées peuvent être des mesures imparfaites d'un concept sous-jacent [Chavance et al., 2010].

4.3.2. Diagramme de chemin

En fonction de la complexité du phénomène à modéliser, une façon simple de représenter les hypothèses formulées est d'utiliser un diagramme de chemin (path diagram). Les modèles SEM pouvant rapidement contenir un grand nombre de variables et de relations, et par conséquent reposer sur un grand nombre d'équations, les diagrammes de chemin constituent un outil majeur pour une représentation graphique des hypothèses formulées, phase cruciale qui sera une base pour tous les choix méthodologiques et l'écriture des modèles mathématiques [Dumas et al., 2014].

Des conventions régissent l'écriture des diagrammes de chemin [Budtz-Jorgensen et al., 2002]. Les ellipses représentent généralement les variables latentes tandis que les rectangles ou carrés sont le plus souvent utilisés pour les variables observées. Les flèches simples représentent le sens de la relation entre deux variables. Les conventions des modèles d'équations structurelles impliquent que les variables observées sont des variables dépendantes de la variable latente à laquelle elles sont associées. Il en résulte graphiquement que les flèches vont de la variable latente vers les variables observées correspondantes. Puis, à partir de la représentation graphique des relations supposées entre variables, sont ensuite écrites l'ensemble des équations structurelles associées au modèle SEM.

4.3.3. Modèle d'équations structurelles complet

Les modèles SEM constituent ainsi une méthode d'analyse intégrative permettant de représenter, estimer et tester un modèle théorique explicatif de relations supposées linéaires entre plusieurs variables observées ou non, et où des effets directs et des effets indirects peuvent être quantifiés [Ditlevsen et al., 2005].

Le modèle SEM complet est composé de deux sous-modèles, le modèle de mesure (measurement model) et le modèle structurel (structural model). Le modèle de mesure permet de définir le nombre de variables latentes ainsi que la manière dont elles sont liées aux données observées. Classiquement, le modèle structurel est une régression linéaire des variables latentes les unes aux autres. La combinaison du modèle de mesure et du modèle structurel constitue le modèle SEM complet [Sanchez et al., 2005].

4.3.4. Exemple d'un diagramme de chemin et équations structurelles associées

A partir de l'exemple de diagramme de chemin illustré en figure 3 ci-après, sont présentées les équations structurelles associées.

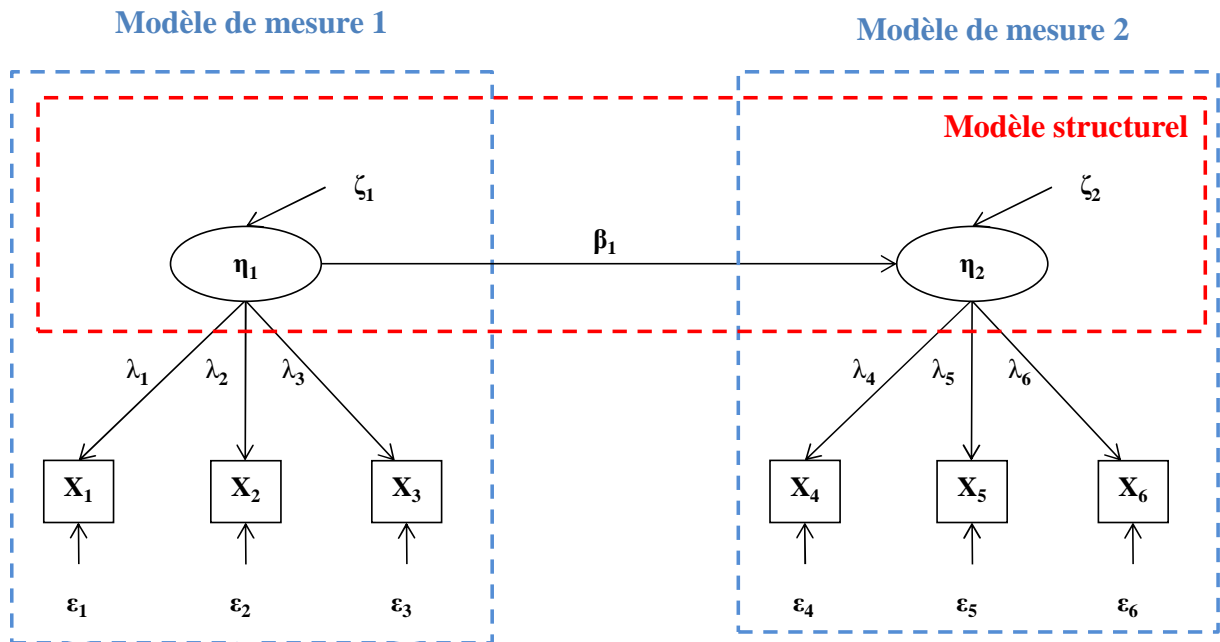


Figure 3. Exemple de diagramme de chemin d'un modèle SEM.

Légende. η : variable latente (ellipse), X : variable observée (rectangle ou carré), λ : coefficient liant une variable latente à une variable observée, β : coefficient de régression liant deux variables latentes, ε : erreur de mesure d'une variable observée, ζ : erreur résiduelle d'une variable latente. Intercepts non représentés.

Les équations correspondant au modèle de mesure 1 peuvent être représentées ainsi :

$$X_1 = \lambda_1 \eta_1 + \varepsilon_1$$

$$X_2 = \lambda_2 \eta_1 + \varepsilon_2$$

$$X_3 = \lambda_3 \eta_1 + \varepsilon_3$$

Les équations correspondant au modèle de mesure 2 peuvent être représentées ainsi :

$$X_4 = \lambda_4 \eta_2 + \varepsilon_4$$

$$X_5 = \lambda_5 \eta_2 + \varepsilon_5$$

$$X_6 = \lambda_6 \eta_2 + \varepsilon_6$$

L'équation du modèle structurel peut être représentée ainsi.

$$\eta_2 = \beta_1 \eta_1 + \zeta_2$$

4.3.5. Construction d'un modèle SEM et sélection du modèle

Une fois le diagramme de chemin théorique des relations supposées établi, le modèle SEM va estimer et tester l'adéquation des données à ce modèle théorique. Pour éviter les problèmes d'identifiabilité, il est généralement recommandé de dimensionner les variables latentes, l'approche la plus classique étant de fixer la variance de la variable latente à 1. Parmi les différentes méthodes d'estimation existantes, l'estimateur du maximum de vraisemblance est couramment utilisé et cherche à maximiser la vraisemblance des estimations par rapport aux données. L'intérêt de la construction d'un modèle SEM est de savoir dans quelle mesure les hypothèses construites sont concordantes avec les données. Une fois les paramètres estimés, l'évaluation globale de l'adéquation du modèle est réalisée via différents index rapportés dans les SEM, dont les plus connus sont ceux rapportés par Hu LT et Bentler PM [Hu et Bentler, 1999], puis la significativité statistique des estimations est le dernier critère employé pour évaluer la pertinence des paramètres individuels pour le modèle. Une fois le modèle final établi, les paramètres standardisés sont généralement présentés, permettant ainsi une comparaison des forces d'associations relatives entre variables observées et/ou latentes.

Les aspects présentés ci-dessus pour la construction d'un modèle SEM sont explicités en détails dans la revue proposée par Sanchez BN et al [Sanchez et al., 2005].

4.4. Application

Ayant à disposition dans l'étude ACTHIV plusieurs marqueurs pour chaque processus immunologique ainsi que des données d'expression des gènes de l'interféron alpha, nous avons investigué plus en avant, via une analyse intégrative par modèles d'équations structurelles, le rôle de la réponse immunitaire induite par le cytomégalovirus, la réponse immunitaire induite par l'auto-immunité et la translocation microbienne dans l'activation immunitaire chronique ainsi que le rôle de l'interféron alpha dans ces relations. L'utilisation de variables latentes a permis de prendre en compte l'ensemble des marqueurs caractérisant un processus immunologique. Ainsi, quatre variables latentes ont été considérées :

- i) une variable latente « réponse immunitaire induite par le cytomégalovirus », à partir des valeurs des tests Quantiféron, pp-65 ELISPOT et des pourcentages de lymphocytes T CD8+ pp65 spécifiques du cytomégalovirus,
- ii) une variable latente « réponse immunitaire induite par l'auto-immunité », à partir des anticorps anti-noyaux et des pourcentages de lymphocytes T CD8+ spécifiques des antigènes issus de produits apoptotiques (actine, vimentine, lamine),

iii) une variable latente « translocation microbienne », à partir des marqueurs rADN 16S et CD14 soluble,

iii) une variable latente « interféron alpha », à partir des données d'expression de 5 gènes de l'interféron alpha.

Les relations supposées entre variables ont été basées sur la littérature et des discussions avec les immunologistes. Nous avons formulé l'hypothèse d'effets directs de la réponse immunitaire induite par le cytomégalovirus, de la réponse immunitaire induite par l'auto-immunité et de la translocation microbienne sur l'activation immunitaire chronique mais aussi d'effets indirects via un rôle médiateur de l'interféron alpha. Le diagramme de chemin théorique proposé est présenté en figure 1 de l'article section 4.4.1.

4.4.1. Article soumis à AIDS

Ce travail a donné lieu à l'écriture d'un article original, actuellement soumis à AIDS et présenté ci-après.

AIDS

Integrative analysis of immunological data to explore chronic immune activation in successfully treated HIV patients.

--Manuscript Draft--

Manuscript Number:	AIDS-D-15-00881
Full Title:	Integrative analysis of immunological data to explore chronic immune activation in successfully treated HIV patients.
Article Type:	Original paper (Clinical)
Keywords:	Cytomegalovirus-induced immune response; autoimmune-induced immune response; microbial translocation; alpha-interferon-stimulated genes; HIV chronic immune activation; integrative analysis; structural equation modelling.
Corresponding Author:	Rodolphe THIÉBAUT, MD, PhD INSERM U897 Bordeaux, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	INSERM U897
Corresponding Author's Secondary Institution:	
First Author:	Marie-Quitterie PICAT
First Author Secondary Information:	
Order of Authors:	Marie-Quitterie PICAT Isabelle PELLEGRIN Juliette BITARD Linda WITTKOP Cécile PROUST-LIMA Benoît LIQUET Jean-François MOREAU Fabrice BONNET Patrick BLANCO Rodolphe THIÉBAUT, MD, PhD
Order of Authors Secondary Information:	
Abstract:	<p>Objectives. To unravel the complex relationships between cytomegalovirus-induced-, autoimmune-induced immune responses, microbial translocation and chronic immune activation (CIA) in successfully treated HIV-infected patients and to explore the mediating role of alpha-interferon in these processes.</p> <p>Design. Cross-sectional study nested in the ANRS CO3 Aquitaine Cohort, a prospective hospital-based cohort of HIV-1-infected patients in South-Western France.</p> <p>Methods. Patients initiated antiretroviral therapy between 2005 and 2008 and were treated with sustained virological suppression for at least two years. CIA was defined by % HLA-DR+/CD38+CD8+T-cells. Integrative analyses were conducted using structural equation modelling (SEM).</p> <p>Results. The main analysis was performed in 57 HLA-A*0201 positive patients, due to availability of % of actin-, vimentin-, lamin-specific CD8+T-cells (HLA-A2-restricted tests) to further characterize autoimmune response. Cytomegalovirus-induced immune response was assessed by Quantiferon and pp-65 ELISPOT. SEM revealed a direct effect of cytomegalovirus-induced immune response on CIA (standardized estimate $\beta_{std}=0.56$, p-value=0.0004). The effect of autoimmune-induced immune response on</p>

	<p>CIA was indirect through alpha-interferon pathway, assessed by expression levels of 5 alpha-interferon-stimulated genes ADAR, ISG15, IFIT1, Mx1 and OAS1 (effect of autoimmune response on alpha-interferon: βstd=0.36, p-value=0.0401; effect of alpha-interferon on CIA: βstd=0.39, p-value=0.0044). Microbial translocation as measured by 16SrDNA and sCD14 was not associated with CIA. Results were consistent in 142 patients in whom cytomegalovirus and auto-immunity responses were measured by Quantiferon and anti-nuclear antibodies, respectively.</p> <p>Conclusion. The role of cytomegalovirus-immune response on CIA was confirmed as well as the involvement of alpha-interferon on CIA. An effect of auto-immunity response on CIA was revealed through the alpha-interferon pathway.</p>
<p>Suggested Reviewers:</p>	<p>Victor APPAY victor.appay@upmc.fr Sorbonne Universités, Université Pierre et Marie Curie, CR7, Centre d'Immunologie et des Maladies Infectieuses (CIMI), F-75013, Paris, France.</p> <p>Raphael GOTTARDO rgottard@fhcrc.org Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M2-C200 Seattle, WA 98109, United States of America</p> <p>Steven Grant DEEKS sdeeks@php.ucsf.edu Department of Medicine, San Francisco General Hospital, University of California, San Francisco, San Francisco, CA 94110, United States of America</p> <p>Laurence WEISS laurence.weiss@egp.aphp.fr Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France</p>
<p>Opposed Reviewers:</p>	

TITLE PAGE

Full title. Integrative analysis of immunological data to explore chronic immune activation in successfully treated HIV patients.

Character count for full title: 118 (120 characters maximum)

Short title/Running head. Chronic immune activation in HIV.

Character count for short title: 33 (40 characters maximum)

Authors and affiliations

Marie-Quitterie PICAT^{1,2,3,4,5}, Isabelle PELLEGRIN⁶, Juliette BITARD⁶, Linda WITTKOP^{1,2,3,4}, Cécile PROUST-LIMA^{1,2}, Benoît LIQUET^{4,5,7}, Jean-François MOREAU^{6,8,9}, Fabrice BONNET^{1,8,10}, Patrick BLANCO^{6,8,9}, Rodolphe THIEBAUT^{1,2,3,4,5} and the ANRS CO3 Aquitaine Cohort

¹ INSERM, ISPED, Centre INSERM U897 - Epidemiologie-Biostatistique, F-33076 Bordeaux, France

² Univ. Bordeaux, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, F-33076Bordeaux, France

³ Service d'Information Médicale, USMR, CHU de Bordeaux - Pôle de Santé Publique, F-33076 Bordeaux, France

⁴ INRIA, Team SISTM, F-33076 Bordeaux, France

⁵ Vaccine Research Institute-VRI, Hôpital Henri Mondor, Créteil, France

⁶ Laboratoire d'Immunologie-Immunogénétique, Pôle de Biologie, CHU Bordeaux, F-33000 Bordeaux, France

⁷ School of Mathematics and Physics, The University Of Queensland, Saint Lucia, Australia

⁸ Univ. Bordeaux, F-33076 Bordeaux, France

⁹ CNRS, UMR 5164, F-33076 Bordeaux, France

¹⁰ Service de Médecine Interne et maladies Infectieuses, CHU de Bordeaux, F-33076 Bordeaux, France

Corresponding author

Prof Rodolphe THIEBAUT, MD PhD

INSERM U897, ISPED

Université de Bordeaux

146 Rue Léo Saignat

33076 Bordeaux, France

E-mail: rodolphe.thiebaut@isped.u-bordeaux2.fr

Tel: +33 5 57 57 45 21 - Fax: +33 5 57 57 15 78

Word count for text: 3439 words (*3500 words maximum*)

Word count for abstract: 250 (*250 words maximum*)

Reference count: 42

Table count: 2

Figure count: 3

Funding

This work was supported by Sidaction [AI20-2-01629], the CHU of Bordeaux [2009-AO1063-54] and the Vaccine Research Institute.

Abstract: 250 words (250 words maximum)

Objectives. To unravel the complex relationships between cytomegalovirus-induced-, autoimmune-induced immune responses, microbial translocation and chronic immune activation (CIA) in successfully treated HIV-infected patients and to explore the mediating role of alpha-interferon in these processes.

Design. Cross-sectional study nested in the ANRS CO3 Aquitaine Cohort, a prospective hospital-based cohort of HIV-1-infected patients in South-Western France.

Methods. Patients initiated antiretroviral therapy between 2005 and 2008 and were treated with sustained virological suppression for at least two years. CIA was defined by % HLA-DR+/CD38+CD8+T-cells. Integrative analyses were conducted using structural equation modelling (SEM).

Results. The main analysis was performed in 57 HLA-A*0201 positive patients, due to availability of % of actin-, vimentin-, lamin-specific CD8+T-cells (HLA-A2-restricted tests) to further characterize autoimmune response. Cytomegalovirus-induced immune response was assessed by Quantiferon and pp-65 ELISPOT. SEM revealed a direct effect of cytomegalovirus-induced immune response on CIA (standardized estimate $\beta_{std}=0.56$, p-value=0.0004). The effect of autoimmune-induced immune response on CIA was indirect through alpha-interferon pathway, assessed by expression levels of 5 alpha-interferon-stimulated genes ADAR, ISG15, IFIT1, Mx1 and OAS1 (effect of autoimmune response on alpha-interferon: $\beta_{std}=0.36$, p-value=0.0401; effect of alpha-interferon on CIA: $\beta_{std}=0.39$, p-value=0.0044). Microbial translocation as measured by 16SrDNA and sCD14 was not associated with CIA. Results were consistent in 142 patients in whom cytomegalovirus and auto-immunity responses were measured by Quantiferon and anti-nuclear antibodies, respectively.

Conclusion. The role of cytomegalovirus-immune response on CIA was confirmed as well as the involvement of alpha-interferon on CIA. An effect of auto-immunity response on CIA was revealed through the alpha-interferon pathway.

Key words: 7 words (5 to 7 maximum)

Cytomegalovirus-induced immune response; autoimmune-induced immune response; microbial translocation; alpha-interferon-stimulated genes; HIV chronic immune activation; integrative analysis; structural equation modelling.

INTRODUCTION

Immune activation is a leading factor for human immunodeficiency virus (HIV) disease progression [1, 2]. In patients with long-term virological suppression while on antiretroviral therapy combination, persistence of immune activation is associated to both acquired immunodeficiency syndrome (AIDS) and non-AIDS comorbidities [3-5]. Noteworthy, chronic type I alpha-interferon (IFN- α) is a hallmark in chronic activation and is reported in disease progression in persistent infections [6-8], including HIV [9-11]. The mechanisms driving chronic immune activation and type I interferon secretion are multifactorial and remain not completely understood [12]. Cytomegalovirus (CMV) and microbial translocation have been proposed as triggering factors of persistent immune activation-associated to HIV-1 infection [13]. Microbial translocation and associated biomarkers such as sCD14 and circulating plasma lipopolysaccharide [14-16] have been implicated in the generation and maintenance of chronic immune activation. Cytomegalovirus is highly prevalent in HIV patients and may reactivate more frequently [17]. Because of its prevalence, the intensity of the immune response to CMV, and its reactivation in immunocompromised patients, interventions against chronic CMV infection have been tested to reduce T-cell activation in HIV-infected individuals with incomplete CD4⁺ T cell recovery on antiretroviral therapy [18].

Auto-immune induced immune activation has also been suspected although less consistently [19]. Hence, an increase of the proportion of autoreactive T-cells has been shown by Rawson et al [20]. The contribution of each of those factors that could vary from patients to patients and over the course of the disease is not well described. Therefore, we previously evaluated the impact of cytomegalovirus-induced, microbial translocation and autoimmune-induced immune responses on chronic immune activation in 191 HIV-1-infected patients with long-term virological suppression while on antiretroviral therapy combination in the ACTHIV substudy of the ANRS CO3 Aquitaine cohort [21]. We reported that the immune response to CMV was independently associated with a higher percentage of CD8⁺ T cell activation as measured by HLA-DR⁺/CD38⁺CD8⁺T-cells. In contrast, autoimmune response and microbial translocation were not found to be associated with chronic immune activation in adjusted analyses. However, two difficulties limited these analyses. Firstly, each process can be measured by several biomarkers. For instance, the immune response to CMV can be measured by Quantiferon-CMV-, CMV-pp65-ELISPOT- and CMV-pp65-specific-CD8⁺T-cell-positivity. Therefore, each analysis considering only one marker does not reflect the whole process. Secondly, the mechanism leading to chronic immune activation is complex,

multifactorial and involved intermediate pathway such as type I interferon. Autoimmunity, autoimmune diseases and the type I interferon system are tightly linked. As example, a type I interferon signature can be found in many autoimmune disorders including lupus, dermatomyositis or type I diabetes. The mechanisms involved in type I interferon secretion in patients implicate the immune complexes-mediated activation of endosomal TLR activation by self DNA [22, 23]. Interestingly, it has been shown in chronic HIV patients that the cross-presentation of caspase-cleaved self-antigen by dendritic cells was promoting the expansion and activation of self-reactive CD8+ T cells against vimentin as an example [20]. A similar mechanism has been described in systemic lupus erythematosus patients where type I interferon promote the differentiation of monocytes into dendritic cells, able to engulf apoptotic bodies and present autoantigens to the adaptive immune system. Conversely, vimentin is an autoantigen targeted by the immune response in systemic lupus erythematosus patients [24]. In parallel, cytomegalovirus infection may also be able to stimulate type I IFN- α production through interferon-stimulated-genes [25].

We further investigated the hypotheses of direct effects of CMV-induced, autoimmune-induced immune responses as measured by percentages of actin-, vimentin-, lamin-specific CD8+T-cells and microbial translocation on chronic immune activation and indirect effects through IFN- α gene expression, in successfully treated HIV-infected patients. We aimed at unravelling the complex relationships between CMV-induced-, autoimmune-induced immune responses, microbial translocation and chronic immune activation in these patients and to explore the mediating role of the IFN- α in these processes. To do so, we performed an integrative analysis using the maximal amount of immunological information available in the ACTHIV study using structural equation modelling [26].

METHODS

Population

The ACTHIV study [21] is a cross-sectional study nested in the ANRS CO3 Aquitaine Cohort [27], a prospective hospital-based cohort of HIV-1-infected patients, in South-Western France. Patients initiated antiretroviral therapy between 2005 and 2008, and were treated with sustained virological suppression (HIV-1 RNA load below the detection limit of 50 copies/mL) for at least two years. Exclusion criteria were two consecutive HIV-1 RNA above 50 copies/mL, hepatitis C virus or hepatitis B virus coinfection, and signs of acute infection. The ACTHIV protocol was approved by the Bordeaux University Institutional Review Board. All patients signed an informed consent before participating in the study.

Measurements

Available biomarkers in the ACTHIV study for assessing CMV-induced-, autoimmune-induced immune responses, microbial translocation, quantitation of IFN- α gene expression and chronic immune activation are presented below.

Cytomegalovirus-induced immune response

The CMV-induced immune response was determined using Quantiferon CMV analysis (Cellestis, Australia), CMV-pp65 enzyme-linked immunosorbent spot (EliSpot) analysis, and CMV-pp65-specific tetramers CD8+. Patients were considered Quantiferon-CMV positive if they obtained a value above 0.2 IU/mL, as defined by the manufacturer.

Autoimmune-induced immune response

Self-antigens-immune response was evaluated with anti-nuclear antibody titers (titers \geq 1:250 were considered as positive) and, in HLA-A*0201 positive patients, with percentages of actin-, vimentin-, lamin-HLA-A*0201 tetramers CD8+T-cells.

Microbial translocation

Bacterial 16SrDNA and soluble CD14 (sCD14) were quantified as markers for microbial translocation.

Quantification of IFN- α -stimulated-genes: ADAR, ISG15, IFIT1, Mx1 and OAS1 gene expression

IFN- α gene expression was assessed using quantitative reverse transcription polymerase chain reaction (qRT-PCR) of five IFN- α -stimulated-genes: ADAR, ISG15, IFIT1, Mx1 and OAS1 [28]. Total RNA was extracted from peripheral blood mononuclear cells (obtained after Ficoll

preparation and storage with Protector RNase inhibitor (Roche applied biosciences), using the High pure RNA isolation kit (Roche applied biosciences). RNA extraction quality control was measure with the Agilent 2100 bioanalyzer in combination with Agilent RNA 6000 Nano kit (RNA sample quality was considered if the obtained RIN was >7). A total of 300 ng of RNA was reverse transcribed into cDNA using the Transcriptor Reverse Transcriptase kit (Roche applied biosciences). PCR was performed (RT2 SYBR Green qPCR MasterMix, SA Biosciences) in 96-wells plates using LightCycler 480 (Roche Diagnostics) to determine the expression levels of ADAR, ISG15, IFIT1, Mx1 and OAS1 target genes. IFN- α gene expression measurements were expressed as fold change which is calculated as a ratio of averages from control and test sample values.

Chronic immune activation (outcome)

Chronic immune T-cell activation was evaluated by measuring by flow cytometry the percentage of HLA-DR+CD38+CD8+T-cells among the CD8+ T-cell subset. Methodology of laboratory measurements by flow cytometry is reported in details in Wittkop et al [21] (Material and Methods subsection entitled “Laboratory measurements” and supplementary material, Table 1).

Statistical method and model assignment

Structural equation modelling

Structural equation modelling with latent variables was applied to perform the integrative analysis of available biomarkers in the ACTHIV study. Latent variables are unobserved variables for which no direct measurements exist, but whose existence may be revealed by associations among measured variables [29]. Latent variables translate the fact that several observed variables could be imperfect measurements of a single underlying concept [30]. The general structural equation model is typically decomposed into two submodels: a measurement model and a structural model [26]. Firstly, the measurement model defines the construction of the latent variables with the corresponding observed measured variables. Secondly, the structural model specifies the pattern by which latent variables and observed variables influence each other, either directly or either indirectly.

We hypothesized direct effects of CMV-induced, autoimmune-induced immune responses and microbial translocation on chronic immune activation and also indirect effects (mediating role) of these factors through IFN- α gene expression.

As biomarkers (tetramers), more relevant than anti-nuclear antibody titers in chronic HIV [20], were available in HLA-A*0201 positive patients to characterize autoimmune-induced immune response, the main integrative analysis was performed in HLA-A*0201 positive patients. We considered four latent variables that are supposed to be linked with chronic immune activation:

- CMV-induced immune response latent variable measured by Quantiferon-CMV, CMV-pp65-ELISPOT and CMV-pp-65-specific-CD8+T-cells,
- autoimmune-induced immune response latent variable measured by anti-nuclear antibody titers and percentages of actin-, vimentin-, lamin-specific CD8+T-cells,
- microbial translocation latent variable measured by 16SrDNA and sCD14,
- IFN- α latent variable measured by expression levels of 5 IFN- α -stimulated genes.

The theoretical path diagram is presented in Figure 1.

Statistical analyses

Qualitative variables were described as numbers (percentages) and quantitative variables as medians (first quartile; third quartile), using SAS 9.13/SAS 9.3 (SAS Institute, Cary, NC, USA).

Data preprocessing

The response variable, i.e. percentage of HLA-DR+/CD38+CD8+T-cells, was \log_{10} transformed. To correct expression data for differences between samples, IFN- α gene expression measurements were normalized to three reference genes GADPH, RPL13A, HPRT1 (housekeeping genes) using the $2^{-\Delta\Delta CT}$ method [31].

Structural equation models

Structural equation models were conducted in a complete case analysis (subjects with no missing values on the modeled variables).

Structural equation modelling results were reported as standardized estimates β_{std} and p-values. Standardized estimates represent the expected change in standard deviation units of a variable, due to one standard deviation increase in another. As a unit free representation, they allow comparisons of the relative strength of associations and thus easier interpretation of the coefficients.

Structural equation models were performed with R statistical software, using lava package for linear latent variable models [32]. Parameters were obtained by maximum likelihood estimation. Goodness of fit was assessed using three indexes [33, 34]: the chi-square test, the

Standardized Root Mean Square Residual (SRMR) and the incremental Bentler's Comparative Fit Index (CFI). Non-significant p-values at a 0.05 threshold for chi-square test, SRMR values below 0.08 and Bentler's CFI value above 0.95 indicate a good fit.

Sensitivity analyses were performed. In HLA-A*0201 positive sample, we considered the use of % CD8+ HLA-DR+T-cells as the outcome to define chronic immune activation. In the overall sample, we considered CMV-induced immune response as measured by positive versus negative Quantiferon-CMV and autoimmune-induced immune response as measured by positive versus negative anti-nuclear antibody titers.

RESULTS

Patients' characteristics

A total of 191 patients were enrolled in the ACTHIV study (women: 25%). Median age at inclusion was 50 years old (43; 58). Median duration of viral suppression was 6.1 years (4; 8.5). Median CD4+ and CD8+ T-cell counts were 517 cells/ μ L (387; 720) and 676 cells/ μ L (480; 977). Patients had a CD4+ nadir of 188 cells/ μ L (89; 275). The expression of HLA-DR+/CD38+ was 16.8% in median among total CD8+ T-cells.

Of the 191 patients, 87 were HLA-A*0201 positive patients. In these patients, the expression of HLA-DR+/CD38+ was 15.8% in median among total CD8+ T-cells. Demographic, clinical and biological characteristics of the 87 HLA-A*0201 positive patients were representative of the overall population of 191 patients (Table 1).

Measurement of the processes

Main patient's characteristics of variables related to CMV-, autoimmune-induced immune responses and microbial translocation are reported in Table 1. In the overall population, Quantiferon-CMV was positive in 78% of patients and antinuclear antibodies were positive in 26% of patients. In HLA-A*0201 positive patients, Quantiferon-CMV was positive in 82% and antinuclear antibodies were positive in 28% of patients. Levels of self-antigen-specific T cells were low in this study. Levels of 16SrDNA and sCD14 for studying microbial translocation did not differ in the HLA-A*0201 positive patients and the overall population (Table 1). Of the 87 HLA-A*0201 positive patients, the main complete analysis was performed in the 57 patients in whom the percentages of actin-, vimentin-, lamin-specific CD8+T-cells as well as Quantiferon, pp-65 ELISPOT for CMV were available (subjects with missing values on the modeled variables excluded). The 30 other patients did not differ from the 57 HLA-A*0201 positive patients included in the final analysis. In the 57 patients, median % of HLA-DR+/CD38+CD8+T-cells was 15.8 (11.7; 23.2). Median % actin-specific-CD8+ T cells, % vimentin-specific-CD8+ T cells, % lamin-specific-CD8+ T were respectively 0.03 (0.02; 0.06), 0.01 (0.01; 0.02) and 0.01 (0.01; 0.02). Median Quantiferon-CMV values were 5.3 IU/mL (0.4; 11.5) and median CMV-pp65-ELISPOT values were 309 (95; 645). IFN- α gene expression measurements were positively and significantly correlated with the percentages of HLA-DR+/CD38+CD8+T-cells (Figure 2).

Main integrative analysis in HLA-A*0201 positive patients

Biomarkers of microbial translocation were not associated with chronic immune activation in univariable analysis (16SrDNA effect: p-value=0.6100; sCD4 effect: p-value=0.9750; microbial translocation latent variable: p-value=0.9922) and in multivariable analysis performed in our previous work [21]. Anti-nuclear antibodies were not contributive for the definition of the autoimmune-induced immune response latent variable (p-value=0.2785) as well as CMV-pp65-specific-CD8+T-cells for the definition of the CMV-induced immune response latent variable (p-value=0.7049). Therefore, further analyses included the effect of CMV-induced immune response as measured by Quantiferon and pp-65 ELISPOT, autoimmune-induced immune response as measured by actin, vimentin, lamin-specific CD8+, interferon- α pathway as measured by the expression level of the five interferon signaling genes (ADAR, ISG15, IFIT1, Mx1 and OAS1) and chronic immune activation.

The analysis of the effect of each single process revealed a direct association between CMV-induced immune response latent variable and chronic immune activation ($\beta_{std}=0.42$, p-value=0.0232), IFN- α latent variable and chronic immune activation ($\beta_{std}=0.31$, p-value=0.0093) but no direct effect of autoimmune-induced immune response latent variable and chronic immune activation ($\beta_{std}=0.04$, p-value=0.7361). However, autoimmune-induced immune response latent variable was associated to IFN- α latent variable ($\beta_{std}=0.33$, p-value=0.0457).

Results from the model including all processes (Figure 3) are presented in Table 2. Quantiferon-CMV ($\beta_{std}=0.61$) and CMV-pp65-ELISPOT ($\beta_{std}=0.51$) contributed significantly to the CMV-induced immune response latent variable. The positive significant standardized estimates mean that higher level of CMV-induced immune response latent variable was associated to higher levels of Quantiferon-CMV and CMV-pp65-ELISPOT. Autoimmune-induced immune response latent variable was composed by % actin-specific-CD8+ T cells, % vimentin-specific-CD8+ T cells and % lamin-specific-CD8+ T cells. All standardized estimates were positive. The percentage of vimentin-specific-CD8+ T cells presented the highest contribution to define the autoimmune-induced immune response latent variable ($\beta_{std}=0.52$, p-value=0.0006). IFN- α latent variable was composed by the expression level of IFN- α -stimulated genes: ADAR ($\beta_{std}=0.59$), ISG15 ($\beta_{std}=0.54$), IFIT1 ($\beta_{std}=0.94$), Mx1 ($\beta_{std}=0.93$), OAS1 ($\beta_{std}=0.77$).

The structural model resented in Figure 3 revealed a direct positive significant effect of CMV-induced immune response latent variable on chronic immune activation ($\beta_{std}=0.56$, p-value=0.0004) but no significant association with the IFN- α latent variable (p-value=0.9069).

There was no direct significant association of autoimmune-induced immune response latent variable with chronic immune activation (p-value=0.3169). The effect of autoimmune-induced immune response latent variable on chronic immune activation was indirect through IFN- α latent variable: AI response latent variable was significantly associated to IFN- α latent variable (β_{std} =0.36, p-value=0.0401) and IFN- α latent variable was associated with chronic immune activation (β_{std} =0.39, p-value=0.0044). The goodness of fit of the model was fair (chi-square p-value=0.32; SRMR=0.07; CFI=0.98).

These findings persisted after adjustment for age, CD4+ T-cell count and regulatory T-cell count (n=57 patients; effect of CMV-induced immune response latent variable on chronic immune activation: β_{std} =0.49, p-value=0.0018; effect of AI response latent variable on IFN- α -stimulated-genes latent variable: β_{std} =0.35, p-value=0.0384; effect of IFN- α -stimulated-genes latent variable on chronic immune activation: β_{std} =0.38, p-value=0.0054).

When defining CD8+ T-cell activation with % CD8+HLA-DR+ T cells (\log_{10} transformed) in place of % CD8+CD38+HLA-DR+ as the outcome, the adjusted results were consistent. There was a direct positive significant effect of CMV-induced immune response latent variable on chronic immune activation (effect of CMV-induced immune response latent variable on chronic immune activation: β_{std} =0.45, p-value=0.0056) as well as an indirect effect of autoimmune-induced immune response latent variable on chronic immune activation through IFN- α -stimulated-genes latent variable. AI response latent variable was significantly associated to IFN- α -stimulated-genes latent variable (β_{std} =0.35, p-value=0.0345) and IFN- α -stimulated-genes latent variable was associated with chronic immune activation (β_{std} =0.35, p-value=0.0111).

Robustness analysis in the overall population

Of the 191 patients included in the ACTHIV study, the robustness analysis was performed in 142 patients in whom Quantiferon-CMV positive versus negative, anti-nuclear antibody titers positive versus negative, interferon-stimulated genes and % HLA-DR+/CD38+CD8+ T-cells were available (including the 57 patients from the main integrative analysis). In these patients, median % of HLA-DR+/CD38+CD8+T-cells was 16.9 (10.8; 23.7), Quantiferon-CMV was positive in 78% of patients and antinuclear antibodies were positive in 23% of patients. The adjusted results tend to be consistent with previous results but with less significant associations. Quantiferon-CMV positivity was associated with higher percentage of CD8+CD38+HLA-DR+ (β_{std} =0.15, p-value=0.0648), positive anti-nuclear antibody titers were associated with higher expression of IFN- α -stimulated-genes latent variable (β_{std} =0.16,

p-value=0.0797) and higher IFN- α -stimulated-gene expression was associated with chronic immune activation (β std=0.22, p-value=0.0120).

DISCUSSION

The structural equation modelling results revealed a direct positive significant effect of CMV-induced immune response latent variable on chronic immune activation. In addition, the approach suggested an indirect effect of autoimmune-induced immune response latent variable on chronic immune activation through IFN- α latent variable. These findings were consistent in various robustness analyses performed in HLA-A*0201 positive patients and in the overall population. Less significant results in the overall population could be explained by the fact that Quantiferon-CMV positivity and positivity of anti-nuclear antibody titers may not be enough accurate for reflecting respectively the CMV-induced immune response and the autoimmune-induced response leading to a dilution bias and underlying the need for an integrative analysis using all available information.

We confirmed the CMV-induced immune response effect on chronic immune activation [21]. According to this new analysis, the effect of CMV-induced immune response on CD8+ T cell activation was not mediated by the IFN- α -stimulated-genes measured in the present study (ADAR, ISG15, IFIT1, Mx1 and OAS1). The direct effect of CMV-induced immune response on chronic immune activation was revealed through the measurements of Quantiferon and pp-65 ELISPOT. CMV-pp-65-specific-CD8+T-cells were not contributive to define the CMV-induced immune response latent variable ($p=0.7049$). Stone et al [35] revealed that percentages of CMV-pp65-specific-CD8+ T-cells were similar in HIV patients with stable undetectable HIV viremia due to highly active antiretroviral therapy and healthy controls. Taking together, these findings indicate that CMV-pp65-specific-CD8+ T-cells may not be accurate when studying CMV-induced immune response in chronic HIV and thus, question the relevance of measuring them in this setting. Interestingly, our results revealed an indirect effect of autoimmune-induced response on chronic immune activation mediated through IFN- α latent variable. There were very few reports in the literature studying auto-immunity in HIV infected patients [19]. However, even in absence of clinical events, low level biological signals may have consequences such as T-cell activation. The effect of autoimmune-induced immune response on IFN- α latent variable was revealed through the measurements of actin-, vimentin-, lamin-specific CD8+T-cells illustrating the important role of caspase-cleaved apoptotic self-antigens in immune activation during chronic HIV infection, as suggested by Rawson et al [20]. Of note, the percentage of vimentin-specific CD8+T-cells was the most contributive to define the autoimmune-induced immune response latent variable. Moreover, the use of a latent variable seemed particularly relevant regarding the low levels of self-antigen-specific T cells in this study. Indeed, anti-nuclear antibodies were not contributive to

define the autoimmune-induced immune response latent variable and were dropped from the final model, revealing the need to measure biomarkers more relevant than anti-nuclear antibodies when studying autoimmune-induced immune response in chronic HIV [20].

The construction of a microbial translocation latent variable with 16SrDNA and sCD14 did not reveal an association with chronic immune activation, which is in agreement with our previous results [21] and results from a study aimed at examining the correlation between different methods to measure microbial translocation and its association with immune activation in long-term suppressed HIV-1-infected individuals [36].

In our study, all analyses highlighted the significant effect of IFN- α latent variable on chronic immune activation, accounting for the evidence of the involvement of the IFN- α pathway reported in chronic immune activation in successfully treated HIV-1-infected patients [9-11]. We only used five quantitative gene expression measurements of IFN- α -stimulated-genes. Some papers report more specific pathways of IFN- α that could be involved in HIV chronic immune activation, i.e. IFN- α MYD88 dependent pathway and IFN- α MYD88 independent pathway [37-39]. Indeed, ongoing efforts are made to identify and better characterize the IFN- α -stimulated-genes landscape [40, 41].

Our study also had some limitations. The ACTHIV study is a cross-sectional study nested in the ANRS CO3 Aquitaine Cohort which design does not allow us to infer the temporality and causal direction of the observed associations. The main complete case integrative analysis was performed in 57 HLA-A*0201, accounting for a potential lack of power. However, our results were confirmed in all robustness analyses.

Thus, further research is needed and could be tested in a longitudinal framework, allowing the inference of causal relationships, combining appropriate biomarkers and adequate quantitative gene expression measurements, including IFN- α and using structural equation modelling that could also be applied to deal with longitudinal data analysis [42].

In conclusion, the role of cytomegalovirus-immune response on the CD8⁺ T cell activation has been confirmed, encouraging intervention against cytomegalovirus in successfully treated HIV-1-infected patients [18], as well as the evidence of the IFN- α involvement in chronic immune activation. Furthermore, the signal found about the auto-immunity response as measured by anti-vimentin especially requires further studies to confirm, evaluate its impact and better explain its role on chronic immune activation.

Acknowledgments

We thank all patients for their participation. We thank the technicians from the immunologic routine laboratory of the CHU Bordeaux for excellent technical assistance.

Authors' contributions

Isabelle PELLEGRIN, Juliette BITARD, Jean-François MOREAU and Patrick BLANCO performed laboratory measurements by flow cytometry as well as quantitation of IFN- α gene expression by quantitative reverse transcription polymerase chain reaction. Linda WITTKOP and Rodolphe THIEBAUT participated in the study design. Isabelle PELLEGRIN, Jean-François MOREAU, Patrick BLANCO, Marie-Quitterie PICAT and Rodolphe THIEBAUT collaborated for designing the theoretical path diagram. Fabrice BONNET provided a clinical point of view on the work provided. Cécile PROUST-LIMA and Benoît LIQUET provided their expertise to validate adequate methods for data preprocessing and integrative analysis. Marie-Quitterie PICAT performed the statistical analysis of the data, under Rodolphe THIEBAUT supervision. All authors critically read the manuscript.

The ANRS CO3 Aquitaine cohort study group

Coordination: F. Bonnet, F. Dabis

Scientific committee: F. Bonnet, S. Bouchet, D. Breilh, G. Chêne, F. Dabis, M. Dupon, H. Fleury, V. Gaborieau, D. Lacoste, D. Malvy, P. Mercié, P. Morlat, D. Neau, I. Pellegrin, JL. Pellegrin, S. Reigadas, S. Tchamgoué

Epidemiology and Methodology: M. Bruyand, G. Chêne, F. Dabis, C. Fagard, S. Lawson-Ayayi, L. Richert, R. Thiébaud, L. Wittkop.

Infectious Diseases and Internal Medicine: K. André, F. Bonnet, N. Bernard, O. Caubet, L. Caunègre, C. Cazanave, J. Ceccaldi, I. Chossat, C. Courtault, FA. Dauchy, S. De Witte, D. Dondia, M. Dupon, A. Dupont, P. Duffau, H. Dutronc, S. Farbos, I. Faure, V. Gaborieau, Y. Gerard, C. Greib, M. Hessamfar-Joseph, Y. Imbert, D. Lacoste, P. Lataste, E. Lazaro, D. Malvy, J. Marie, M. Mechain, JP. Meraud, P. Mercié, E. Monlun, P. Morlat, D. Neau, A. Ochoa, JL. Pellegrin, M. Pillot-Debelleix, T. Pistone, I. Raymond, MC. Receveur, P. Rispal, L. Sorin, S. Tchamgoué, C. Valette, MA. Vandenhende, MO. Vareil, JF. Viillard, H. Wille, G. Wirth.

Immunology: JF. Moreau, I. Pellegrin.

Virology: H. Fleury, ME. Lafon, S. Reigadas, P. Trimoulet.

Pharmacology: S. Bouchet, D. Breilh, F. Haramburu, G. Miremont-Salamé

Data collection, Project Management and Statistical Analyses: MJ. Blaizeau, I. Crespel, M. Decoin, S. Delveaux, F. Diarra, C. D'Ivernois, C. Hanappier, D. Lacoste, S. Lawson-Ayayi, O. Leleux, F. Le Marec, E. Lenaud, J. Mourali, E. Pernot, A. Pougetoux, B. Uwamaliya-Nziyumvira, A. Tsaranazy, A. Valdes.

IT department and eCRF development: V. Conte, I. Louis, G. Palmer, V. Sapparrart, D. Touchard.

Centre for Biological Ressources-CHU de Bordeaux

C. Le Scanf-Terrien, I. Pellegrin.

Funding

This work was supported by Sidaction [AI20-2-01629], the CHU of Bordeaux [2009-AO1063-54] and the Vaccine Research Institute. The ANRS CO3 Aquitaine cohort was supported by the Agence nationale de recherches sur le sida et les hépatites virales [ANRS, Action Coordonnée no.7, Cohortes].

Conflicts of interest

The authors declare no conflicts of interest. All authors have completed the ICMJE Uniform Disclosure Form for Potential Conflicts of Interest.

References

1. Deeks SG, Kitchen CM, Liu L, Guo H, Gascon R, Narvaez AB, et al. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. *Blood* 2004 Aug 15;104(4):942-7.
2. Sooda DL, Silvestri G. Immune activation and AIDS pathogenesis. *AIDS* 2008 Feb 19;22(4):439-46.
3. Hsue PY, Deeks SG, Hunt PW. Immunologic basis of cardiovascular disease in HIV-infected adults. *J Infect Dis* 2012 Jun;205 Suppl 3:S375-82.
4. Plaeger SF, Collins BS, Musib R, Deeks SG, Read S, Embry A. Immune activation in the pathogenesis of treated chronic HIV disease: a workshop summary. *AIDS Res Human Retroviruses* May 2012;28(5):469-77.
5. Tebas P, Henry WK, Matining R, Weng-Cherng D, Schmitz J, Valdez H, et al. Metabolic and immune activation effects of treatment interruption in chronic HIV-1 infection: implications for cardiovascular risk. *PloS One* 2008;3(4):e2021.
6. Teijaro JR, Ng C, Lee AM, Sullivan BM, Sheehan KC, Welch M, et al. Persistent LCMV infection is controlled by blockade of type I interferon signaling. *Science* 2013 Apr 12;340(6129):207-11.
7. Wilson EB, Yamada DH, Elsaesser H, Herskovitz J, Deng J, Cheng G, et al. Blockade of chronic type I interferon signaling to control persistent LCMV infection. *Science* 2013 Apr 12;340(6129):202-7.
8. Zhou Y, Zhang Y, Moorman JP, Yao ZQ, Jia ZS. Viral (hepatitis C virus, hepatitis B virus, HIV) persistence and immune homeostasis. *Immunology* 2014 Nov;143(3):319-30.
9. Acchioni C, Marsili G, Perrotti E, Remoli AL, Sgarbanti M, Battistini A. Type I IFN--a blunt spear in fighting HIV-1 infection. *Cytokine Growth Factor Rev* 2015 Apr;26(2):143-58.
10. Cha L, Berry CM, Nolan D, Castley A, Fernandez S, French MA. Interferon-alpha, immune activation and immune dysfunction in treated HIV infection. *Clin Transl Immunology* 2014 Feb 28;3(2):e10.
11. Ries M, Pritschet K, Schmidt B. Blocking type I interferon production: a new therapeutic option to reduce the HIV-1-induced immune activation. *Clin Dev Immunol* 2012;2012:534929.
12. Rajasuriar R, Khoury G, Kamarulzaman A, French MA, Cameron PU, Lewin SR. Persistent immune activation in chronic HIV infection: do any interventions work? *AIDS* 2013 May 15;27(8):1199-208.

13. Appay V, Sauce D. Immune activation and inflammation in HIV-1 infection: causes and consequences. *J Pathol* 2008 Jan;214(2):231-41.
14. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med* 2006Dec;12(12):1365-71.
15. Klatt NR, Funderburg NT, Brenchley JM. Microbial translocation, immune activation, and HIV disease. *Trends Microbiol* 2013 Jan;21(1):6-13.
16. Jiang W, Lederman MM, Hunt P, Sieg SF, Haley K, Rodriguez B, et al. Plasma Levels of Bacterial DNA Correlate with Immune Activation and the Magnitude of Immune Restoration in Persons with Antiretroviral-Treated HIV Infection. *J Infect Dis* 1999(8):1177-85.
17. Appay V, Fastenackels S, Katlama C, Ait-Mohand H, Schneider L, Guihot A, et al. Old age and anti-cytomegalovirus immunity are associated with altered T-cell reconstitution in HIV-1-infected patients. *AIDS* 2011 Sep 24;25(15):1813-22.
18. Hunt PW, Martin JN, Sinclair E, Epling L, Teague J, Jacobson MA, et al. Valganciclovir reduces T cell activation in HIV-infected individuals with incomplete CD4+ T cell recovery on antiretroviral therapy. *J Infect Dis* 2011 May 15;203(10):1474-83.
19. Stratton R, Slapak G, Mahungu T, Kinloch-de Loes S. Autoimmunity and HIV. *Curr Opin Infect Dis* 2009 Feb;22(1):49-56.
20. Rawson PM, Molette C, Videtta M, Altieri L, Franceschini D, Donato T, et al. Cross-presentation of caspase-cleaved apoptotic self antigens in HIV infection. *Nat Med* 2007 Dec;13(12):1431-9.
21. Wittkop L, Bitard J, Lazaro E, Neau D, Bonnet F, Mercie P, et al. Effect of cytomegalovirus-induced immune response, self antigen-induced immune response, and microbial translocation on chronic immune activation in successfully treated HIV type 1-infected patients: the ANRS CO3 Aquitaine Cohort. *J Infect Dis* 2013 Feb 15;207(4):622-22.
22. Banchereau J, Pascual V. Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity* 2006 Sep;25(3):383-92.
23. Blanco P, Palucka AK, Pascual V, Banchereau J. Dendritic cells and cytokines in human inflammatory and autoimmune diseases. *Cytokine Growth Factor Rev* 2008 Feb;19(1):41-52.
24. Kinloch AJ, Chang A, Ko K, Henry Dunand CJ, Henderson S, Maienschein-Cline M, et al. Vimentin is a dominant target of in situ humoral immunity in human lupus tubulointerstitial nephritis. *Arthritis Rheumatol* 2014 Dec;66(12):3359-70.

25. Yang S, Netterwald J, Wang W, Zhu H. Characterization of the elements and proteins responsible for interferon-stimulated gene induction by human cytomegalovirus. *J Virol* 2005 Apr;79(8):5027-34.
26. Sanchez BN, Budtz-Jorgensen E, Ryan LM, and Hu H. Structural Equation Models: A Review With Applications to Environmental Epidemiology. *JASA* 2005;100(472):1443-55.
27. Thiebaut R, Morlat P, Jacqmin-Gadda H, Neau D, Mercie P, Dabis F, et al. Clinical progression of HIV-1 infection according to the viral response during the first year of antiretroviral treatment. Groupe d'Epidemiologie du SIDA en Aquitaine (GECSA). *AIDS* 2000 May 26;14(8):971-8.
28. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nature reviews Immunology* 2014 Jan;14(1):36-49.
29. Grace J B, Schoolmaster DR, Guntenspergen GR, Little AM, Mitchell BR, Miller KM, Schweiger EW. Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* 2012;3(8):73.
30. Chavance M, Escolano S, Romon M, Basdevant A, de Lauzon-Guillain B, Charles MA. Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology. *BMC Med Res Methodol* 2010 Apr 30;10:37.
31. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001 Dec;25(4):402-8.
32. Holst KK JE. Linear latent variable models: the lava-package. *Comput Stat* 2013;28:1385-452.
33. Hoyle RH. *Handbook of Structural Equation Modeling*. New York: Guildford Press; 2012.
34. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6(1):1-55.
35. Stone SF, Price P, Khan N, Moss PA, French MA. HIV patients on antiretroviral therapy have high frequencies of CD8 T cells specific for Immediate Early protein-1 of cytomegalovirus. *AIDS* 2005 Mar 24;19(6):555-62.
36. Abad-Fernandez M, Vallejo A, Hernandez-Novoa B, Diaz L, Gutierrez C, Madrid N, et al. Correlation between different methods to measure microbial translocation and its association with immune activation in long-term suppressed HIV-1-infected individuals. *J Acquir Immune Defic Syndr* 2013 Oct 1;64(2):149-53.

37. Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004Jul;4(7):499-511.
38. Kader M, Smith AP, Guiducci C, Wonderlich ER, Normolle D, Watkins SC, et al. Blocking TLR7- and TLR9-mediated IFN-alpha production by plasmacytoid dendritic cells does not diminish immune activation in early SIV infection. *PLoS Pathog* 2013;9(7):e1003530.
39. Meier A, Alter G, Frahm N, Sidhu H, Li B, Bagchi A, et al. MyD88-dependent immune activation mediated by human immunodeficiency virus type 1-encoded Toll-like receptor ligands. *J Virol* 2007 Aug;81(15):8180-91.
40. Schneider WM, Chevillotte MD, Rice CM. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* 2014;32:513-45.
41. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol* 2011 Dec;1(6):519-25.
42. Dunson DB. Bayesian methods for latent trait modelling of longitudinal data. *Stat Methods Med Res* 2007 Oct;16(5):399-415.

Tables

Table 1. Demographic, clinical and biological patients' characteristics. ACTHIV study.

Characteristics	Statistics*	
	All patients n = 191	HLA-A2 patients n = 87
Female	47 (25)	20 (23)
Age at inclusion	50 (43; 58)	50 (44; 60)
Risk factor		
Homo-bisexual – MSM	92 (48)	43 (49)
IDU/Blood products	1/1 (0.5/0.5)	1/0 (1.5/0)
Heterosexual	81 (42)	34 (39)
Others	16 (8)	9 (10.5)
CDC Stage		
A	101 (53)	40 (46)
B	48 (25)	29 (33)
C	42 (22)	18 (21)
Duration since first reported seropositivity, years	12.3 (8; 18)	12.5 (7.5; 19)
Duration of viral suppression, years	6.1 (4; 8.5)	5.8 (4.4; 7.7)
Duration of ART exposure, years	9.7 (6; 13)	10.5 (5.9; 13.4)
cART at baseline		
2NRTI+1NNRTI	65 (34)	24 (28)
2NRTI+1PI	100 (52)	48 (55)
Others	26 (14)	15 (17)

Table 1 (following). Demographic, clinical and biological patients' characteristics. ACTHIV study.

Characteristics	Statistics*	
	All patients n = 191	HLA-A2 patients n = 87
<i>CD4+ cells/μL</i>		
CD4+ nadir, cells/μL	188 (89; 275)	172 (47; 249)
CD4+, cells/μL	517 (387; 720)	441 (328; 660)
CD4+, %	42 (33; 50)	42 (33; 48)
% HLA-DR+/CD38+ among CD4+ T-cells	4.5 (3.0; 6.5)	4.6 (3.0; 6.3)
<i>CD8+ cells/μL</i>		
CD8+, cells/μL	676 (480; 977)	652 (478; 896)
% CD8+	53 (46; 62)	54 (48; 62)
% HLA-DR+/CD38+ among CD8+ T-cells	16.8 (11.1; 23.2)	15.8 (11.1; 22.6)
Regulatory T-cells/μL	38 (25; 51)	35 (22; 49)
B lymphocytes, cells/μL	197 (137; 309)	197 (129; 274)
Natural killer lymphocytes, cells/μL	80 (30; 168)	81 (27; 192)
<u>Cytomegalovirus-related variables</u>		
CMV Serology positive	177 (93)	79 (91)
CMV PCR positive (whole blood)	1 (0.5)	0 (0)
Quantiferon-CMV positive	148 (78)	71 (82)
Quantiferon-CMV values, IU/mL	3.3 (0.3 ; 9.8)	5.3 (0.9 ; 9.9)
CMV-pp65-ELISPOT†	377.5 (120.0; 689.0)	452 (120; 728)
% CMV-pp-65-specific-CD8+T-cells	NA¶	1.1 (0.1; 2.1)‡

Table 1 (following). Demographic, clinical and biological patients' characteristics. ACTHIV study.

Characteristics	Statistics*	
	All patients n = 191	HLA-A2 patients n = 87
<u>Autoimmune response related variables</u>		
Antinuclear antibody positive	43 (26)‡	21 (28)‡
% Actin-specific-CD8+ T-cells	NA	0.03 (0.02; 0.06)‡
% Vimentin-specific-CD8+ T-cells	NA	0.01 (0.01; 0.02)‡
% Lamin-specific-CD8+ T-cells	NA	0.01 (0.01; 0.02)‡
<u>Microbial translocation related variables</u>		
sCD14 µg/mL	1.9 (1.6; 2.2)	1.8 (1.6; 1.9)
16S rDNA log ₁₀ copies/mL	3.2 (3.0; 3.5)	3.3 (3.1; 3.4)

Legend: *for categorical variables n (%), for quantitative variables median (first quartile; third quartile) are presented. NRTI, nucleoside reverse-transcriptase inhibitor; NNRTI, non-nucleoside reverse-transcriptase inhibitor; PI, protease inhibitor; IDU, injection drug users; MSM, men who have sex with men; CDC, Centers for Disease Control and Prevention; cART, antiretroviral therapy combination.

† Median spot-forming cells / 2.5×10⁵ peripheral blood mononuclear cells. ‡ Less than 10% of missing values. ¶ Not applicable.

Table 2. Effects of cytomegalovirus-, autoimmune-induced immune responses on chronic immune activation, n=57 patients HLA-A*0201 positive patients. Main integrative analysis to read in conjunction with Figure 3. ACTHIV study.

Process	β std*	p value
<u>Measurement models (latent variable definitions)</u>		
Cytomegalovirus(CMV)-induced immune response latent variable		
CMV response latent variable → Quantiferon-CMV	0.61	< 0.0001
CMV response latent variable → CMV-pp-65-ELISPOT	0.51	< 0.0001
Autoimmune(AI)-induced immune response latent variable		
AI response latent variable → % actin-specific-CD8+	0.23	0.0862
AI response latent variable → % vimentin-specific-CD8+	0.52	0.0006
AI response latent variable → % lamin-specific-CD8+	0.39	0.0184
IFN-α-stimulated-genes latent variable		
IFN- α -stimulated-genes latent variable → ADAR	0.59	< 0.0001
IFN- α -stimulated-genes latent variable → ISG15	0.54	< 0.0001
IFN- α -stimulated-genes latent variable → IFIT1	0.94	< 0.0001
IFN- α -stimulated-genes latent variable → Mx1	0.93	< 0.0001
IFN- α -stimulated-genes latent variable → OAS1	0.77	< 0.0001
<u>Structural model</u>		
CMV response latent variable → IFN- α -stimulated-genes latent variable	0.02	0.9069
CMV response latent variable → Chronic immune activation	0.56	0.0004
AI response latent variable → IFN- α -stimulated-genes latent variable	0.36	0.0401
AI response latent variable → Chronic immune activation	-0.10	0.3169
IFN- α -stimulated-genes latent variable → Chronic immune activation	0.39	0.0044

Legend: *Standardized estimates

Figure legends

Figure 1. Theoretical hypothesized path diagram of effects of cytomegalovirus-induced-, autoimmune-induced immune responses and microbial translocation on chronic immune activation (model adjusted on age, CD4+ T-cell count and regulatory T-cell count).

Legend: Boxes: observed variables; ellipses: latent variables; ε : measurement errors.

Figure 2. Spearman correlations between percentages of CD8+CD38+HLA-DR+ T-cells and IFN- α gene expression measurements. n=57 patients HLA-A*0201 positive patients. ACTHIV study.

Figure 3. Effects of cytomegalovirus-induced-, autoimmune-induced immune responses latent variables on chronic immune activation (final model). Figure to read in conjunction with Table 2. ACTHIV study.

*Legend: Boxes: observed variables; ellipses: latent variables; measurement errors deleted for clarity. *p-values < 0.05. **p-values < 0.0001.*

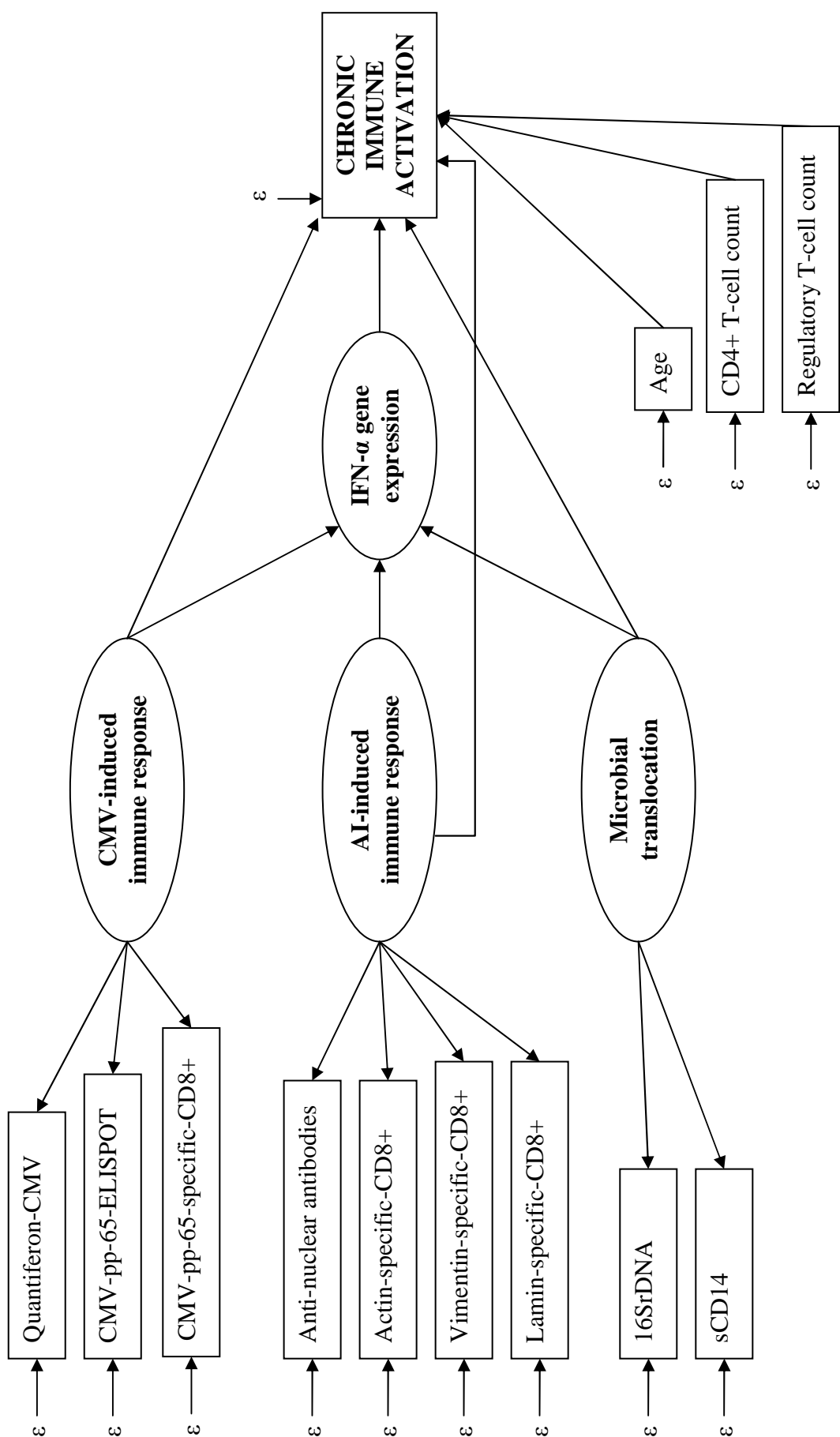
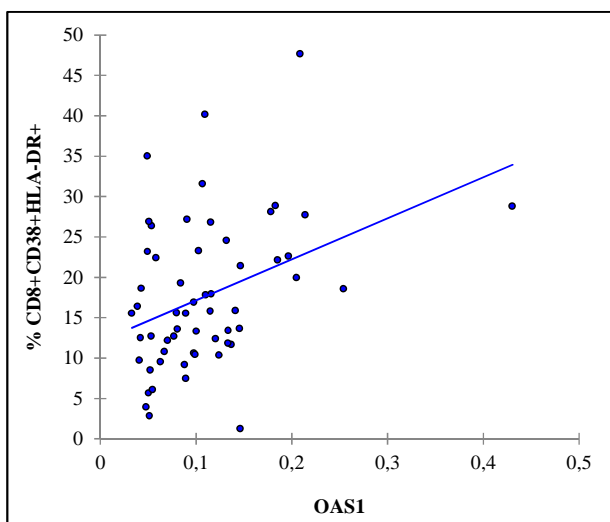
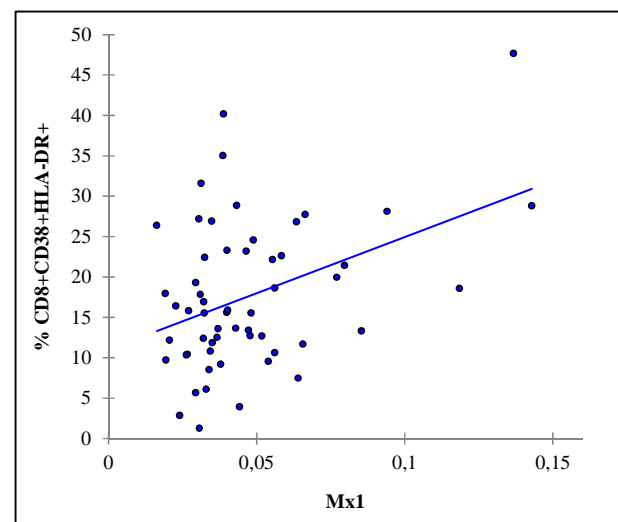
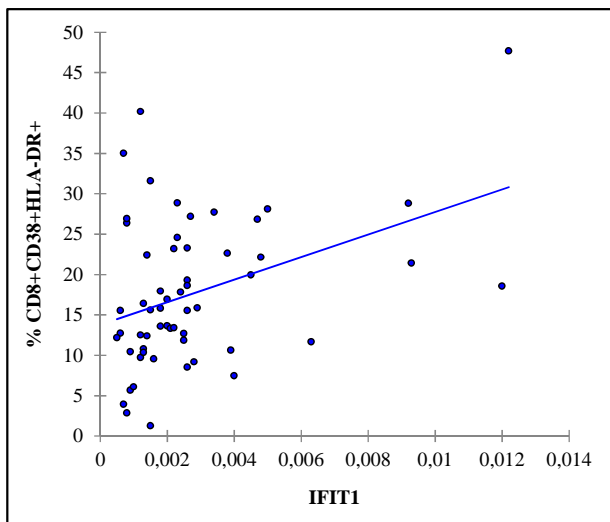
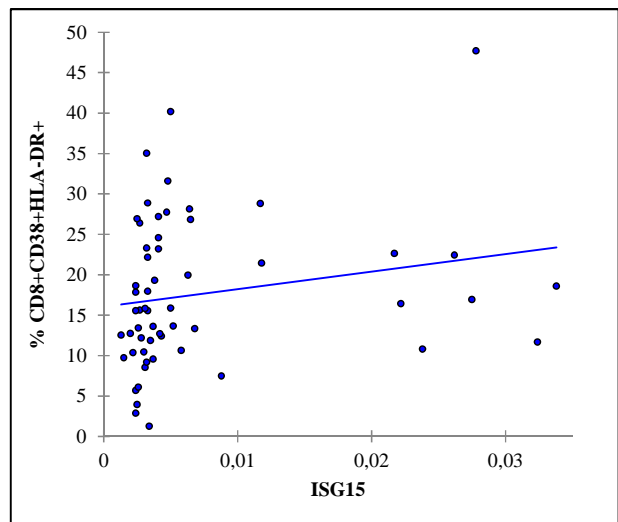
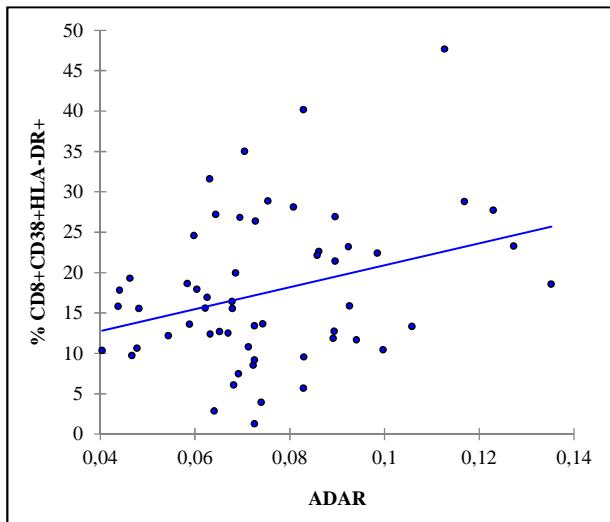


Figure 1.



Spearman correlation coefficients with % CD8+CD38+HLA-DR+

ADAR	r=0.31, p-value=0.0180
ISG15	r=0.34, p-value=0.0085
IFIT1	r=0.32, p-value=0.0166
Mx1	r=0.29, p-value=0.0301
OAS1	r=0.35, p-value=0.0077

Figure 2.

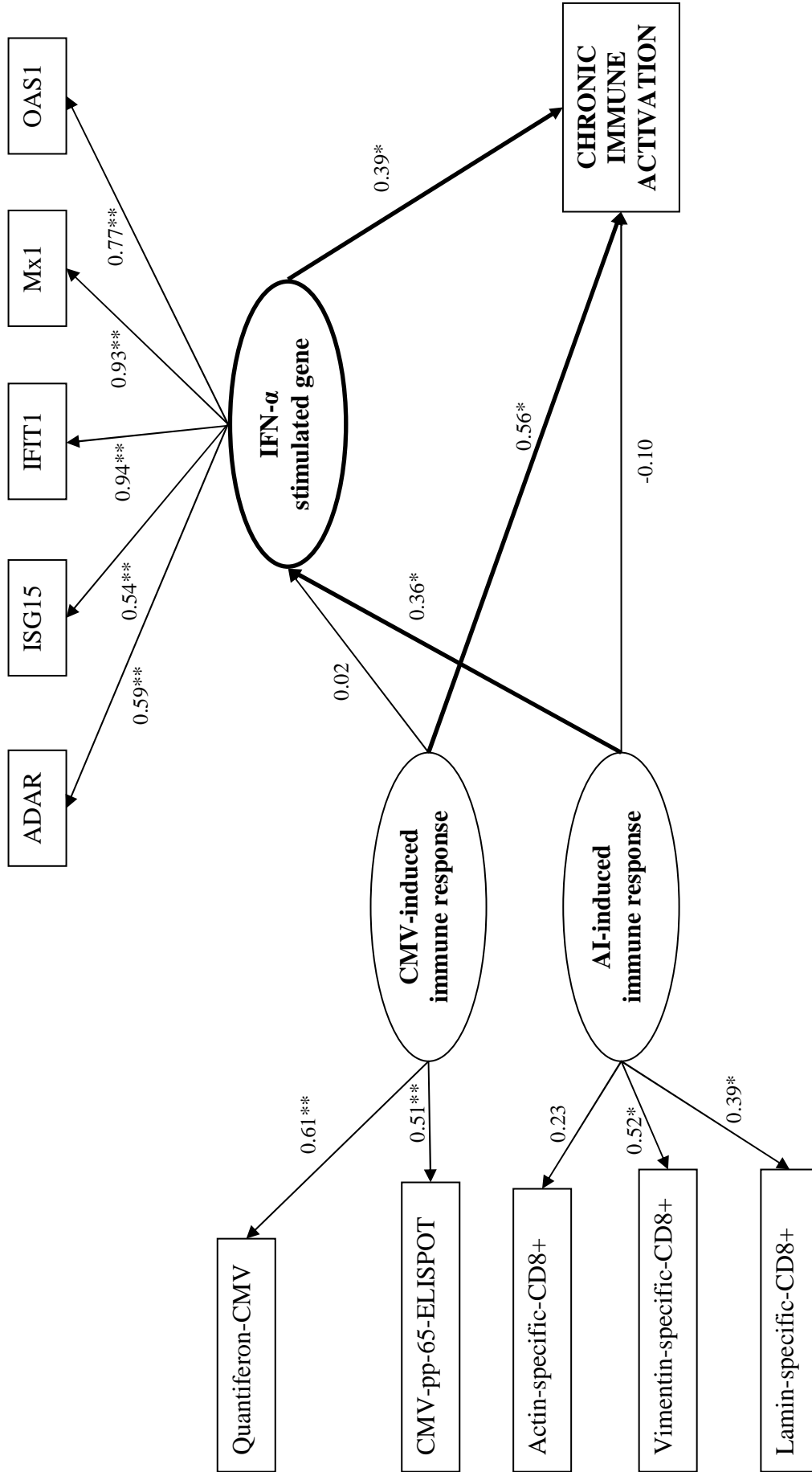


Figure 3.

4.5. Discussion

Dans le cadre de ce travail, l'analyse principale a été effectuée chez 57 patients HLA-A*0201 positifs, en raison de la disponibilité des pourcentages de lymphocytes T CD8+ spécifiques des antigènes issus de produits apoptotiques pour mieux caractériser la réponse immunitaire induite par l'auto-immunité (tests restreints aux patients HLA-A*0201 positifs).

Dans les étapes de sélection vers l'élaboration d'un modèle final, les variables observées qui ne contribuaient pas à la définition des variables latentes dans les modèles de mesure n'ont pas été conservées pour la modélisation ultérieure, soit les anticorps anti-noyaux ($p=0,2785$) et les lymphocytes T CD8+ pp65 spécifiques du cytomégalovirus ($p=0,7049$). Les marqueurs séparés de translocation microbienne et la variable latente « translocation microbienne » n'étant pas associés à l'activation immunitaire chronique, ils n'ont pas été conservés pour la modélisation ultérieure.

Les résultats du modèle final incluant l'ensemble des processus sont présentés ci-après et sont synthétisés dans le tableau 2 et la figure 3 de l'article section 4.4.1. Les modèles SEM ont révélé un effet direct de la variable latente « réponse immunitaire induite par le cytomégalovirus » sur l'activation immunitaire chronique ($\beta_{std}=0,56$, $p=0,0004$) mais pas d'effet sur la variable latente interféron alpha ($p=0,9069$). On ne retrouvait pas d'effet direct de la variable latente « réponse immunitaire induite par l'auto-immunité » sur l'activation immunitaire chronique ($p=0,3169$). L'effet de la variable latente « réponse immunitaire induite par l'auto-immunité » sur l'activation immunitaire chronique était indirect via la variable latente « interféron alpha » (effet de la variable latente « réponse immunitaire induite par l'auto-immunité » sur la variable latente « interféron alpha » : $\beta_{std}=0,36$, $p=0,0401$; effet de la variable latente « interféron alpha » sur l'activation immunitaire chronique : $\beta_{std}=0,39$, $p=0,0044$). Les résultats étaient robustes chez 142 patients pour lesquels les réponses immunitaires induites par le cytomégalovirus et l'auto-immunité étaient mesurées par la positivité du test Quantiféron et les anticorps anti-noyaux respectivement.

Notre étude a ainsi confirmé le rôle de la réponse immunitaire induite par le cytomégalovirus dans l'activation immunitaire chronique des patients infectés par le VIH traités efficacement, encourageant ainsi les interventions contre le cytomégalovirus, ainsi que le rôle de l'interféron alpha. Un effet indirect de la réponse immunitaire induite par l'auto-immunité sur l'activation immunitaire chronique a été révélé via la voie de l'interféron alpha.

Plusieurs points méritent d'être soulignés quant à la réflexion menée autour de l'application des modèles SEM comme analyse intégrative autour de l'exemple concret sur l'activation immunitaire chronique.

D'un point de vue méthodologique, contrairement aux modèles de régression qui ne considèrent que des associations indépendantes entre les variables explicatives et la variable à expliquer, les modèles SEM présentent l'avantage d'utiliser des variables observées ou latentes, mais aussi de modéliser des relations directes ou indirectes, avec des variables pouvant être d'une part des variables à expliquer et d'autre part des variables explicatives [VanderWeele, 2012]. De plus, l'écriture de ces modèles est sous-tendue par un modèle sous-jacent de relations supposées entre variables, l'absence de flèche entre deux variables sur le diagramme de chemin devant être considérée comme une hypothèse en elle-même. Une autre différence notable entre les modèles de régression et les modèles SEM est la prise en compte de termes d'erreurs multiples sur les marqueurs considérés (voir écriture du modèle SEM complet section 4.3.4). Les modèles SEM permettent ainsi de modéliser une erreur de mesure sur les variables explicatives alors que dans un modèle de régression classique, seul un terme d'erreur sur la variable à expliquer est estimé. Ceci présente l'avantage de pouvoir s'affranchir du biais de dilution lié à l'effet des erreurs de mesure sur les variables explicatives, ce phénomène étant connu pour rendre les coefficients de régression instables et les faire tendre vers 0 [Hutcheon et al., 2010 ; Liu, 1988]. Néanmoins, si les modèles SEM présentent de nombreux avantages, l'interprétation de ces modèles doit être faite avec précaution. Un modèle SEM doit être interprété comme un modèle dont les hypothèses sont cohérentes avec les données et non comme la preuve de la véracité des mécanismes testés, un certain nombre de modèles alternatifs étant susceptibles de fournir des indices de qualité aussi bons que le modèle considéré comme final. En parallèle, l'adéquation du modèle aux données doit reposer sur la validation explicite des différents index référencés dans les SEM, notamment ceux influencés par la taille de l'échantillon, particulièrement quand le modèle testé est complexe. Dans le cadre de l'étude ACTHIV, les modèles SEM ont été réalisés pour les patients pour lesquels l'ensemble des données pour l'analyse intégrative étaient renseignées. En particulier, les 57 patients considérés pour le modèle final étaient représentatifs des 87 patients HLA-A*0201 positifs de notre échantillon. Au total, dans notre étude, l'utilisation des modèles SEM sur le jeu de données préalablement analysé [Wittkop et al., 2013], a permis, en plus de confirmer les résultats identifiés sur le cytomégalovirus via des méthodes classiques de régression, de confirmer également le rôle majeur de l'interféron alpha et de mettre en évidence un effet indirect de la réponse immunitaire induite par l'auto-

immunité sur l'activation immunitaire chronique, via la voie de l'interféron alpha. Le signal identifié sur l'auto-immunité, malgré des valeurs faibles de pourcentages de lymphocytes T CD8+ spécifiques des antigènes issus de produits apoptotiques (actine, vimentine, lamine), souligne en particulier l'apport des variables latentes pour prendre en compte l'ensemble des mesures disponibles pour étudier ce processus. De plus, les anticorps anti-noyaux n'étaient pas contributifs pour la définition de la variable latente « réponse immunitaire induite par l'auto-immunité », soulignant la nécessité de prendre en compte des marqueurs d'auto-immunité plus pertinents dans l'étude du VIH chronique [Rawson et al., 2007]. Ainsi, même en l'absence de signes cliniques d'auto-immunité, des valeurs biologiques de marqueurs de réponse immunitaire induite par l'auto-immunité, même faibles, pourraient contribuer à l'activation immunitaire chronique dans le VIH. Ce signal identifié sur l'auto-immunité mérite des recherches futures, pour mieux comprendre le rôle de l'auto-immunité sur l'activation immunitaire chronique. Néanmoins, très peu d'études de la littérature ont étudié l'auto-immunité dans le VIH [Stratton et al., 2009 ; Rawson et al., 2007]. A notre connaissance, ce résultat potentiel sur l'effet indirect de la réponse immunitaire induite par l'auto-immunité sur l'activation immunitaire chronique, via la voie de l'interféron alpha, n'a pas été rapporté dans la littérature dans le VIH, c'est pourquoi il nous est apparu important de le rapporter à la communauté scientifique. C'est ce qui a motivé en particulier notre choix de soumettre ce travail à un Journal scientifique clinique tel que le Journal AIDS.

Bien que nos résultats nécessitent des explorations ultérieures, l'analyse intégrative par modèles SEM a permis une meilleure compréhension des déterminants de l'activation immunitaire chronique, compréhension indispensable pour mieux guider les stratégies actuellement discutées pour réduire l'activation immunitaire chronique chez les patients atteints de VIH traités efficacement [Martinson et al., 2010 ; Rajasuriar et al., 2013]. Une des limites de l'étude est son caractère transversal ne permettant pas d'inférence causale. Plus spécifiquement, une étude longitudinale permettrait peut-être de montrer l'effet de l'auto-immunité sur la mobilisation de la voie interféron alpha suivie de l'activation des lymphocytes T CD8+. Cependant, le schéma d'étude nécessaire pour faire cette observation et notamment l'intensité de répétition des mesures n'est pas trivial.

4.6. Conclusion et perspectives

Les modèles d'équations structurelles constituent un outil précieux pour les épidémiologistes pour étudier des relations complexes entre variables et prendre en compte des processus à

plusieurs dimensions. L'extension de l'utilisation de ces modèles dans le cadre d'analyse de données longitudinales [Dunson, 2007] ouvre des perspectives pour l'approche d'inférence causale. La mise en œuvre de ces méthodes est néanmoins complexe et nécessite l'appropriation des étapes pas à pas qui concourent à la construction d'un modèle pertinent [Tu, 2009]. L'application de ces modèles trouve particulièrement son sens pour l'étude multifactorielle des troubles immunologiques des maladies du système immunitaire et à un sens plus large, dans le champ de l'Immunologie du fait de l'existence de nombreux processus mesurés par de multiples marqueurs.

5. ETUDE DE LA RECONSTITUTION IMMUNITAIRE A LONG TERME, APRES MISE SOUS TRAITEMENTS ANTIRETROVIRAUX, CHEZ LES ENFANTS INFECTES PAR LE VIH DANS LES PAYS A RESSOURCES LIMITEES

5.1. Problématique

L'infection par le VIH est responsable d'une atteinte du système immunitaire, ciblant en particulier des lymphocytes T CD4+. Chez les patients non traités, la destruction du système immunitaire occasionne, au cours du temps, une diminution du nombre de lymphocytes T CD4+ (cellules/mm³) et une augmentation de la charge virale plasmatique, entraînant la survenue potentielle de pathologies dites opportunistes dont les plus graves définissent le stade SIDA. L'avènement des traitements antirétroviraux hautement actifs a significativement diminué la morbi-mortalité liée au VIH dans les pays où l'accès aux soins est généralisé [UNAIDS, 2015]. L'efficacité des traitements antirétroviraux est basée sur la chute de la charge virale et, en conséquence, la reconstitution immunitaire du nombre de lymphocytes T CD4+ vers des valeurs attendues les plus proches de celles d'un sujet sain. Néanmoins, dans les pays à ressources limitées, seulement 33 % des enfants et 42 % des adultes infectés par le VIH nécessitant un traitement y ont accès en 2014 [UNAIDS, 2015].

En 2014, le nombre d'enfants de moins de 15 ans infectés par le VIH étaient de 2,6 millions dans le monde, 88 % vivant en Afrique sub-saharienne [UNAIDS, 2015]. La décision d'initier un traitement antirétroviral chez ces enfants est majoritairement basée sur les recommandations de l'Organisation Mondiale de la Santé (OMS) qui définissent, selon des âges distincts, des seuils de lymphocytes T CD4+ en dessous desquels la mise en place d'un traitement est recommandée pour diminuer le risque à court terme de progression de la maladie et de mortalité [WHO, 2010]. Ces recommandations sont présentées dans le tableau 2 ci-après.

Tableau 2. Recommandations OMS 2010 pour initier une mise sous traitements antirétroviraux chez l'enfant.

Age des enfants	Moins de 2 ans	De 2 à 5 ans	De plus de 5 ans
Nombre de lymphocytes T CD4+ en cellules/mm ³	Initiation d'un traitement antirétroviraux*	≤ 750	≤ 350

* Tous les enfants de moins de 2 ans avec un diagnostic de VIH devraient bénéficier d'une mise sous traitement en raison d'un risque accru de progression à court terme du VIH.

Ainsi, la mise sous traitements antirétroviraux des enfants ayant un nombre de lymphocytes T CD4+ au-dessus des seuils recommandés pourrait être différée, la finalité étant de réduire les risques de toxicité à long terme des médicaments et la résistance virale aux traitements [Lewis et al., 2012]. Par conséquent, les recommandations actuelles ne prennent pas en compte que différer la mise sous traitements antirétroviraux pourrait avoir des conséquences sur la reconstitution à long terme du système immunitaire. En particulier, cette question apparaît majeure avec l'efficacité accrue des traitements antirétroviraux, leur diversité et leur acceptabilité, ayant pour conséquence une survie augmentée des enfants infectés par le VIH vers l'âge adulte.

Dans ce contexte, le défi auquel sont désormais confrontés les praticiens pour la prise en charge de ces enfants est d'optimiser l'initiation de la mise sous traitements antirétroviraux pour diminuer, non seulement le risque à court terme de progression de la maladie, mais également permettre une reconstitution optimale de leur système immunitaire sur le long terme. Les études qui se sont intéressées à la reconstitution du système immunitaire dans le VIH pédiatrique dans les pays à ressources limitées se sont plus particulièrement intéressées au risque à court terme de progression de la maladie (< 2 ans) [Sauvageot et al., 2010 ; Sutcliffe et al., 2008]. Celles qui se sont intéressées à la reconstitution à long terme sous traitement sont rares et ont inclus un nombre limité d'enfants (< 200 enfants) [De Beudrap et al., 2008]. Ainsi, l'amélioration des connaissances sur la reconstitution immunitaire à long terme, après mise sous traitements antirétroviraux, chez les enfants infectés par le VIH dans les pays à ressources limitées, requiert une grande considération, notamment dans le contexte des questions posées dans l'essai clinique START (Strategic Timing of Antiretroviral Treatment), chez l'adulte, d'initier les traitements antirétroviraux mêmes si le nombre de lymphocytes T CD4+ est supérieur aux seuils préconisés par les recommandations [Babiker et al., 2013].

5.2. Modélisation longitudinale de la reconstitution immunitaire dans le VIH : exemple de l'adulte

Les mesures répétées des marqueurs immuno-virologiques tels que le nombre de lymphocytes T CD4+ ou la charge virale plasmatique permettent d'évaluer l'efficacité des traitements au cours du temps. Les mesures peuvent être réalisées aux mêmes dates pour tous les sujets ou à des dates différentes. Le nombre de mesures peut être variable d'un sujet à l'autre. La

modélisation de la reconstitution immunitaire dans le VIH est ainsi basée sur l'analyse de données répétées, notamment du nombre de lymphocytes T CD4+.

L'objectif d'une telle analyse est de décrire l'évolution du nombre de lymphocytes T CD4+ au cours du temps et d'identifier les facteurs qui influencent cette évolution. Cette analyse ne peut être effectuée avec les méthodes statistiques classiques car les mesures réalisées sur les mêmes sujets à des temps différents ne sont pas indépendantes, la répétition des mesures d'une variable sur les mêmes sujets induisant une corrélation intra-sujet. Les modèles qui permettent de prendre en compte de telles données sont les modèles mixtes [Diggle et al., 2002 ; Laird et Ware, 1982]. Ces modèles sont composés de deux parties :

- les effets fixes, identiques pour tous les sujets de la population, représentant la tendance moyenne pour la population,
- les effets aléatoires représentant l'écart de chaque individu par rapport à la tendance de la population.

Ils permettent de prendre en compte plusieurs sources de variabilité : i) la variabilité inter-individuelle (effets aléatoires), ii) la variabilité intra-individuelle (erreur autocorrélée), iii) l'erreur de mesure (erreur indépendante). En parallèle, l'avantage de ces modèles est de pouvoir utiliser en une seule analyse toute l'information disponible au cours des suivis.

Le modèle de reconstitution immunitaire sous traitements antirétroviraux le plus étudié dans le VIH est celui de l'adulte, modélisé à partir de modèles linéaires mixtes [Thiébaud et al., 2002 ; Drylewicz et al., 2008 ; Balestre et al., 2012]. En particulier, la reconstitution immunitaire des lymphocytes T CD4+ après mise sous traitements antirétroviraux est connue pour être biphasique chez l'adulte, liée à une augmentation rapide des lymphocytes T mémoires suivie par une augmentation plus lente des lymphocytes T naïfs [Pakker et al., 1998]. La modélisation est ainsi réalisée le plus souvent par des modèles linéaires mixtes à deux pentes [Le Moing et al., 2002 ; Thiébaud et al., 2006 ; Thiébaud et al., 2005].

5.3. Hypothèse d'une reconstitution immunitaire asymptotique chez l'enfant

Bien que des études incluant un nombre considérable de patients se soient intéressées aux effets des traitements antirétroviraux sur la reconstitution à long terme du système immunitaire chez les adultes vivant dans des pays à ressources limitées [Achhra et al., 2011 ; Mugenyi et al. ; 2010], la reconstitution du système immunitaire diffère significativement chez l'enfant [Gibb et al., 2000 ; Resino et al., 2006], en raison d'une thymopoïèse plus active

[Bains et al., 2009 ; Steinmann et al., 1985]. Par conséquent, les conclusions formulées chez l'adulte n'apparaissent pas applicables pour l'enfant. En particulier, la reconstitution du compartiment lymphocytaire T sous traitement, sa vitesse, sa cinétique et les proportions relatives des lymphocytes T naïfs et mémoires différent.

Chez l'enfant, la reconstitution biphasique du compartiment lymphocytaire T sous traitement est moins prononcée. Cette reconstitution sous traitement s'initie d'emblée grâce à une augmentation rapide des lymphocytes T naïfs, du fait de la thymopoïèse [Gibb et al., 2000 ; De Rossi et al., 2002]. De plus, un âge plus jeune a été rapporté comme associé à une augmentation plus rapide des lymphocytes T CD4+ après mise sous traitement [Sabin et al., 2008].

A la vue de ces éléments, une modélisation biphasique de la reconstitution immunitaire sous traitement, par modèles linéaires mixtes, telle qu'utilisée chez l'adulte avec un modèle à changement de pente, n'apparaît dès lors pas optimale chez l'enfant, dont le profil de reconstitution immunitaire sous traitement s'apparente plus à une reconstitution asymptotique sur le long terme. Pour modéliser l'évolution du nombre de lymphocytes T CD4+ en fonction du temps chez l'enfant, plusieurs modèles, qui sont des extensions des modèles linéaires mixtes, auraient pu être utilisés, permettant une modélisation plus souple de l'évolution des lymphocytes T CD4+ en fonction du temps, comme par exemple des modèles utilisant une fonction polynomiale ou des splines [Jacqmin-Gadda et al., 2002]. Néanmoins, il était nécessaire d'avoir un modèle dont les paramètres seraient aisément interprétables en termes cliniques et biologiques. En plus de répondre à ces critères, l'utilisation d'une fonction exponentielle, solution d'un système d'équations différentielles ordinaires (suite à des hypothèses qui doivent être explicitées), offrait l'avantage de pouvoir établir des liens ultérieurs avec une approche plus mécanistique. Pour exemple, le modèle bi-exponentiel proposé par Perelson AS et al [Perelson et al., 1996], publié dans Science en 1996, est une solution d'un système d'équations différentielles décrivant la dynamique de la charge virale dans le VIH. En particulier, dans le modèle mécanistique proposé, l'estimation des paramètres, à partir de données observées de la charge virale en réponse à un inhibiteur de protéase, a apporté des connaissances considérables dans la pathogenèse du VIH, en révélant des demi-vies courtes des cellules infectées et du virion. Cela signifiait une réplication virale intense, associée à une probabilité de mutations très élevée, ce qui expliquait l'échappement aisé du virus aux monothérapies antirétrovirales. On peut également noter que la forme exponentielle est également utilisée pour les modèles pharmacocinétiques [Rochetti et al., 2013 ; Simeoni et al., 2004], la structure mathématique de ces modèles étant le plus souvent

un système d'équations différentielles représentant les variations de chaque acteur dans le temps et/ou dans l'espace [Torneo et al., 2004].

Ainsi, sous l'hypothèse d'une reconstitution immunitaire non linéaire des lymphocytes T CD4+, après mise sous traitements antirétroviraux, tendant vers un plateau stable sur le long terme (asymptote), nous avons proposé un modèle de reconstitution immunitaire asymptotique utilisant une fonction exponentielle.

Pour un enfant i au temps t_{ij} après l'initiation du traitement, le modèle s'écrit :

$$\ln(CD4_{forage_{ij}}) = asy_i - (asy_i - int_i) \exp(-c_i t_{ij})$$

Dans ce modèle, les paramètres spécifiques à chaque enfant ont clairement une pertinence clinique et biologique :

- l'intercept (int_i) correspond au nombre de lymphocytes T CD4+ ajusté sur l'âge (transformation pour normalisation), au moment de l'initiation du traitement,
- l'asymptote (asy_i) correspond au nombre de lymphocytes T CD4+ ajusté sur l'âge, prédit sur le long terme,
- le paramètre c_i correspond au temps jusqu'à reconstitution, avec $\ln(2)/c$ étant la moitié du temps nécessaire pour augmenter le nombre de lymphocytes T CD4+ ajusté sur l'âge, de l'intercept à l'asymptote.

Les paramètres estimés spécifiques à chaque enfant incluent les effets aléatoires représentant l'écart de chaque enfant par rapport à la tendance de la population.

Une représentation schématique du modèle est présentée en figure 4 ci-après.

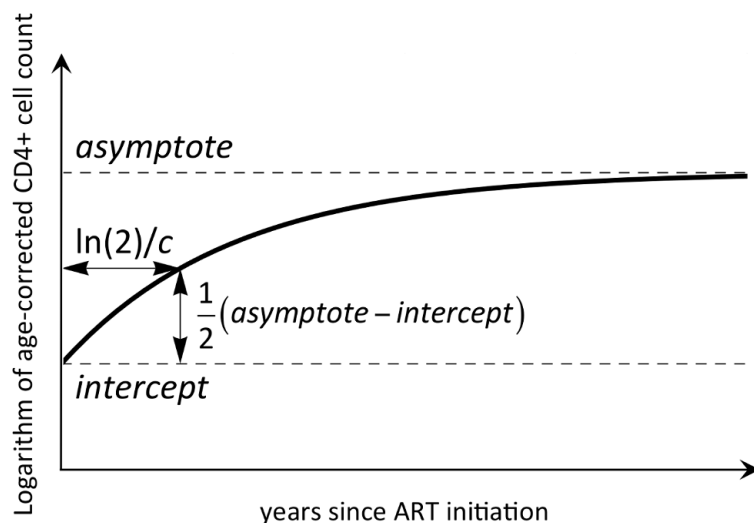


Figure 4. Représentation schématique du modèle mathématique utilisé pour étudier la reconstitution immunitaire des lymphocytes T CD4+ sur le long terme après mise sous traitements antirétroviraux chez l'enfant.

Ce modèle a été appliqué pour étudier la reconstitution immunitaire des lymphocytes T CD4+ sur le long terme, après mise sous traitements antirétroviraux, chez les enfants infectés par le VIH en Afrique sub-saharienne (section 5.4). Une discussion de ces travaux est proposée en section 5.5.

5.4. Application

Dans le cadre de la mobilité à l'Institute of Child Health, University College London et de la collaboration avec la Medical Research Council Clinical Trials Unit de Londres, les données de l'essai clinique ARROW ont été analysées (Antiretroviral Research for Watoto trial, <http://www.arrowtrial.org/>; ISRCTN 24791884) [Kekitiinwa et al., 2013]. L'objectif de cet essai clinique pédiatrique randomisé, ouvert, avec 5 ans de suivi, était de comparer deux stratégies de monitoring après mise sous traitements antirétroviraux chez 1200 enfants infectés par le VIH en Uganda et Zimbabwe et d'évaluer si l'initiation de 4 molécules antirétrovirales sur une courte période était supérieure sur le long terme à une initiation à 3 molécules. Le diagramme de flux du protocole ARROW, illustrant les critères d'éligibilité et les stratégies de randomisation, est présenté en figure 5 ci-après.

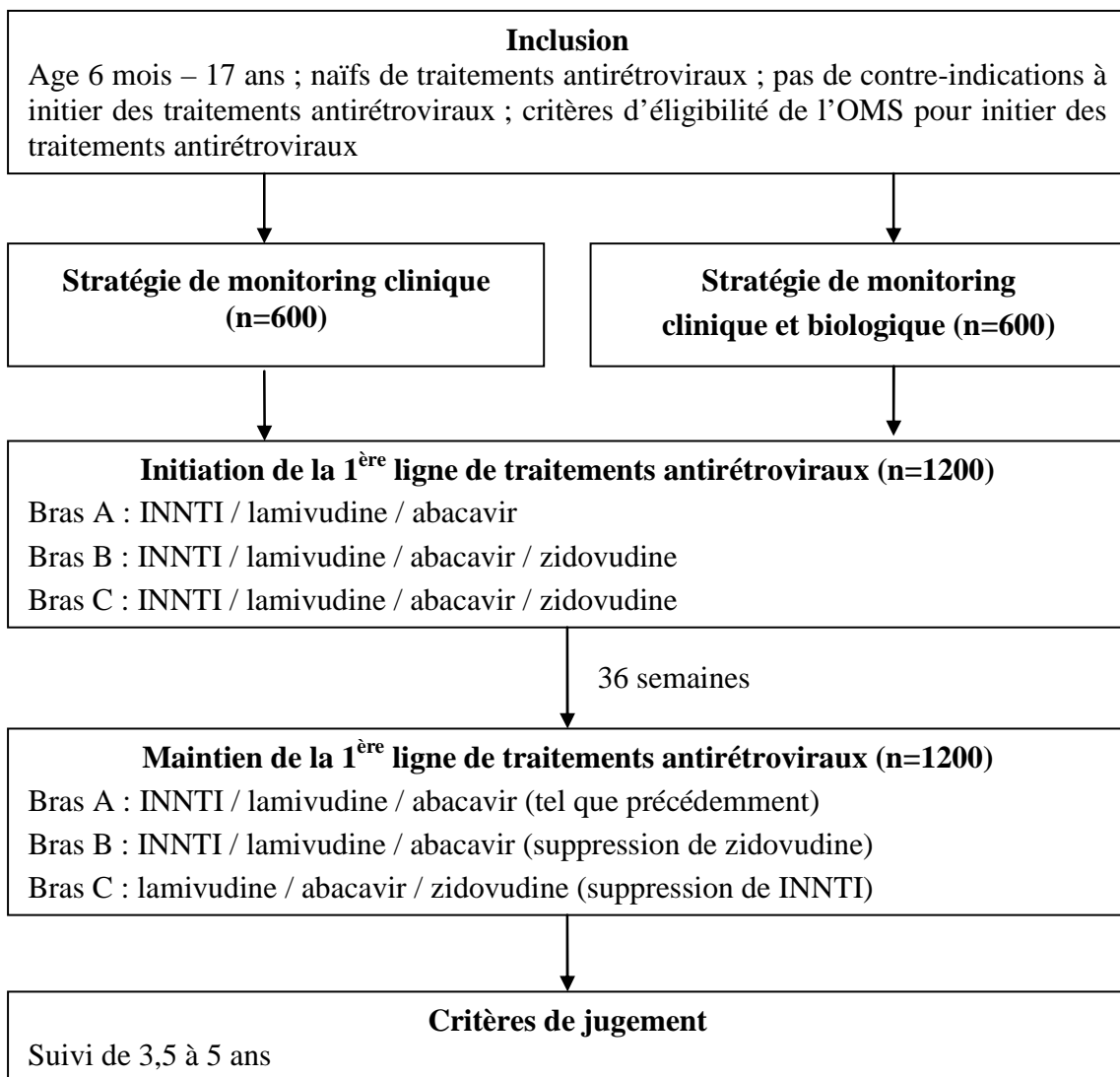


Figure 5. Diagramme de flux du protocole ARROW.

INNTI : inhibiteur non nucléosidique de la transcriptase inverse

A partir des données longitudinales disponibles de cet essai clinique, nous avons mené une analyse exploratoire, non spécifiée initialement au protocole, dont l'objectif était d'identifier et mieux comprendre des profils de reconstitution immunitaire des lymphocytes T CD4+ sur le long terme, après mise sous traitements antirétroviraux (cohort-based modelling study). L'utilisation de modèles non linéaires mixtes nous a permis d'appliquer le modèle de reconstitution asymptotique présenté, pour investiguer l'effet de l'âge, du nombre de lymphocytes T CD4+ et d'autres covariables au moment de la mise sous traitements antirétroviraux sur la reconstitution immunitaire à long terme.

Les paramètres estimés ont ensuite permis de représenter l'effet d'une mise sous traitements antirétroviraux à des âges distincts sur le profil de reconstitution attendu sur le long terme et la mise en parallèle avec les recommandations d'initiation de traitements antirétroviraux formulées par l'OMS.

5.4.1. Article publié dans Plos Medicine

Ce travail a donné lieu à un article original, publié dans Plos Medicine et présenté ci-après.

Predicting Patterns of Long-Term CD4 Reconstitution in HIV-Infected Children Starting Antiretroviral Therapy in Sub-Saharan Africa: A Cohort-Based Modelling Study

Marie-Quitterie Picat^{1,2,3}, Joanna Lewis^{1,4*}, Victor Musiime⁵, Andrew Prendergast^{6,7}, Kusum Nathoo⁸, Addy Kekitiinwa⁹, Patricia Nahirya Ntege¹⁰, Diana M. Gibb⁶, Rodolphe Thiebaut^{2,3}, A. Sarah Walker^{1,6}, Nigel Klein^{1,1}, Robin Callard^{1,4,†}, the ARROW Trial Team[‡]

1 Institute of Child Health, University College London, London, United Kingdom, **2** Institut de Santé Publique, d'Épidémiologie et de Développement, Centre Inserm U897—Épidémiologie-Biostatistique, University of Bordeaux, Bordeaux, France, **3** Department of Medical Information, Bordeaux University Hospital, Bordeaux, France, **4** Centre for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London, United Kingdom, **5** Joint Clinical Research Centre, Kampala, Uganda, **6** MRC Clinical Trials Unit, Medical Research Council, London, United Kingdom, **7** Centre for Paediatrics, Blizard Institute, Queen Mary University of London, London, United Kingdom, **8** University of Zimbabwe Medical School, University of Zimbabwe, Harare, Zimbabwe, **9** Paediatric Infectious Diseases Centre, Mulago, Uganda, **10** MRC/UVRI Uganda Research Unit on AIDS, Uganda Virus Research Institute, Entebbe, Uganda

Abstract

Background: Long-term immune reconstitution on antiretroviral therapy (ART) has important implications for HIV-infected children, who increasingly survive into adulthood. Children's response to ART differs from adults', and better descriptive and predictive models of reconstitution are needed to guide policy and direct research. We present statistical models characterising, qualitatively and quantitatively, patterns of long-term CD4 recovery.

Methods and Findings: CD4 counts every 12 wk over a median (interquartile range) of 4.0 (3.7, 4.4) y in 1,206 HIV-infected children, aged 0.4–17.6 y, starting ART in the Antiretroviral Research for Watoto trial (ISRCTN 24791884) were analysed in an exploratory analysis supplementary to the trial's pre-specified outcomes. Most ($n=914$; 76%) children's CD4 counts rose quickly on ART to a constant age-corrected level. Using nonlinear mixed-effects models, higher long-term CD4 counts were predicted for children starting ART younger, and with higher CD4 counts ($p<0.001$). These results suggest that current World Health Organization–recommended CD4 thresholds for starting ART in children ≥ 5 y will result in lower CD4 counts in older children when they become adults, such that vertically infected children who remain ART-naïve beyond 10 y of age are unlikely ever to normalise CD4 count, regardless of CD4 count at ART initiation. CD4 profiles with four qualitatively distinct reconstitution patterns were seen in the remaining 292 (24%) children. Study limitations included incomplete viral load data, and that the uncertainty in allocating children to distinct reconstitution groups was not modelled.

Conclusions: Although younger ART-naïve children are at high risk of disease progression, they have good potential for achieving high CD4 counts on ART in later life provided ART is initiated following current World Health Organization (WHO), Paediatric European Network for Treatment of AIDS, or US Centers for Disease Control and Prevention guidelines. In contrast, to maximise CD4 reconstitution in treatment-naïve children >10 y, ART should ideally be considered even if there is a low risk of immediate disease progression. Further exploration of the immunological mechanisms for these CD4 recovery profiles should help guide management of paediatric HIV infection and optimise children's immunological development.

Please see later in the article for the Editors' Summary.

Citation: Picat M-Q, Lewis J, Musiime V, Prendergast A, Nathoo K, et al. (2013) Predicting Patterns of Long-Term CD4 Reconstitution in HIV-Infected Children Starting Antiretroviral Therapy in Sub-Saharan Africa: A Cohort-Based Modelling Study. PLoS Med 10(10): e1001542. doi:10.1371/journal.pmed.1001542

Academic Editor: Steven G. Deeks, San Francisco General Hospital, United States of America

Received: July 17, 2012; **Accepted:** September 12, 2013; **Published:** October 29, 2013

Copyright: © 2013 Picat et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: ARROW is funded by the UK Medical Research Council and the UK Department for International Development (DFID). ViiV Healthcare/GlaxoSmithKline donates first-line drugs for ARROW. JL was supported by an Engineering and Physical Sciences Research Council Life Sciences Interface Doctoral Training Centre studentship at the Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), and also received support from the National Institute for Health Research Biomedical Research Centre Funding Scheme at Great Ormond Street Hospital and the Institute of Child Health. NK was partly supported by Great Ormond Street Hospital Charity. The work was also supported by funding from an MRC Grant G1001190. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: AJP is a coinvestigator on the ARROW trial but did not receive any funds for his participation in this trial. AJP is a member of the WHO paediatric ARV working group but receives no financial support for this activity. AJP is funded by the Wellcome Trust. All other authors have declared that no competing interests exist.

Abbreviations: ARROW, Antiretroviral Research for Watoto; ART, antiretroviral therapy; CDM, clinically driven monitoring; IQR, interquartile range; LCM, laboratory and clinical monitoring; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; OR, odds ratio.

* E-mail: joanna.lewis@ucl.ac.uk

‡ These authors contributed equally to this work.

¶ These authors are joint senior authors on this work.

‡ Membership of the ARROW Trial Team is listed in the Acknowledgments.

Introduction

By December 2011, an estimated 3.3 million children under 15 y of age were living with HIV worldwide, over 90% of whom lived in sub-Saharan Africa [1]. While the 28% coverage of antiretroviral therapy (ART) for children in need lags behind the 58% coverage in adults, the number of HIV-infected children receiving ART has nevertheless increased to around 563,000, of whom 88% are in sub-Saharan Africa. Most receive ART under a public health approach with little, if any, laboratory monitoring [2,3]. Current guidelines recommend initiation of ART at age-related CD4 thresholds that are based on the risk of short-term disease progression. However, with increasing likelihood of survival into adulthood, a major challenge is to better understand how timing of ART initiation in childhood influences long-term immune reconstitution and immunological health [3]. To date, studies describing ART outcomes in resource-limited settings have usually reported only short-term (<1–2 y) disease progression and CD4 changes as outcomes [2,4]. Studies of longer-term immune reconstitution after ART initiation are scarce, and have typically included fewer than 200 children [5–8].

Although some large studies have assessed the long-term immunological effects of ART in adults in resource-limited settings [9,10], immune reconstitution differs significantly in children [11–14], most likely as a result of their more active thymopoiesis [15–17]. Successful approaches for studying long-term immune reconstitution in adults [18], and the conclusions drawn, may therefore not extend to children, and particularly not across all age groups in childhood. In addition, factors such as greater burden of co-infections and higher rates of malnutrition may result in differences in CD4 response between children in well-resourced and resource-limited settings.

We previously developed a statistical model for CD4 reconstitution in 127 European/South American HIV-infected children starting ART [19]. It assumes that the rate of increase in CD4 count (standardised for age) is fastest immediately following treatment initiation, falling exponentially with treatment duration so that CD4 count tends to a constant value in the long term, ideally approaching the count in a healthy child of the same age. By incorporating factors including age at treatment initiation in our model, we found that current guidelines for ART initiation, which are designed to reduce short-term risk of disease progression, may not always be optimal in terms of long-term immune reconstitution.

With the aim of understanding the kinetics of long-term CD4 reconstitution, we have investigated a large cohort of HIV-infected, ART-naïve children initiating ART in Uganda and Zimbabwe and followed up for a median (interquartile range [IQR]) 4.0 (3.7, 4.4) y. We identified qualitatively different CD4 recovery profiles using linear and nonlinear regression. We also present predictive models of long-term CD4 status based on age and CD4 count at ART initiation, constructed using nonlinear mixed-effects models. By taking into consideration long-term immunological health, this approach should better inform decisions for ART initiation, which to date have been based mainly on risks of short-term disease progression in untreated children before the era of ART.

Methods

Participants and Follow-Up

We used data from the Antiretroviral Research for Watoto (ARROW) trial, an open-label, randomised 5-y paediatric trial

comparing two monitoring strategies during first-line ART (<http://www.arrowtrial.org/>; ISRCTN 24791884 [20]). 1,206 HIV-infected, ART-naïve children aged 3 mo to 18 y and meeting WHO criteria for ART were enrolled at three centres in Uganda (Joint Clinical Research Centre, Paediatric Infectious Diseases Centre, and Medical Research Council/Uganda Virus Research Institute Uganda Research Unit on AIDS) and one in Zimbabwe (University of Zimbabwe, Harare). The trial was approved by research ethics committees in Uganda, Zimbabwe, and the United Kingdom. Written informed consent was obtained from children's parents/guardians, and written assent from the children as appropriate.

The ARROW trial's primary aim was to compare two monitoring strategies in children starting ART. Accordingly, children were randomised to either the "clinically driven monitoring" (CDM) or "laboratory and clinical monitoring" (LCM) group. In both groups, children had a routine full blood count, lymphocyte subsets (CD4, CD8), and liver and renal function tests (bilirubin, urea, creatinine, aspartate transaminase, alanine transaminase) performed every 12 wk. Demographic (age, sex) and clinical (WHO staging, height, weight) data were also collected. An additional week 4 lymphocyte subset measurement was available in children participating in an immunology substudy. In the CDM group, test results were returned to the clinician only if he or she requested them for clinical reasons, or for a grade 4 adverse event as defined in the trial protocol [21]. No total lymphocyte or CD4 counts were ever returned for CDM participants. Independently of monitoring randomisation, all children were also randomised to one of three first-line ART strategies. A control group (Arm A) received two nucleoside reverse transcriptase inhibitor (NRTI) drugs (lamivudine and abacavir) with one non-nucleoside reverse transcriptase inhibitor (NNRTI)—nevirapine in children aged <3 y and either nevirapine or efavirenz for older children based on clinician choice—according to WHO guidelines. The remaining children received these three drugs and also a third NRTI, zidovudine. After 36 wk, children who had started four-drug ART dropped either zidovudine (Arm B) or their NNRTI (nevirapine or efavirenz; Arm C). Criteria for failure of first-line ART and switch to second-line ART were clinical in both monitoring groups, or immunological in the LCM group only, and followed WHO guidelines [3]. Participants were followed under these monitoring/ART strategies until the end of the overarching ARROW trial on 16 March 2012.

No HIV viral loads were assayed in real time, following national guidelines, although samples were stored, and retrospective testing was carried out at a subset of time points and in a subset of children to address specific focussed questions [20]. The sampling design involved running assays for samples taken at 0, 4, 24, 36, 48, and 144 wk after ART initiation in all children aged <5 y at enrolment, plus in older children enrolled after 1 June 2008 and participating in the immunology substudy (assays 70% complete). Samples were also assayed at the time of a secondary randomisation to once versus twice daily lamivudine+abacavir, and 48 and 96 wk subsequently (assays 95% complete), and at 36, 48, and 84 wk in 41 children enrolled in a pharmacokinetic substudy of once versus twice daily lamivudine+abacavir (assays 100% complete). One centre also systematically performed viral load assays in all children as they exited the trial (i.e., did not select according to possible failure). In children for whom some post-enrolment viral loads were available, we characterised virological response to ART by calculating the mean \log_{10} viral load (in RNA copies/millilitre) ("average log-viral load") and the proportion of a

child's viral load measurements that were above 80 RNA copies/ml ("proportion of viral loads detectable").

The primary and secondary randomised comparisons have been reported separately [20]. This analysis is exploratory, was not specified in the protocol, and was conducted with the aim of understanding and modelling qualitatively different patterns of immune reconstitution.

Age-Adjusted CD4 Count

CD4 T cell counts were adjusted for age by dividing by the CD4 count that would be expected in HIV-uninfected children of the same age [22,23]. Thus, a ratio ("CD4-for-age") of 1 indicates the expected CD4 count for age, and ratios <1 indicate a lower than expected CD4 count for age. The CD4 z -score (normalised centile of the observed CD4 count within a reference population of uninfected, HIV-exposed children of the same age) [24] was not used here because z -scores do not allow the youngest children to have extremely low scores (e.g., 1 CD4 cell/ μl has z -score -8.1 in a 1-y-old, but -62 in a 10-y-old), and the z -score is also very sensitive to small errors in CD4 count at low values (Figure S1). Although CD4 percentage is more stable with age and has been widely used [25,26], CD4 count has been shown to have greater prognostic value [27] and so was the preferred variable in this analysis.

Statistical Analyses

Our mathematical model, which assumes nonlinear CD4-for-age recovery over time, tending towards a final stable level (asymptote) in the long term, has been previously described [19]. It is summarised as follows for CD4-for-age in child i at time t_{ij} :

$$\ln(\text{CD4-for-age}_{ij}) = \text{asy}_i - (\text{asy}_i - \text{int}_i) \exp(-c_i t_{ij}) \quad (1)$$

The child-specific intercept (int_i) is the $\ln(\text{CD4-for-age})$ value at treatment initiation. The asymptote (asy_i) is the long-term $\ln(\text{CD4-for-age})$ value. The proportional recovery rate c_i characterises the time for recovery, with $\ln(2)/c$ being the time for half the total recovery from int to asy to occur (illustrated in Figure 1). Transformation of CD4-for-age onto the log scale improves normality.

Identifying children with asymptotic CD4 reconstitution profiles. In preliminary analysis we found that the model did not converge onto the entire dataset. To identify children whose

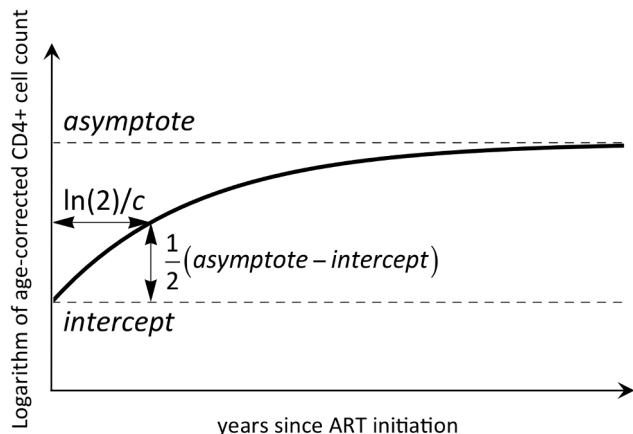


Figure 1. An illustrative representation of the mathematical model used to describe CD4 reconstitution. doi:10.1371/journal.pmed.1001542.g001

CD4-for-age trajectories followed Equation 1 ("asymptotic reconstitution"), each child's CD4 counts were modelled individually, using least squares to fit each child's data to Equation 1. Children were divided into two groups: one in which the least-squares fit had converged, and the other in which it had not. The first group thus consisted of children with asymptotic reconstitution, while the other comprised children with qualitatively different CD4 reconstitution profiles.

Modelling asymptotic reconstitution profiles. In the group with asymptotic CD4 reconstitution, recovery was analysed using nonlinear mixed-effects models for $\ln(\text{CD4-for-age})$, based on Equation 1. The mixed-effects model assumes that in each child the values of the three parameters int , asy , and c are the sum of a fixed effect—which describes the average response in the whole population of children with asymptotic reconstitution—and a normally distributed random effect—which accounts for and describes inter-individual variability. A normally distributed error term ϵ_{ij} is added to Equation 1 to represent noise, measurement error, and model mis-specification.

We investigated the predictive value of age, sex, WHO disease stage, height- and weight-for-age z -scores at ART initiation, and randomisation to monitoring and first-line ART strategies on proportional recovery rate (c) and CD4-for-age at ART initiation (int) and in the long term (asy), using a backwards-selection procedure based on an exit $p=0.01$ to favour parsimony and avoid including small significant but clinically unimportant effects given the size of the dataset. Height- and weight-for-age z -scores were calculated using UK norms, which, unlike WHO references, cover the whole childhood age range [28].

Characterising non-asymptotic reconstitution. In children without asymptotic CD4 reconstitution, we used linear models to fit CD4-for-age trajectories and identify those with increasing or decreasing CD4-for-age, those in whom CD4-for-age did not change significantly, and those with insufficient data to characterise a recovery profile.

Analyses were performed in R [29], using the nlme package [30] for nonlinear modelling. Graphs interpreting the fitted parameter values in terms of predicted CD4 trajectories were produced using Mathematica [31].

Results

1,206 ART-naïve children (51% girls) were enrolled in the ARROW trial between 15 March 2007 and 18 November 2008 and followed for a median (IQR) 4.0 (3.7, 4.4) y. 806 (67%) were from Uganda; 400 (33%) were from Zimbabwe. There were 600 children in the LCM group versus 606 children in the CDM group in the monitoring randomisation, and 397 children in Arm A (198 LCM, 199 CDM) versus 404 in Arm B (201 LCM, 203 CDM) versus 405 in Arm C (201 LCM, 204 CDM) in the ART strategy randomisation.

Median (IQR) age at ART initiation was 6.0 (2.4, 9.3) (range: 0.4–17.6) y, and the median (IQR) CD4 count and CD4-for-age were 360 (164, 694) cells/ μl and 0.28 (0.14, 0.43), respectively. At ART initiation, 683 (57%) and 169 (14%) children were at WHO disease stage 3 and 4, respectively. Children were moderately wasted (median [IQR] weight-for-age z -score -2.2 [-3.3 , -1.3]) and stunted (median [IQR] height-for-age z -score -2.4 [-3.4 , -1.5]).

At least one post-enrolment viral load measurement was available in 950 children (79%), who had a median (IQR) age of 4.4 (2.0, 7.7) y, a median CD4 count of 439 (197, 780) cells/ μl , and a median CD4-for-age of 0.29 (0.16, 0.46) at ART initiation. In these 951 children, the median (IQR) number of measurements

per child was 4 (3, 7), and samples were collected after a median (IQR) 84 (36, 144) wk on ART.

Contrasting Long-Term Profiles of CD4 Reconstitution Following ART Initiation

Our model identified two distinct groups according to long-term CD4 reconstitution profiles on ART. The first group included 914 (76%) children with asymptomatic CD4 reconstitution (average response in this group shown in Figure 2B). Among the remaining 292 (24%) children, CD4-for-age values followed qualitatively different trajectories. At ART initiation, children with asymptomatic reconstitution were younger ($p < 0.001$) and had lower CD4-for-age ($p < 0.001$) and height-for-age z -score ($p < 0.001$) (Table 1). In multivariable logistic models, experiencing asymptomatic reconstitution was independently associated with being younger at ART initiation (odds ratio [OR] [95% CI] = 0.94 [0.91, 0.97] per year older, $p < 0.001$) and having lower height-for-age (OR = 0.86 [0.78, 0.95] per unit higher height-for-age, $p = 0.003$). Children randomised to four-drug, two-class induction followed by two-NRTI+NNRTI maintenance (Arm B) were more likely to have asymptomatic reconstitution than the standard-of-care arm (Arm A) (OR = 1.48 [1.05, 2.07] higher, $p = 0.024$), but there was no evidence of a difference between three-NRTI maintenance (Arm C) and Arm A (OR = 0.93 [0.68, 1.28] higher, $p = 0.65$). Notably, monitoring strategy (CDM versus LCM) did not predict allocation to asymptomatic or non-asymptomatic reconstitution groups before ($p = 0.61$, Table 1) or after adjustment for other factors (OR [CDM:LCM] = 0.93 [0.71, 1.21], $p = 0.57$).

Clinically, children with non-asymptomatic recovery showed generally worse disease progression than those with asymptomatic recovery. 42 (14%) as opposed to 21 (2.3%) switched to second-line therapy (OR 7.13 [4.04, 12.92], $p < 0.001$) (none in the first year on ART), and 53 (18%) versus 59 (6.5%) experienced WHO stage 3 or 4 HIV events (OR 3.21 [2.11, 4.87], $p < 0.001$).

Post-enrolment viral loads were available in 750 (82%) of the children with asymptomatic reconstitution (median [IQR] 4 [3, 7] values available per child) and 200 (68%) of the children with non-asymptomatic reconstitution (median [IQR] 3 [2, 6] values available per child). There was some marginal evidence of poorer virological response to ART in the non-asymptomatic responders: median (IQR) average log-viral load was 2.07 (1.90, 3.45) compared to 2.03 (1.90, 2.39) in asymptomatic responders ($p = 0.050$), and the proportion of viral loads detectable (> 80 copies/ml) was 0.33 (0.00, 1.00) compared to 0.25 (0.00, 0.50) ($p = 0.014$).

Children with Asymptomatic CD4 Reconstitution

Figure 2A shows mean CD4-for-age throughout follow-up in this group, who had a median (IQR) age of 5.6 (2.3, 9.0) y and a median (IQR) CD4 count of 345 (176, 655) cells/ μ l at ART initiation (Table 1). Table 2 gives population average (fixed-effects) parameters as fitted in the nonlinear mixed-effects model, and Figure 2B and 2C illustrate the impact of age on expected CD4-for-age trajectory. There was no evidence of nonlinearity in the relationships between age at ART initiation and CD4 count either at ART initiation or in the long term. From the model, the estimated median (IQR) CD4 count in this group after 3.9 y on ART was 903 (658, 1,281) cells/ μ l. The proportional recovery rate, c , was 0.038 (95% CI: 0.036, 0.040) per week, corresponding to a half-life $t_{1/2} = \ln(2)/c$ (the time taken for half of the ultimate on-ART increase in CD4-for-age from intercept to asymptote to occur) of 18 (95% CI: 17, 19) wk.

A strong correlation was estimated between CD4-for-age at ART initiation and in the long term (correlation [95% CI] = 0.37 [0.31, 0.43], $p < 0.001$; Figure 2D): that is, children who started

ART with higher CD4-for-age achieved higher CD4-for-age in the long term. Interestingly, children with lower CD4-for-age at ART initiation had faster initial recovery (correlation = -0.35 [$-0.43, -0.27$], $p < 0.001$; Figure 2D). These two correlations combined so that those with faster initial recovery also ended up with lower long-term CD4-for-age (correlation = -0.44 [$-0.54, -0.33$], $p < 0.001$; Figure 2D). Such results are consistent with reconstitution processes that are sensitive to a low concentration of CD4 T cells, driving faster recovery in response to lower CD4-for-age, but that are nonetheless unable to reverse significant HIV-related impairment, which is reflected in lower CD4-for-age both before and after treatment. The second important observation was that in multivariable analyses (Table 2), lower CD4-for-age at ART initiation and in the long term were each independently associated with older age at ART initiation (both $p < 0.001$; Figure 2B). Lower CD4-for-age at ART initiation was also associated with lower weight-for-age z -score ($p < 0.001$). Boys achieved lower long-term CD4-for-age values than girls, as did children randomised to zidovudine-containing three-NRTI maintenance (Arm C) compared to those in Arm A (both $p < 0.001$). Using backwards selection, we found no covariates whose inclusion in the model for proportional recovery rate c substantially improved the fit.

Because CD4 counts in adults starting ART have often been fitted using linear mixed-effects models with a change in slope some weeks after treatment initiation [18], we fitted a similar model in our CD4-for-age data. It proved a substantially worse fit (Akaike information criterion 976 units higher) than the asymptotic model, confirming use of our Equation 1 as more appropriate. Because CD4 recovery in adults is known to be biphasic [32], we also investigated a model that was the sum of two asymptotic functions with different proportional rates of recovery, but were unable to fit it to our data.

Using the fitted model parameters, we can predict CD4 trajectories for children with asymptomatic reconstitution starting ART at different ages and with different CD4 counts. Figure 3A illustrates predicted trajectories in children starting ART having reached current (2010) WHO ART initiation criteria aged 2, 6, or 10 y. While the 2-y-old is predicted to recover CD4 count well, the 6- and 10-y-olds show progressively impaired recovery. Thus, starting ART at older ages appears to impair long-term CD4 count when treatment is initiated at the WHO CD4 count thresholds for ART initiation, which differ substantially between children aged < 5 y and > 5 y. From Figure 3B it is apparent that the fixed CD4 threshold for starting ART in children over 5 y old, while satisfactorily reducing risk of disease progression, increasingly limits the potential for optimal CD4 recovery as the child approaches adulthood. In early life, however, relatively large changes in the thresholds for ART initiation have a fairly limited impact on long-term CD4 count.

Children with Non-Asymptomatic CD4 Reconstitution

At ART initiation, the 292 children whose CD4 recovery differed qualitatively from the asymptomatic reconstitution described above had a median (IQR) age of 7.3 (2.8, 10.7) y and a median CD4 count of 401 (155, 837) cells/ μ l. Fitting individual linear models to CD4-for-age measurements in each of these children allowed us to identify four subgroups within this population based on a per-child test of significance for the change in CD4-for-age over follow-up (Figure 4; Table 3).

Of these 292 children, 23 (8%) (median [IQR] age: 9.2 [6.6, 11.6] y, CD4 count: 281 [89, 1,013] cells/ μ l, and CD4-for-age: 0.28 [0.08, 0.43], all at ART initiation) had a significant ($p < 0.05$) decline in CD4-for-age over follow-up. These children's mean

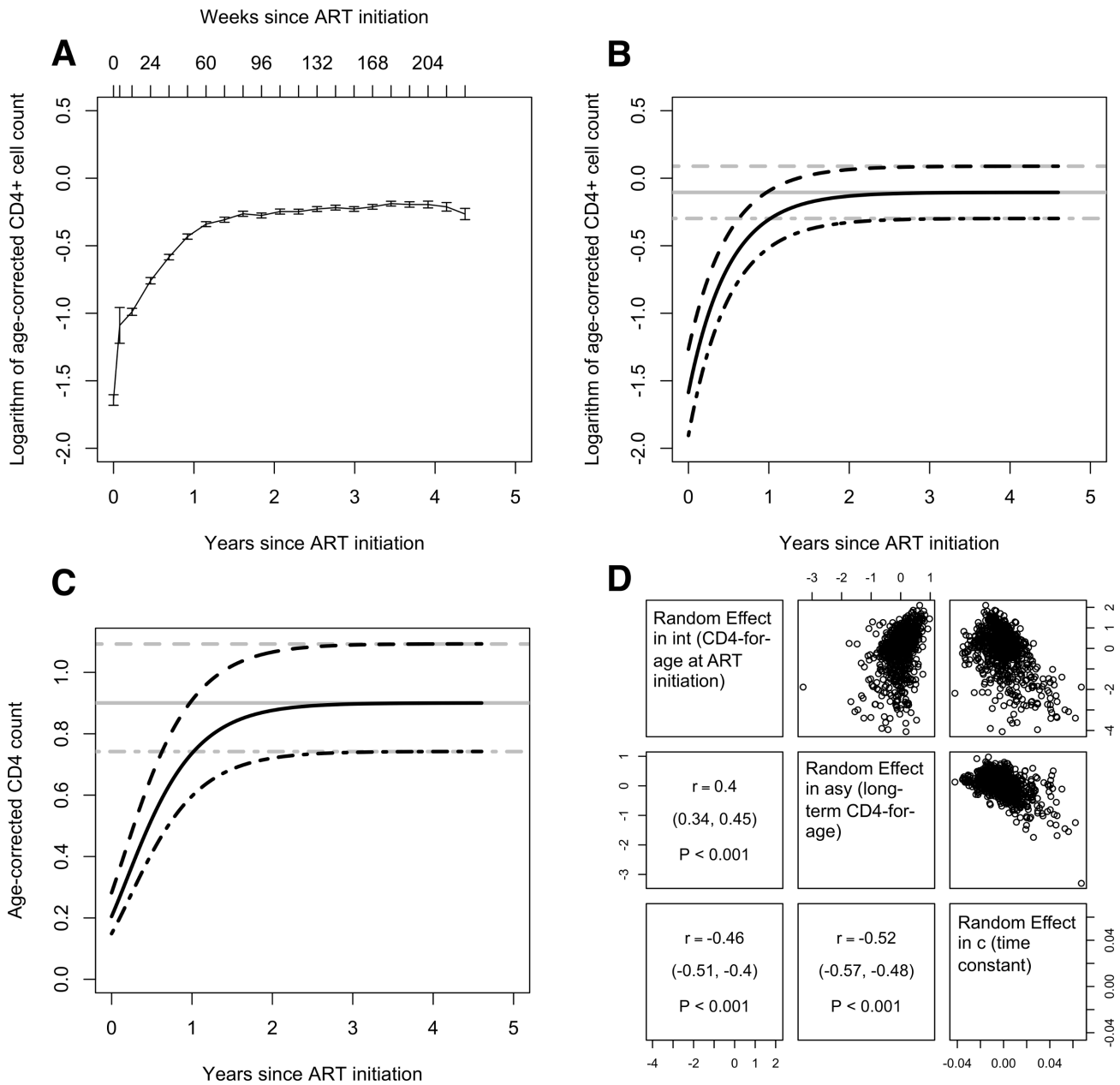


Figure 2. Modelling CD4-for-age in children showing asymptotic reconstitution. (A) Mean CD4-for-age over time since ART initiation for 914 children with asymptotic CD4 reconstitution. Errors bars give standard error of the mean. Points are shown for times where >20% of children in this group remained in the trial. Only children participating in a substudy had measurements available at week 4, accounting for the wider error bar. The numbers of children with measurements available at 0, 4, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, and 228 wk were 914, 153, 897, 893, 900, 893, 894, 898, 989, 900, 902, 898, 899, 897, 896, 892, 861, 790, 516, 354, and 232, respectively. (B) Population-average (fixed-effects) $\ln(\text{CD4-for-age})$ trajectories predicted by our nonlinear mixed-effects model (Table 2) for “average” HIV-infected children (female, weight-for-age z-score at ART initiation -2.2) aged 2 (dashed line), 6 (solid line), and 10 (dot-dashed line) y. (C) Predicted CD4-for-age trajectories as shown in (B), but plotted on the linear scale (i.e., $\text{CD4}/\text{expected CD4}$). (D) Correlations between child-level random effects (difference from population average) in the three model parameters. Each panel refers to correlations between the estimated random effects (differences between child-specific parameters and the population average) in two of the three parameters int (CD4-for-age at ART initiation), asy (long-term CD4-for-age), and c (proportional recovery rate). The pair of parameters being considered is identified by the row and column labels in the panels on the diagonal. The upper-right panels plot the values of the random effects against one another. The lower-left panels give Pearson correlations between the estimated random effects, with 95% confidence intervals and p -values. doi:10.1371/journal.pmed.1001542.g002

CD4-for-age values showed an initial, small response to treatment in the first few weeks on ART (Figure 4A). After week 12, however, mean CD4-for-age began to fall, possibly as a result of treatment failure. Finally, as non-response or failure was detected

and treatment switched to second-line ART, the CD4-for-age began to recover again. Of note, in univariable tests (Table S1) these children differed from the larger group with asymptotic CD4 reconstitution only in their older age ($p = 0.002$). There was no

Table 1. Characteristics at ART initiation of the two groups of HIV-infected ART-naïve children in the ARROW trial with qualitatively different CD4 reconstitution profiles.

Characteristic	Asymptotic CD4 Reconstitution (Figure 1) (n=914)	Non-Asymptotic CD4 Reconstitution (n=292)	p-Value
Female, n (percent)	474 (52%)	136 (49%)	0.13
Age (years)	5.6 (2.3, 9.0)	7.3 (2.8, 10.7)	<0.001
WHO stage, n (percent)			0.72
1	17 (2%)	3 (1%)	
2	251 (27%)	83 (30%)	
3	514 (56%)	169 (61%)	
4	132 (14%)	37 (13%)	
CD4 count (cells/μl)^a	350 (170, 646)	401 (155, 837)	0.062
CD4-for-age (count/healthy count)^a	0.26 (0.14, 0.41)	0.32 (0.14, 0.55)	<0.001
CD4-for-age ln(count/healthy count)^a	-1.3 (-2.0, -0.9)	-1.1 (-1.9, -0.6)	<0.001
Weight-for-age z-score	-2.3 (-3.3, -1.3)	-2.0 (-3.2, -1.0)	0.023
Height-for-age z-score	-2.5 (-3.5, -1.6)	-2.0 (-3.1, -1.1)	<0.001
Monitoring randomisation, n (percent)			0.61
LCM	459 (50%)	141 (48%)	
CDM	455 (50%)	151 (52%)	
First-line ART randomisation, n (percent)			0.024
Arm A	294 (32%)	103 (35%)	
Arm B	325 (36%)	79 (27%)	
Arm C	295 (32%)	110 (38%)	

Data are median (IQR) unless otherwise indicated. p-Values are for Fisher's exact (WHO stage and first-line ART randomisation), chi-squared (sex and monitoring randomisation), and Wilcoxon rank sum (continuous variables) tests for a difference between the two groups.

^aCD4 count at ART initiation was missing in one child from each group.

doi:10.1371/journal.pmed.1001542.t001

Table 2. Predictors of CD4-for-age after ART for 914 children with asymptotic CD4 reconstitution.

Variable	Estimate (95%CI)		p-Value
	Natural Scale (ln[CD4-for-Age])	Linear Scale (CD4-for-Age)	
Intercept (CD4-for-age at ART initiation)	-1.59 (-1.66, -1.52)	0.20 (0.19, 0.22)	<0.001
Age at ART initiation (per year older)	-0.08 (-0.10, -0.06)	0.92 (0.91, 0.94)	<0.001
Weight-for-age at ART initiation (per unit z-score higher)	0.16 (0.12, 0.20)	1.17 (1.12, 1.22)	<0.001
Asymptote (long-term CD4-for-age)	-0.21 (-0.26, -0.16)	0.81 (0.77, 0.85)	<0.001
Age at ART initiation (per year older)	-0.048 (-0.055, -0.042)	0.95 (0.95, 0.96)	<0.001
Sex (female versus male)	0.11 (0.06, 0.16)	1.11 (1.06, 1.17)	<0.001
ART strategy (Arm B versus Arm A)	0.044 (-0.017, 0.105)	1.04 (0.98, 1.11)	0.16
ART strategy (Arm C versus Arm A)	-0.14 (-0.20, -0.08)	0.87 (0.82, 0.93)	<0.001
c	0.038 (0.036, 0.040)	—	<0.001

Results are expressed relative to the reference category (boy, aged 6 y, with weight-for-age z-score -2.2 and randomised to first-line ART strategy Arm A [standard of care two-NRTI+NNRTI throughout]). Analysis was carried out on the natural logarithm scale (ln[CD4/expected CD4]), but results are reported both on the natural scale and back-transformed to the linear scale (CD4/expected CD4). Estimated effect sizes of covariates on the baseline (intercept) and the long-term (asymptote) CD4-for-age are additive on the natural log scale, but multiplicative on the linear scale—for example, a 1-y increase in age is associated with a -0.08 unit lower ln(CD4/expected CD4) at ART initiation or, equivalently, a CD4/expected CD4 lower by a factor of $\exp(-0.08) = 0.92$. Similar effect sizes were seen in univariable models. After adjusting for the factors shown, there was no additional predictive value on intercept of sex, WHO disease stage 3 or 4, or first-line ART or monitoring strategy; on asymptote of WHO disease stage 3 or 4, weight-for-age z-score, or monitoring strategy, or on c of age, sex, WHO disease stage 3 or 4, or ART or monitoring strategy (likelihood ratio tests, $p > 0.01$; further details are given in Table S2). Although weight-for-age was not selected in the backwards selection, likelihood ratio tests to assess additional predictors on top of the final multivariable model suggested weight-for-age might be a predictor for c ($p = 0.003$ in likelihood ratio test; estimate 9.31×10^{-4} higher per one-unit increase in weight-for-age [95% CI: -7.01×10^{-5} , 1.93×10^{-3}], i.e., faster proportionate recovery speed in those with higher weight-for-age at ART initiation). However, other effect size estimates were essentially unchanged, and examining closely related covariate models suggests this apparent predictive value may be the result of overfitting. Thus, the final backwards selection is presented as the final model, as described in the Methods.

doi:10.1371/journal.pmed.1001542.t002

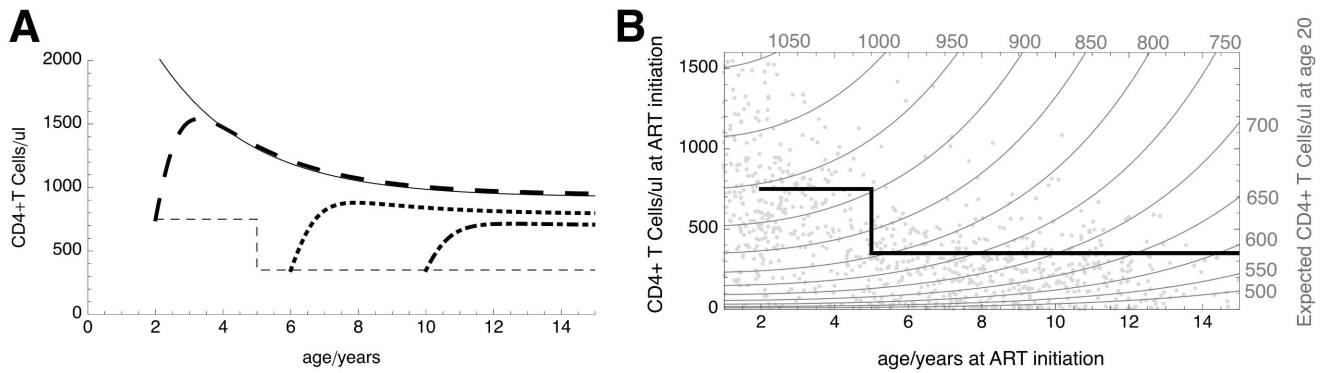


Figure 3. Predicted long-term CD4 counts in children starting ART at different ages and CD4 levels. (A) CD4 trajectories predicted for children starting ART having reached WHO CD4 count thresholds at age 2 (dashed line), 6 (dotted line), or 10 (dot-dashed line) y. The thin dashed line indicates WHO thresholds for ART initiation, and the thin solid line the trajectory in CD4 count with age expected in a healthy child. (B) Expected CD4 count on immunological maturity (estimated at age 20 y) for different ages and CD4 counts at ART initiation. Values at the ends of the grey contour lines indicate expected adult CD4 count in children starting ART at the ages and CD4 counts given on the horizontal and vertical axes. The black line indicates the current WHO CD4 thresholds for ART initiation. Grey point markers show the age and fitted CD4 count at ART initiation of the 914 children on whom the model is based. They indicate at which ages/CD4 counts the model has substantial evidence, and where it represents an extrapolation from the available data. doi:10.1371/journal.pmed.1001542.g003

evidence of between-group differences in other parameters, including CD4-for-age at ART initiation, and monitoring and ART strategy randomisations ($p > 0.08$). Clinical progression was worse than in the asymptomatic group: five children (22%) switched to second-line treatment (none in the first year on ART), 10 (43%) experienced WHO stage 3 or 4 events, and six (26%) died (all $p < 0.001$ for comparison to the large asymptomatic group; Table S1). Viral loads were available in 11 children in this group (48%; median [IQR] 2 [1.5, 3] measurements per child). These children had poorer virological control than asymptomatic responders: average log-viral load was higher (median [IQR] 4.52 [2.02, 4.76] RNA copies, $p = 0.014$) and a greater proportion of viral loads were detectable (1.00 [0.83, 1.00], $p < 0.001$).

79 (27%) of the 292 children (median [IQR] age: 5.5 [2.4, 8.8] y, CD4 count: 360 [83, 737] cells/ μ l, and CD4-for-age: 0.28 [0.07, 0.41] at ART initiation) showed continuing improvement in CD4-for-age with time, characterised by a significant ($p < 0.05$) positive linear slope (Figure 4B). Like the children with asymptomatic reconstitution, these children's CD4-for-age increased with time on ART. Unlike the asymptomatic group, however, their CD4-for-age increased more slowly and continuously over 4–5 y on ART, although it is possible that CD4 count will plateau in the years following the last available measurement. These children may be showing CD4 recovery that is slower for reasons connected to the mechanism of recovery, and/or to qualitatively different functional impairments at ART initiation. Compared to the large group with asymptomatic reconstitution (Table S1), there was evidence in univariate comparisons that these children had lower (less severe) WHO disease stage ($p = 0.0048$ for both stage 3 and stage 4 versus stage 2) and were more likely to be in Arm A and less likely to be in Arm B ($p = 0.024$). Other characteristics were similar ($p > 0.05$). Clinical progression was worse than in the asymptomatic group: 21 (27%) children switched to second-line therapy (none in the first year on ART), and 16 (20%) experienced WHO stage 3 or 4 HIV events (both $p < 0.001$), although only one child died (Table S1). 60 (76%) of these children had at least one viral load recorded, with a median (IQR) 5 (2, 9) measurements per child. Their virological response was similar to that of the asymptomatic group: median [IQR]

average log-viral load was 1.96 (1.84, 3.10) ($p = 0.70$), and the proportion of viral loads detectable was 0.17 (0.00, 0.53) ($p = 0.60$).

Linear models in 153 (52%) of the 292 children (median [IQR] age: 8.1 [3.1, 11.3] y; CD4 count: 479 [230, 980] cells/ μ l; CD4-for-age: 0.40 [0.21, 0.66] at ART initiation) showed no evidence of a significant change in CD4-for-age over the first 4–5 y on ART ($p > 0.05$) (Figure 4C). These children started ART with a higher CD4 count and CD4-for-age (both $p < 0.001$ versus those with asymptomatic CD4 reconstitution) and had a small, fast increase in CD4-for-age in the first few weeks following ART, compatible with redistribution of memory CD4 T cells. Following this initial increase, however, CD4-for-age showed no further recovery over 4–5 y of follow-up. Moreover, when ages, weights-for-age, and CD4 counts in these children were used to predict their long-term CD4-for-age according to the nonlinear mixed-effects model for asymptomatic reconstitution described above, we found a median (IQR) predicted long-term CD4-for-age of -0.39 ($-0.21, +0.00$)—somewhat higher than the steady level observed (around -0.5). While these children did show some response to treatment, they seemed to be less able to reconstitute CD4 cells in the long term than the asymptomatic group. In addition to their higher CD4-for-age at ART initiation, in univariate comparisons (Table S1) these children were older, heavier, and taller than the large group with asymptomatic CD4 reconstitution ($p < 0.001$). Clinical progression was worse than in the asymptomatic responders: 16 children (10%) switched to second-line therapy ($p < 0.001$), 21 (14%) experienced WHO stage 3 or 4 HIV events ($p = 0.004$), and eight died (5.2%, $p < 0.001$). (ORs for clinical progression are given in Table S1.) 113 (74%) of these children had viral loads recorded, with a median [IQR] 4 (2, 7) measurements available per child. Their virological response was again similar to that of the asymptomatic group: median (IQR) average log-viral load was 2.03 (1.90, 3.18) ($p = 0.28$), and the proportion of viral loads detectable was 0.33 (0.00, 0.75) ($p = 0.18$).

Finally, 37 (13%) of the 292 children (median [IQR] age at ART initiation: 5.4 [1.8, 9.1] y) had very low baseline CD4 counts (median [IQR] 130 [22, 671] cells/ μ l)—and thus low CD4-for-age (median [IQR] 0.12 [0.02, 0.41])—and ≤ 1 CD4 measurements

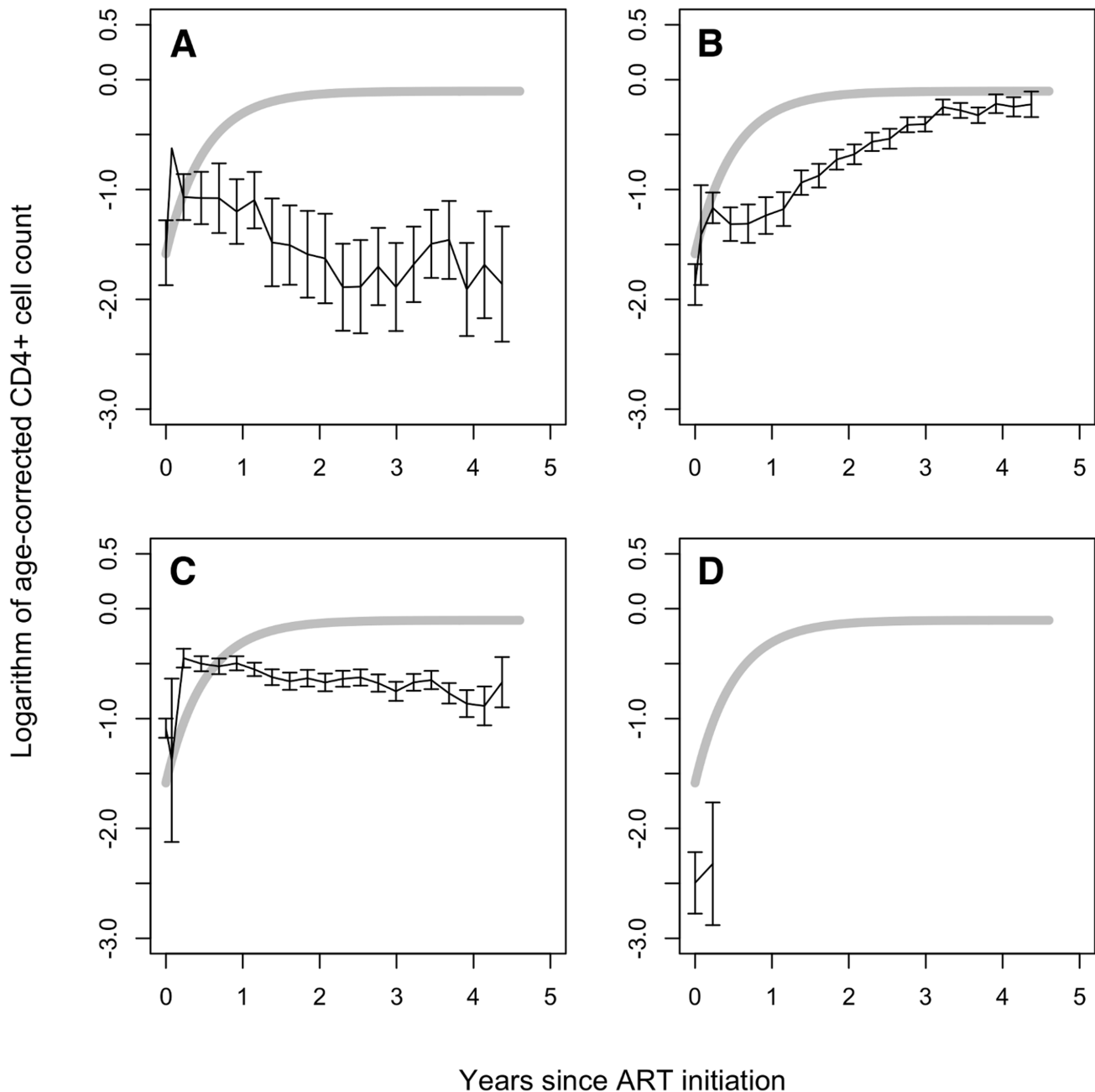


Figure 4. CD4-for-age in children not showing asymptotic reconstitution. Mean CD4 cell count for age with time since ART initiation (error bars: standard error of the mean) for: (A) 23 children with significant ($p < 0.05$) decreasing CD4-for-age with time; (B) 79 children with significant ($p < 0.05$) increasing CD4-for-age with time but not fitting the asymptotic model; (C) 153 children with no significant change ($p > 0.05$) in CD4-for-age with time; and (D) 37 children with only baseline CD4 count available. For comparison to the asymptotic group, grey lines indicate average (fixed-effects, for a 6-y-old girl with weight-for-age z-score -2.2) CD4-for-age in the 914 children with asymptotic CD4 reconstitution. In (A–C), data points are shown where $>20\%$ of children in each group remained in the trial. doi:10.1371/journal.pmed.1001542.g004

available after ART initiation (Figure 4D). 32 of these children died, and five were lost to follow-up, early in the trial.

Discussion

At present it is recommended that children initiating ART should continue treatment for life. Understanding the long-term impact of treatment initiation at different CD4 counts and ages is

therefore essential to ensure that children reach adulthood with immune systems as intact as possible. We have conducted an exploratory study characterising qualitatively and quantitatively different CD4 responses to ART in a large sample of children initiating ART. The analysis, which was not pre-planned or specified in the trial protocol, has allowed us to identify subpopulations of children who respond in qualitatively different ways to ART over the long term.

Table 3. Characteristics at ART initiation of four subgroups of HIV-infected ART-naïve children who did not show asymptotic CD4-for-age reconstitution on long-term ART ($n = 292$).

Characteristic	CD4-for-Age Decreasing ($p < 0.05$) ($n = 23$)	CD4-for-Age Increasing ($p < 0.05$) ($n = 79$)	No Significant Change in CD4-for-Age ($p > 0.05$) ($n = 153$)	≤ 2 Measurements ($n = 37$)	p -Value
Female, n (percent)	12 (52%)	40 (51%)	68 (44%)	16 (43%)	0.59
Age (years)	9.2 (6.6, 11.6)	5.5 (2.4, 8.8)	8.1 (3.1, 11.3)	5.4 (1.8, 9.1)	0.009
WHO stage, n (percent)					0.16
1	0 (0%)	1 (1%)	2 (1%)	0 (0%)	
2	6 (26%)	32 (41%)	39 (25%)	6 (16%)	
3	13 (57%)	39 (49%)	100 (65%)	17 (46%)	
4	4 (17%)	7 (9%)	12 (8%)	14 (38%)	
CD4 count (cells/μl)^a	281 (89, 1013)	360 (83, 737)	479 (230, 980)	130 (22, 671)	0.016
CD4-for-age (count/healthy count)^a	0.28 (0.08, 0.43)	0.28 (0.07, 0.41)	0.40 (0.21, 0.66)	0.12 (0.01, 0.34)	0.002
CD4-for-age ln(count/healthy count)^a	-1.3 (-2.5, -0.9)	-1.3 (-2.7, -0.9)	-0.9 (-1.6, -0.4)	-2.1 (-4.2, -1.1)	0.002
Weight-for-age z-score	-1.7 (-2.8, -1.4)	-2.0 (-3.2, -0.8)	-1.8 (-2.8, -0.8)	-3.5 (-5.0, -2.4)	0.55
Height-for-age z-score	-2.1 (-3.4, -1.1)	-2.3 (-3.3, -1.0)	-1.9 (-2.8, -1.2)	-2.7 (-3.4, -1.4)	0.73
Monitoring randomisation, n (percent)					0.33
LCM	7 (30%)	38 (48%)	71 (46%)	25 (68%)	
CDM	16 (70%)	41 (52%)	82 (54%)	12 (32%)	
First-line ART randomisation, n (percent)					0.14
Arm A	7 (30%)	35 (44%)	48 (31%)	13 (35%)	
Arm B	10 (43%)	20 (25%)	42 (27%)	7 (19%)	
Arm C	6 (26%)	24 (30%)	63 (41%)	17 (46%)	

Data are median (IQR) unless otherwise indicated. p -Values are for Fisher's exact (categorical variables) or Kruskal-Wallis (continuous variables) tests for a difference between the four groups.

doi:10.1371/journal.pmed.1001542.t003

The majority of children (76%) had asymptotic CD4 reconstitution following ART initiation, with an initial steep increase in CD4-for-age that slowed with a half-life of ~ 18 wk, tending towards a constant level over the long term of (for a child of average demographics) $\sim 80\%$ of the CD4 count expected in an HIV-uninfected child of the same age. CD4-for-age both before ART and in the long term were higher for children starting ART at younger ages, and long-term CD4 reconstitution was better in children with higher CD4-for-age at ART initiation. These results in resource-limited settings reiterate findings in European cohorts [19,33], and demonstrate the need for earlier diagnosis in all settings, the value of testing children at multiple points of access to healthcare services, and the importance of early availability of ART at minimum at the recommended thresholds for ART initiation. Of note, we found no evidence in these more sophisticated models that CD4 reconstitution (asymptotic or non-asymptotic recovery, or any model parameter) depended on whether children received ART with or without CD4 monitoring. This finding is similar to that of the main randomised comparison [20] and contrasts with findings in adults, in whom there is evidence that CD4 monitoring has benefits for long-term immune reconstitution [9].

Unexpectedly, none of the pre-ART factors considered, nor the randomised groups, were associated with the parameter characterising speed of proportional CD4 recovery. Although this might at first appear inconsistent with previous studies showing faster recovery in younger children [33–35], the measure of recovery used elsewhere has generally been an absolute one rather than the proportional rate (ϵ), which we used here. Thus, absolute reconstitution may be faster in younger children in that a given

increase in CD4 count or percentage takes less time, but a given proportion of the total long-term improvement achieved may still occur in the same time, regardless of age. Nonetheless, we did find significant inter-individual variability in proportional recovery rate that was unexplained by the pre-ART factors studied. Genetic differences between children and/or viral strains may be one explanation for this, and the qualitative composition of the CD4 count when ART began—for example, the percentages of activated and proliferating cells—may also play a role.

As in European/US children, we found that the asymptotic model was a better fit to the data from African children than the two-slope linear mixed-effects model that is optimal in adults [18]. This finding adds further weight to the hypothesis that CD4 recovery in children differs qualitatively and mechanistically from that in adults, perhaps reflecting greater involvement of the naive T cell pool and greater dependence on thymopoiesis; this finding is supported by the difference in the impact of CD4 monitoring on long-term reconstitution in adults [9] compared to children [20]. Some of the conclusions from adult studies do, however, apply also in children. For example, like the children studied here, adults with low CD4 counts at ART initiation have been observed to have generally lower CD4 counts 3 y later [36].

Using parameters obtained from the nonlinear mixed-effects model for asymptotic CD4 reconstitution, we were able to predict CD4 trajectories for children starting ART at different ages and with different CD4 counts, reflecting the response that would be expected from the majority of children initiating ART. Qualitatively, our findings are similar to those obtained in previous work on European and Brazilian children [19], but the far larger numbers in the present study now provide much greater certainty

and also demonstrate that factors such as co-infection burden, malnutrition, and trial recruitment criteria do not appear to cause large differences in the CD4 response to ART. From Figure 3B, predicted CD4 count at age 20 y (i.e., in adulthood) can be estimated for every combination of age and CD4 count at ART initiation. Although our predictions extrapolate the data forward in time, the wide age and CD4 range at ART initiation, and the stability of results compared to analysis through 3 y, suggest that the extrapolations are reasonable. Nevertheless, further analysis of long-term CD4 response in adolescents will be important as validation. The figure shows how starting ART at the same CD4 count but at progressively older age is increasingly detrimental to long-term immunological health. For example, to achieve a CD4 count of >700 cells/ μl at 20 y of age, children under 8 y need to start ART with a CD4 count >223 cells/ μl , but a 10-y-old would need to start before CD4 count falls below 347 cells/ μl , and a 12-y-old before reaching 557 cells/ μl . In contrast, a 2-y-old would need only a CD4 count >96 cells/ μl , and a 5-y-old, CD4 count >130 cells/ μl . To optimise immune reconstitution when children reach adulthood, our findings suggest that ART may be warranted in vertically HIV-infected children who are still ART-naïve at ages over 10 y, regardless of CD4 count. In young children, however, current guidelines designed to minimise risk of disease progression also preserve the potential for good CD4 counts in the long term.

Clearly, the aim of an ART strategy for an HIV-infected child is not limited to maximising CD4 count, but is a matter of balancing overall immunological health—encompassing lymphocyte dynamics, population structure and receptor diversity, and the effectiveness of the immune response overall—against short-term side effects, long-term toxicity, regimen durability (particularly in resource-limited settings where frequently only two regimens are available), cost, treatment adherence, and social factors. Even a CD4 count as low as 500 cells/ μl is above the 2.5th centile for CD4 count in a population of healthy adults, raising questions about the appropriateness of aiming for a long-term CD4 count as high as, for example, 700 cells/ μl , especially where multiple other factors might argue against ART. However, even relatively small reductions from normal CD4 counts in HIV-infected patients have been associated with increased risk of cardiovascular disease and malignancies [37–39]. We also note that our predictions are of CD4 response expected on average in HIV-infected children starting ART. While an increase in the CD4 count threshold for ART initiation might result in an only marginally higher long-term count for the “average” child, the shift in the whole distribution of CD4 counts to higher levels results in fewer children being likely to experience lower counts.

Interestingly, we found slightly, but significantly, lower long-term CD4-for-age values in children receiving three-NRTI maintenance (including zidovudine) (Arm C), despite no differences in CD4 cell percentage over the long term [20] and no difference in the proportion achieving asymptotic response compared to standard two-NRTI+NNRTI treatment. One explanation is the known haematological toxicity of zidovudine, which may modestly suppress all components of the white blood count, leaving proportionate (percentage) recovery unchanged. Whether these small differences in long-term asymptote ($\sim 13\%$ reduction compared to Arm A) are clinically important is unknown. However, we also found that children randomised to Arm B (four-drug, two-class induction followed by standard two-NRTI + NNRTI maintenance) were more likely to have asymptotic than non-asymptotic reconstitution profiles, a novel finding not identified by the standard comparison of mean responses [20]. This supports the potential of enhanced initial treatment to improve long-term CD4 responses.

Approximately one-quarter of children did not have asymptotic CD4 reconstitution. The fact that subgroups within the population had very different patterns of reconstitution, and varied in both CD4 count and age at ART initiation, highlights the issue of modelling one single population mean response across qualitatively different types of reconstitution. Those who had a constant or increasing CD4-for-age following ART had a higher CD4-for-age at ART initiation. We suggest that the differences between these children and those showing asymptotic reconstitution may arise from different reconstitution mechanisms, or qualitatively different CD4 T cell populations at ART initiation. Other children whose CD4 count declined on ART had similar CD4-for-age at ART initiation but were older than the group with asymptotic CD4 reconstitution. Children with ≤ 2 CD4 measurements after ART initiation had had very low CD4-for-age values when ART was begun; most died early in the trial. Factors affecting reconstitution profiles in these 292 children may have included clinical factors, adherence, and viral dynamics, and investigating these and other factors/mechanisms influencing CD4 recovery in children with atypical CD4 responses will provide interesting directions for further research. Our findings regarding the different CD4 reconstitution groups should also be validated in independent studies.

Children in each of the non-asymptotic recovery groups experienced more regimen switching, WHO disease stage 3 and 4 events, and deaths than the large group of asymptotic responders. While this is an interesting and important observation, it does not allow any conclusions about causation to be drawn. On the one hand, it is possible that more frequent clinical events might have impaired children's CD4 recovery and led to their different reconstitution profiles. Equally, however, since children in the non-asymptotic groups had CD4-for-age values below those of the asymptotic responders at almost all time points (Figure 4), it is unsurprising that they experienced more regimen switching, clinical events, and deaths. What can be stated is that, since logistic regression models showed little absolute difference in age-corrected CD4 counts between the two groups at ART initiation, the CD4 count when ART begins does not explain the substantial difference between asymptotic and non-asymptotic recovery.

One of the limitations of this study was the lack of systematic viral load measurements, which have only been retrospectively assayed in subsets of children at restricted time points [20] (further work ongoing). Previous work suggests that viral loads before ART are likely to predict CD4 count at ART initiation (though not in the long term) [19], and that joint models incorporating CD4 count and viral load are superior to univariate models of CD4 response only [40]. In the absence of systematic viral load measurements on all children, we were not able to assess conclusively whether patterns of viraemia could be contributing to the patterns of CD4 response. Nevertheless, the incomplete results that were available suggested that the children with falling CD4-for-age (Figure 4A, 23 children) had poorer virological control than the group with asymptotic recovery. However, children with slowly increasing (Figure 4B, 79 children) or constant (Figure 4C, 153 children) CD4-for-age had similar virological responses to the asymptotic responders. There were examples of viral rebound in all response groups. Hence, although virological control may be an important factor for children with falling CD4 counts, it does not appear to explain the differences in CD4 responses between the asymptotic group and the two major non-asymptotic groups.

A limitation of the modelling approach is that uncertainty as to children's group allocation to asymptotic or non-asymptotic recovery is not accounted for. This problem is illustrated by the

observation that an earlier analysis—of data up to 1 April 2011 (i.e., missing the last year of the trial)—reversed the allocations of a small number of children. The uncertainty in allocation could be modelled using mixture models or latent class models, and this will be an important further step in developing this kind of analysis. However, the robustness of allocation in the vast majority of children when data from the last year of the trial were included or excluded suggests it is unlikely that failing to model the uncertainty has materially affected our conclusions. In previous work, we fitted a nonlinear mixed-effects model with an asymptotic recovery function to a whole cohort of HIV-infected children starting ART. In spite of modelling recovery as asymptotic in all children, the same conclusion was reached: that age affects CD4 count both before ART and in the long term, and that there is an additional correlation between unexplained differences in counts at ART initiation and in the long term [19]. Thus, although parameter estimates might be revised slightly by a more rigorous model, it seems likely that our conclusions would remain unchanged. The allocation of children to asymptotic versus non-asymptotic recovery groups also raises a problem with predictions: while predictions like those illustrated in Figure 3B are helpful for children who will go on to experience asymptotic recovery, they are less relevant to those with different recovery profiles. This is a particular example of a general and widespread problem with using population-level responses to inform treatment guidelines. In this context, at the level of individual children, it highlights the challenge of identifying baseline predictors of non-asymptotic recovery so that those children can be treated differently.

In conclusion, our analysis of data from a large cohort of HIV-infected, ART-naïve African children initiating ART provides important information to inform clinical practice by considering long-term CD4 reconstitution in response to therapy in resource-limited settings, and the roles of age and CD4 count at ART initiation in determining this response. Given increased availability of simpler ART formulations and regimens for children, we have reinforced the importance of considering long-term immunological health, as well as short-term disease progression, in the decision to initiate ART and the development of treatment guidelines. Our data suggest that even though the short-term risks of morbidity and mortality are greater in younger untreated children, survivors have excellent potential for immune reconstitution when they initiate ART. In contrast, whilst not subject to the same immediate risks of disease progression/death, older children who wait to initiate ART may never reconstitute immunologically to the same degree. The current CD4 count thresholds for ART initiation are predominantly based on these short-term risks, and in older children could lead to delays in ART administration that might compromise immune reconstitution in adulthood.

Our results indicate that although younger ART-naïve children are at high risk of disease progression, provided they start ART following current WHO/Paediatric European Network for Treatment of AIDS/US Centers for Disease Control and Prevention guidelines, they have good potential for achieving high CD4 levels in later life. However, to attain maximum immune reconstitution in older children, particularly those >10 y, ART may need to be initiated regardless of CD4 cell count. The approach taken in this study to characterise long-term immunological outcomes has the potential, by identifying qualitatively different CD4 recovery

profiles, to improve our understanding of the factors that determine children's CD4 counts on ART and provide input into future management guidelines in both resource-rich and resource-limited settings.

Supporting Information

Figure S1 Relationship between very low CD4 count and CD4-for-age z-score in children of different ages.

(PDF)

Table S1 Univariate comparisons between individual non-asymptotic response subgroups and the large (n = 914) group of asymptotic responders.

(DOC)

Table S2 Additional predictors of CD4-for-age profiles in the asymptotic recovery group.

(DOC)

Acknowledgments

We thank the children, carers, and staff from all the centres participating in the ARROW trial, and the ARROW Trial Steering Committee for access to data.

MRC/UVRI Uganda Research Unit on AIDS, Entebbe, Uganda: P. Munderi, P. Nahirya-Ntege, R. Katuramu, L. Matama, F. Nankya, G. Nabulime, A. Ruberantwari, R. Sebukyu, G. Tushabe, D. Nakitto-Kesi. **Joint Clinical Research Centre, Kampala, Uganda:** P. Mugenyi, V. Musiime, R. Keishanyu, V. D. Afayo, J. Bwomezi, J. Byaruhanga, P. Erimu, C. Karungi, H. Kizito, W. S. Namala, J. Namusanje, R. Nandugwa, T. K. Najjuko, E. Natukunda, M. Ndigendawani, S. O. Nsiyona, K. Robinah, B. Bainomuhwezi, D. Sseremba, J. Tezkyabbiri, C. S. Tumusiime, A. Balaba, A. Mugumya. **University of Zimbabwe, Harare, Zimbabwe:** K. J. Nathoo, M. F. Bwakura-Dangarembizi, E. Chidziva, T. Mhute, T. Vhembo, R. Mandidewa, D. Nyoni, G. C. Tinago, J. Bhiri, D. Muchabaiwa, M. Phiri, V. Masore, C. C. Marozva, S. J. Maturure, S. Tsikirayi, L. Munetsi, K. M. Rashirai, J. Steamer, R. Nhema, W. Bikwa, B. Tambawoga, E. Mufuka. **Baylor College of Medicine Children's Foundation Uganda, Mulago Hospital, Kampala, Uganda:** A. Kekitiinwa, P. Musoke, S. Bakeera-Kitaka, R. Namuddu, P. Kasirye, A. Babirye, J. Asello, S. Nakalanzi, N. C. Ssemambo, J. Nakafeero, J. Tikabibamu, G. Musoba, J. Ssanyu, M. Kisekka. **MRC Clinical Trials Unit, London, UK:** D. M. Gibb, M. J. Thomason, A. S. Walker, A. D. Cook, B. Naidoo, M. J. Spyer, C. Male, A. J. Glabay, L. K. Kendall, A. Prendergast.

Independent ARROW Trial Monitors: I. Machingura, S. Ssenyonjo. **Trial Steering Committee:** I. Weller (Chair), E. Luyirika, H. Lyall, E. Malianga, C. Mwansambo, M. Nyathi, F. Miiro, D. M. Gibb, A. Kekitiinwa, P. Mugenyi, P. Munderi, K. J. Nathoo. **Observers:** S. Kinn, M. McNeil, M. Roberts, W. Snowden. **Data and Safety Monitoring Committee:** A. Breckenridge (Chair), A. Pozniak, C. Hill, J. Matenga, J. Tumwine, A. S. Walker. **Endpoint Review Committee:** G. Tudor-Williams (Chair), H. Barigye, H. A. Mujuru, G. Ndeezi.

Author Contributions

Conceived and designed the experiments: RC NK MQP JL ASW DMG RT. Analyzed the data: MQP JL. Wrote the first draft of the manuscript: MQP JL. Contributed to the writing of the manuscript: MQP JL RC NK ASW DMG VM AP KN AK PNN RT. ICMJE criteria for authorship read and met: MQP JL VM AP KN AK PNN RT DMG ASW NK RC. Agree with manuscript results and conclusions: MQP JL VM AP KN AK PNN RT DMG ASW NK RC. Enrolled patients: VM KN PNN AK.

References

1. Joint United Nations Programme on HIV/AIDS (2012) UNAIDS report on the global AIDS epidemic. Geneva: Joint United Nations Programme on HIV/AIDS.
2. Sutcliffe CG, van Dijk JH, Bolton C, Persaud D, Moss WJ (2008) Effectiveness of antiretroviral therapy among HIV-infected children in sub-Saharan Africa. *Lancet Infect Dis* 8: 477–489.

3. World Health Organization (2010) Antiretroviral therapy of HIV infection in infants and children: towards universal access: recommendations for a public health approach—2010 revision. Geneva: World Health Organization.
4. Sauvageot D, Schaefer M, Olson D, Pujades-Rodriguez M, O'Brien DP (2010) Antiretroviral Therapy Outcomes in Resource-Limited Settings for HIV-Infected Children 5 Years of Age. *Pediatrics* 125: e1039–e1047.
5. de Beaudrap P, Rouet F, Fassinou P, Kouakoussui A, Mercier S, et al. (2008) CD4 cell response before and after HAART initiation according to viral load and growth indicators in HIV-1-infected children in Abidjan, Cote d'Ivoire. *J Acquir Immune Defic Syndr* 49: 70–76.
6. Rouet F, Fassinou P, Inwoley A, Anaky M-F, Kouakoussui A, et al. (2006) Long-term survival and immuno-virological response of African HIV-1-infected children to highly active antiretroviral therapy regimens. *AIDS* 20: 2315–2319.
7. Kariyo PC, Mbukenakamwe MJ, Baramperanye E, Barutwanayo L, Nyatanyi J, et al. (2006) Antiretroviral therapy in a hundred Burundian children with AIDS [abstract]. Abstract number MOPE0208. AIDS 2006—XVI International AIDS Conference; 13–18 Aug 2006; Toronto, Ontario, Canada.
8. Anabwani GM, Lowenthal MM, Jibril HB, Sechele M, Schwarzwald H, et al. (2006) Long term response to highly active antiretroviral therapy among treatment naïve children in Botswana [abstract]. Abstract number MOPE0261. AIDS 2006—XVI International AIDS Conference; 13–18 Aug 2006; Toronto, Ontario, Canada.
9. DART Trial Team (2010) Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* 375: 123–31.
10. Achhra AC, Phanuphak P, Amin J (2011) Long-term immunological outcomes in treated HIV-infected individuals in high-income and low-middle income countries. *Curr Opin HIV AIDS* 6: 258–265.
11. Gibb DM, Newberry A, Klein N, De Rossi A, Grosch-Woerner I, et al. (2000) Immune repopulation after HAART in previously untreated HIV-1-infected children. *Lancet* 355: 1331–1332.
12. Resino S, Seoane E, Pérez A, Ruiz-Mateos E, Leal M, et al. (2006) Different profiles of immune reconstitution in children and adults with HIV-infection after highly active antiretroviral therapy. *BMC Infect Dis* 6: 112.
13. Sleasman JW, Nelson RP, Goodenow MM, Wilfred D, Hutson A, et al. (1999) Immune reconstitution after ritonavir therapy in children with human immunodeficiency virus infection involves multiple lymphocyte lineages. *J Pediatr* 134: 597–606.
14. The Collaboration of Observational HIV Epidemiological Research Europe (COHERE) Study Group, Sabin CA, Smith CJ, d'Arminio Monforte A, Battegay M, et al. (2008) Response to combination antiretroviral therapy: variation by age. *AIDS* 22: 1463–1473.
15. Bains I, Thiebaut R, Yates AJ, Callard R (2009) Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *J Immunol* 183: 4329–4336.
16. De Rossi A, Walker AS, Klein N, Forni DD, King D, et al. (2002) Increased thymic output after initiation of antiretroviral therapy in human immunodeficiency virus type 1-infected children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *J Infect Dis* 186: 312–320.
17. Steinmann GG, Klaus B, Muller-Hermelink H-K (1985) The involution of the ageing human thymic epithelium is independent of puberty. *Scand J Immunol* 22: 563–575.
18. Thiebaut R, Jacqmin-Gadda H, Walker S, Sabin C, Prins M, et al. (2006) Determinants of response to first HAART regimen in antiretroviral-naïve patients with an estimated time since HIV seroconversion. *HIV Med* 7: 1–9.
19. Lewis J, Walker AS, Castro H, De Rossi A, Gibb DM, et al. (2012) Age and CD4 count at initiation of antiretroviral therapy in HIV-infected children: effects on long-term T-cell reconstitution. *J Infect Dis* 205: 548–556.
20. ARROW Trial Team, Kekitiinwa A, Cook A, Nathoo K, Mugenyi P, et al. (2013) ARROW: a 5-year randomised factorial trial of routine vs clinically driven laboratory monitoring and first-line antiretroviral therapy strategies in HIV-infected African children. *Lancet* 381:1391–1403.
21. National Institutes of Health Division of AIDS (2009) Division of AIDS table for grading the severity of adverse events, version 1.0, December, 2004; clarification August 2009. Available: http://rsc.tech-res.com/Document/safetyandpharmacovigilance/Table_for_Grading_Severity_of_Adult_Pediatric_Adverse_Events.pdf. Accessed 24 September 2013.
22. Hulstaert F, Hannel I, Deneys V, Munhyeshuli V, Reichert T, et al. (1994) Age-related changes in human blood lymphocyte subpopulations. *Clin Immunol Immunopathol* 70: 152–158.
23. Hueenecke S, Behl M, Fadler C, Zimmermann SY, Bochenek K, et al. (2008) Age-matched lymphocyte subpopulation reference values in childhood and adolescence: application of exponential regression analysis. *Eur J Haematol* 80: 532–539.
24. Wade AM, Ades AE (1994) Age related reference ranges: significance tests for models and confidence intervals for centiles. *Stat Med* 13: 2359–2367.
25. Patel K, Hernan MA, Williams PL, Seeger JD, McIntosh K, et al. (2008) Long-term effects of highly active antiretroviral therapy on CD4+ cell evolution among children and adolescents infected with HIV: 5 years and counting. *Clin Infect Dis* 46: 1751–1760.
26. Resino S, Correa R, Bellon JM, Sanchez-Ramon S, Muñoz-Fernández MÁ (2002) Characterizing immune reconstitution after long-term highly active antiretroviral therapy in pediatric AIDS. *AIDS Res Hum Retroviruses* 18: 1395–1406.
27. HIV Paediatric Prognostic Markers Collaborative Study, Boyd K, Dunn DT, Castro H, Gibb DM, et al. (2010) Discordance between CD4 cell count and CD4 cell percentage: implications for when to start antiretroviral therapy in HIV-1 infected children. *AIDS* 24: 1213–1217.
28. Cole TJ, Freeman JV, Preece MA (1998) British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 17: 407–429.
29. R Development Core Team (2011) R: a language and environment for statistical computing [computer program]. Vienna: R Foundation for Statistical Computing.
30. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2011) nlme: linear and nonlinear mixed effects models. Vienna: R Foundation for Statistical Computing.
31. Wolfram Research (2008) Mathematica 7.0 [computer program]. Champaign (Illinois): Wolfram Research.
32. Pakker NG, Notermans DW, de Boer RJ, Roos MTL, de Wolf F, et al. (1998) Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nat Med* 4: 208–214.
33. Walker AS, Doerholt K, Sharland M, Gibb DM, for the Collaborative HIV Paediatric Study (CHIPS) Steering Committee (2004) Response to highly active antiretroviral therapy varies with age: the UK and Ireland Collaborative HIV Paediatric Study. *AIDS* 18: 1915–1924.
34. van Rossum AMC, Scherpbier HJ, van Lochem EG, Pakker NG, Sliker WAT, et al. (2001) Therapeutic immune reconstitution in HIV-1-infected children is independent of their age and pretreatment immune status. *AIDS* 15: 2267–2275.
35. Hainaut M, Ducarme M, Schandene L, Peltier CA, Marissens D, et al. (2003) Age-related immune reconstitution during highly active antiretroviral therapy in human immunodeficiency virus type 1-infected children. *Pediatr Infect Dis J* 22: 62–69.
36. The Antiretroviral Therapy Cohort Collaboration (ART-CC), Lanoy E, May M, Mocroft A, Phillip A, et al. (2009) Prognosis of patients treated with cART from 36 months after initiation, according to current and previous CD4 cell count and plasma HIV-1 RNA measurements. *AIDS* 23: 2199–2208.
37. Lichtenstein KA, Armon C, Buchacz K, Chmiel JS, Buckner K, et al. (2010) Low CD4+ T cell count is a risk factor for cardiovascular disease events in the HIV Outpatient Study. *Clin Infect Dis* 51: 435–447.
38. Triant VA, Regan S, Lee H, Sax PE, Meigs JB, et al. (2010) Association of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. *Lancet Oncol* 10: 1152–1159.
40. Thiebaut R, Jacqmin-Gadda H, Lepout C, Katlama C, Costagliola D, et al. (2003) Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left censoring of HIV RNA measures. *J Biopharm Stat* 13: 271–282.

Editors' Summary

Background. Worldwide, about 3.3 million children under 15 years old are infected with HIV, the virus that causes AIDS. More than 90% of these children live in sub-Saharan Africa, where nearly 600 children become infected with HIV every day, usually acquiring the virus from their mother during pregnancy, birth, or breastfeeding. HIV gradually reduces the numbers of CD4 lymphocytes in the immune system, leaving infected individuals susceptible to other infections. HIV infection can be kept in check but not cured with antiretroviral therapy (ART)—cocktails of drugs that have to be taken every day throughout life. ART reduces the amount of virus in the blood (viral load), which allows the immune system to recover (long-term immune reconstitution). Unfortunately, ART is very expensive, but concerted international efforts over the past decade mean that about a third of children who need ART are now receiving it, including half a million children in sub-Saharan Africa.

Why Was This Study Done? World Health Organization (WHO) guidelines recommend initiation of ART at age-related CD4 cell count thresholds based on the risk of short-term disease progression. The guidelines recommend that all HIV-positive children under two years old begin ART as soon they receive a diagnosis of HIV infection. For children aged 2–5 years, ART initiation is recommended once the CD4 count drops below 750 cells/ μ l blood, whereas for older children the threshold for ART initiation is 350 CD4 cells/ μ l. Because of improved ART coverage, many more HIV-infected children now survive into adulthood than in the past. It is therefore important to know how the timing of ART initiation in childhood affects long-term immune reconstitution. Unfortunately, although several studies have examined the effect of ART on immune reconstitution in adults, the results of these studies cannot be extrapolated to children because of age-related differences in immune reconstitution. In this cohort-based modelling study, the researchers investigate long-term CD4 recovery in a cohort (group) of HIV-infected children initiating ART in Uganda and Zimbabwe, and present statistical models that predict patterns of long-term CD4 status based on age and CD4 count at ART initiation.

What Did the Researchers Do and Find? To investigate long-term CD4 reconstitution in children, the researchers used CD4 counts collected during the ARROW trial, a study designed to investigate monitoring strategies during first-line ART in 1,206 HIV-positive children. In three-quarters of the children, CD4 reconstitution following ART initiation was asymptotic—CD4 counts increased rapidly immediately after ART initiation, then slowed before eventually reaching a constant level of about 80% of the CD4 count expected in an uninfected child of the same age. Using a nonlinear mixed-effects statistical model that fitted this pattern of immune reconstitution, the researchers predicted CD4 trajectories for children starting ART at different ages and with different CD4 counts. Higher long-term counts were predicted for children starting ART earlier and with higher CD4 counts. Thus, to achieve a CD4 count greater than 700 cells/ μ l at age 20 years, CD4 counts of at least 96 cells/ μ l, 130 cells/ μ l, and 557 cells/

μ l are needed for children aged two, five, and 12 years, respectively, when they initiate ART. Qualitatively distinct reconstitution patterns were seen in the remaining children in the study.

What Do These Findings Mean? These findings suggest that young HIV-positive, ART-naïve children can achieve high CD4 counts in later life, provided ART is initiated as recommended in the current WHO guidelines. However, the recommended CD4 count thresholds for ART initiation are unlikely to maximize immune reconstitution in treatment-naïve children over ten years old. Rather, these findings suggest that ART initiation should be considered in these older children when their CD4 count is still high—even though they have a low risk of immediate disease progression—in order to achieve higher long-term CD4 levels. The omission of viral load measurements in the researchers' model may limit the accuracy of these findings. Moreover, although the predictions made by the model apply to children who will go on to experience asymptotic recovery, they are less relevant to those with different recovery profiles, who cannot currently be accurately identified. Further exploration of the immunological mechanisms underlying the CD4 recovery profiles described here should improve our understanding of the factors that determine the response of HIV-positive children to ART and provide information to guide the management of HIV infections in children.

Additional Information. Please access these websites via the online version of this summary at <http://dx.doi.org/10.1371/journal.pmed.1001542>.

- Information is available from the US National Institute of Allergy and Infectious Diseases on HIV infection and AIDS
- NAM/aidsmap provides basic information about HIV/AIDS and summaries of recent research findings on HIV care and treatment
- Information is available from Avert, an international AIDS charity, on many aspects of HIV/AIDS, including information on HIV and AIDS in Africa and on HIV infection in children (in English and Spanish)
- The *UNAIDS World AIDS Day Report 2012* provides up-to-date information about the AIDS epidemic and efforts to halt it; the UNAIDS's *2013 Progress Report on the Global Plan* provides information on progress towards eliminating new HIV infections among children by 2015
- The World Health Organization provides information about universal access to AIDS treatment (in several languages); its 2010 guidelines for ART in infants and children can be downloaded
- Information about the ARROW trial is available
- Personal stories about living with HIV/AIDS, including stories from young people infected with HIV, are available through Avert, through Nam/aidsmap, and through the charity website Healthtalkonline

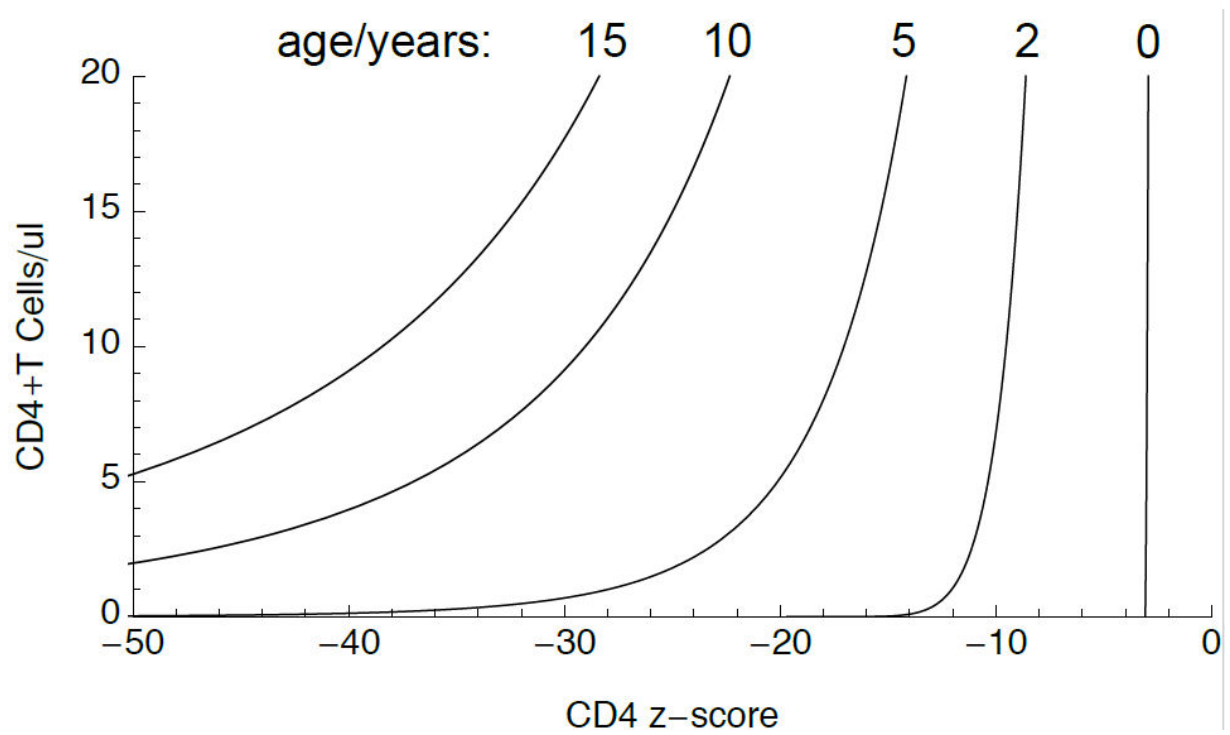


Figure S1. Relationship between very low CD4 count and CD4-for-age z-score in children of different ages.

Table S1: Univariate comparisons between individual non-asymptotic response sub-groups and the large (n=914) group of asymptotic responders.

Characteristics	CD4-for-age decreasing (n=23) (value (95% CI); p-value)	CD4-for-age increasing (n=79) (value (95% CI); p-value)	No significant change in CD4-for-age (n=153) (value (95% CI); p-value)	≤2 measurements (n=37) (value (95% CI); p-value)
Pre-ART age (years)	+3.2 (+1.1, +5.1); 0.002	+0.13 (-0.67,+0.93); 0.74	+1.6 (+0.8, +2.4); <0.001	-0.12 (-1.32, +1.27); 0.84
Pre-ART CD4 count (cells/μL)	-25 (-167, +115); 0.70	-7 (-84,+71); 0.80	+134 (+69, +199); <0.001	-111 (-220, -11); 0.02258
Pre-ART CD4-for-age (count/healthy count)	+0.01 (-0.09, +0.10); 0.83	+0.00 (-0.05, +0.05); 1.00	+0.14 (+0.09, +0.18); <0.001	+0.10 (-0.17, -0.03); 0.002
Pre-ART CD4-for-age ln(count/healthy count)	+0.05 (-0.42, +0.51); 0.83	+0.00 (-0.22, +0.22); 1.00	+0.46 (+0.31, +0.60); <0.001	-0.64 (-1.24, -0.22); 0.002
Pre-ART weight-for-age z-score	+0.29 (-0.28, +0.88); 0.30	+0.29 (-0.08, +0.66); 0.14	+0.51 (+0.25, +0.77); <0.001	-1.2 (-1.8, +0.7); <0.001
Pre-ART height-for-age z-score	+0.40 (-0.31, +1.01); 0.25	+0.34 (-0.01, +0.69); 0.060	+0.52 (+0.27, +0.76); <0.001	-0.10 (-0.58, +0.38); 0.69
Sex				
female vs male	0.99 (0.39, 2.47); 1.00	1.05 (0.64, 1.71); 0.91	1.35 (0.94, 1.93); 0.10	0.74 (0.36, 1.53); 0.40
3 vs 2	0.95 (0.29, 2.70); 1.00	1.68 (0.99, 2.82); 0.048	0.80 (0.52, 1.21); 0.28	1.66 (0.58, 5.82); 0.37
4 vs 2	0.79 (0.18, 3.87); 0.74	2.40 (1.00, 6.62); 0.048	1.71 (0.84, 3.71); 0.15	5.30 (1.76, 19.22); <0.001
Monitoring randomisation				
CDM vs LCM	0.43 (0.15, 1.13); 0.089	0.92 (0.56, 1.49); 0.73	0.86 (0.60, 1.23); 0.43	0.50 (0.23, 1.06); 0.062
First-line ART randomisation				
B vs A	0.77 (0.25, 2.29); 0.63	1.93 (1.06, 3.62); 0.024	1.26 (0.79, 2.02); 0.31	0.49 (0.16, 1.34); 0.17
C vs A	1.17 (0.33, 4.27); 1.00	1.46 (0.82, 2.64); 0.18	0.76 (0.49, 1.17); 0.21	1.23 (0.54, 2.82); 0.70

Table S1 (following): Univariate comparisons between individual non-asymptotic response sub-groups and the large (n=914) group of asymptotic responders.

Characteristics	CD4-for-age decreasing (n=23) (value (95% CI); p-value)	CD4-for-age increasing (n=79) (value (95% CI); p-value)	No significant change in CD4-for-age (n=153) (value (95% CI); p-value)	≤2 measurements (n=37) (value (95% CI); p-value)
WHO class 3/4 events	11.1 (4.2, 28.7); <0.001	3.67 (1.86, 6.93); <0.001	2.30 (1.28, 4.00); 0.004	2.80 (0.91, 7.18); 0.035
Deaths	44.7 (11.2, 174.4); <0.001	1.66 (0.04, 13.19); 0.49	7.13 (2.22, 23.47); <0.001	222 (778, 4110); <0.001

The upper rows of the table (pre-ART age to pre-ART height-for-age) refer to comparisons of continuous variables. The values given are estimated (95% confidence interval, 95% CI) differences between each non-asymptotic response group and the large, asymptotic group, followed by a p-value for a Wilcoxon rank sum test with the null hypothesis of zero difference. For example, the age of children with decreasing CD4-for-age was estimated to be 3.2 (95% CI 1.1 to 5.1) years older than the children with asymptotic reconstitution, corresponding to a p-value of 0.00182. The lower rows (sex to first-line ART randomisation) refer to categorical variables, and the values given are an odds ratio with 95% CI and p-value for Fisher's exact test, with the null hypothesis of odds ratio 1. None of the children with only one or two CD4 counts available

Table S2: Additional predictors of CD4-for-age profiles in the asymptotic recovery group.

		Estimate	Standard Error	p-value (Likelihood ratio test)
Predictors of intercept				
Sex	Female	0.137	0.068	0.018
WHO stage	Stage 3	0.0679	0.0659	0.079
	Stage 4	-0.121	0.099	0.046
Monitoring strategy	CDM	0.104	0.065	0.029
First-line ART strategy	Arm-B	-0.0409	0.0823	0.21
	Arm-C	-0.0791	0.0845	
Predictors of asymptote				
WHO stage	Stage 3	0.0142	0.0257	0.12
	Stage 4	0.0809	0.0367	0.012
Pre-ART weight-for-age	Per unit higher	-0.0220	0.0083	0.034
Monitoring strategy	CDM	-0.0197	0.0255	0.078
Predictors of c				
Pre-ART age	Per year older	7.64×10^{-4}	2.50×10^{-4}	0.034
Sex	Female	8.46×10^{-4}	1.81×10^{-3}	0.18
WHO stage	Stage 3	1.53×10^{-3}	1.71×10^{-3}	0.11
	Stage 4	-9.68×10^{-4}	2.24×10^{-3}	0.089
Pre-ART weight-for-age	Per unit higher	9.31×10^{-4}	5.11×10^{-4}	0.003
Monitoring strategy	CDM	-2.29×10^{-3}	1.70×10^{-3}	0.024
First-line ART strategy	Arm-B	-5.65×10^{-3}	2.18×10^{-3}	0.020
	Arm-C	-2.94×10^{-3}	2.33×10^{-3}	

We examined the effects of including additional predictors of parameter values in the nonlinear mixed-effects model described in Table 2. Predictive effects on each of the three model parameters were included one at a time. The table gives the estimated and standard error effect sizes, and the results of likelihood ratio tests comparing each extended model to the model in Table 2.

5.5. Discussion

Les principaux résultats de l'étude sont rapportés ci-après. Au total, les données de 1206 enfants infectés par le VIH, ayant initié un traitement antirétroviral en Uganda et Zimbabwe, ont été analysées (âge de 0,4 à 17,6 ans ; lymphocytes T CD4+ mesurés en cellules/ μ L toutes les 12 semaines). L'application du modèle de reconstitution asymptotique, via les modèles non linéaires mixtes, a permis d'identifier deux groupes distincts. Parmi les 1206 enfants, 914 (76 %) présentaient une phase de reconstitution lymphocytaire asymptomatique, avec une augmentation rapide du nombre de lymphocytes CD4+, ajusté sur l'âge, après mise sous traitement. Un nombre de lymphocytes T CD4+, ajusté sur l'âge, plus élevé à long terme était rapporté pour les enfants ayant initié un traitement antirétroviral à un âge plus jeune et ceux ayant initié un traitement avec un nombre de lymphocytes T CD4+ initial plus élevé ($p < 0,001$). En particulier, les prédictions de reconstitution à long terme, basées sur l'âge au moment de la mise sous traitements antirétroviraux, suggéraient que les seuils préconisés par l'OMS pour initier le traitement chez des enfants de plus de 5 ans résulteraient en des nombres de lymphocytes T CD4+ plus bas à l'âge adulte que pour les enfants de moins de 5 ans, ceci étant plus marqué encore chez les enfants de plus de 10 ans. Pour les 292 (24 %) autres enfants, quatre modèles de reconstitution qualitativement distincts ont été observés :

- i) 23 (8 %) enfants présentaient une diminution dans le temps du nombre de lymphocytes T CD4+, ajusté sur l'âge, témoignant d'un probable échec du traitement,
- ii) 79 (27 %) enfants présentaient une reconstitution immunitaire lente, sans augmentation rapide du nombre de lymphocytes T CD4+, ajusté sur l'âge, après mise sous traitement mais avec un nombre de lymphocytes T CD4+, ajusté sur l'âge, avoisinant l'asymptote sur le long terme,
- iii) 153 (52 %) enfants présentaient un nombre de lymphocytes T CD4+, ajusté sur l'âge, inférieur à l'asymptote, témoignant d'une incapacité potentielle du système immunitaire à reconstituer le compartiment lymphocytaire T CD4+ sur le long terme,
- iiii) 37 (13 %) enfants présentaient des mesures insuffisantes de lymphocytes T CD4+ pour identifier des profils de reconstitution (≤ 2 mesures).

En raison du coût et de la complexité d'acquisition de la charge virale en temps réel dans les pays à ressources limitées, les recommandations OMS ne préconisent pas la mesure de la charge virale en systématique, ni son utilisation dans les décisions pour initier une mise sous traitements antirétroviraux. L'absence de cette variable dans la modélisation est une limite de l'étude.

Ainsi, bien que les plus jeunes enfants soient à haut risque de progression de la maladie, ils ont un bon potentiel pour reconstituer un nombre de lymphocytes T CD4+ élevé à l'âge adulte. En revanche, selon l'initiation des traitements antirétroviraux préconisée par les recommandations de l'OMS, pour des enfants de plus de 10 ans en particulier sans risque de progression à court terme de la maladie, différer la mise sous traitement pourrait avoir des conséquences considérables sur la capacité de reconstitution du système immunitaire sur le long terme. Des explorations plus approfondies, pour améliorer la compréhension des mécanismes immunologiques responsables des différents profils de reconstitution des lymphocytes T CD4+, sont nécessaires pour permettre d'orienter les stratégies de mise en place des traitements antirétroviraux, afin d'optimiser la reconstitution du système immunitaire à court et long terme.

Plusieurs points méritent d'être soulignés quant à la modélisation réalisée pour ce travail. Dans les analyses préliminaires, le modèle asymptotique proposé, appliqué à la base de données entière des 1206 enfants, ne convergait pas. Ainsi, afin d'identifier les enfants qui suivaient le profil de reconstitution asymptotique supposé, les profils de reconstitution des enfants ont été modélisés séparément (fonction `nlsList` du package R `nlme`), permettant ainsi d'identifier le groupe des 974 enfants avec un profil de reconstitution asymptotique et celui des 292 ne présentant pas ce profil. Parmi les 292 enfants restants, l'étude des profils de reconstitution a été réalisée à l'aide de modèles linéaires mixtes pour identifier ceux qui avaient un profil de reconstitution linéaire ascendant, descendant ou stable.

Les modèles non linéaires mixtes utilisés dans le cadre de ce travail ne sont pas des méthodes nouvelles. Plus que la mise en œuvre et l'application de ces modèles en eux-mêmes, l'originalité de ce travail réside plus dans la réflexion préalable menée autour du choix du modèle de reconstitution immunitaire asymptotique utilisant une fonction exponentielle. L'exploration complémentaire des données à l'aide d'un modèle linéaire mixte à changement de pente, tel qu'utilisé chez l'adulte [Thiébaud et al., 2006], a révélé une moins bonne adéquation aux données, confirmant l'emploi plus optimal du modèle de reconstitution asymptotique chez l'enfant.

Dans le cadre de ce travail, l'application du modèle de reconstitution immunitaire asymptotique nous a permis d'améliorer les connaissances sur la reconstitution immunitaire à long terme, après mise sous traitements antirétroviraux, chez les enfants infectés par le VIH dans les pays à ressources limitées. Les résultats concernant le nombre de lymphocytes T

CD4+ plus élevé à long terme chez les enfants ayant initié un traitement antirétroviral à un âge plus jeune sont à souligner.

5.6. Conclusion et perspectives

Les enfants infectés par le VIH survivant de plus en plus à l'âge adulte, deux objectifs spécifiques ont guidé la réalisation de cette analyse supplémentaire sur l'étude de la reconstitution immunitaire à long terme, non initialement prévue au protocole ARROW. Tout d'abord, un modèle caractérisant des profils distincts de reconstitution immunitaire à long terme selon les caractéristiques des enfants au moment de la mise sous traitement antirétroviral, incluant l'âge et le nombre de lymphocytes T CD4+, constituait un outil pertinent pour guider les stratégies thérapeutiques, incluant les seuils définis pour l'initiation du traitement. Ensuite, ce modèle devait constituer une base de réflexion ultérieure pour les mécanismes biologiques influençant les différents profils de reconstitution observés. Le choix d'un tel modèle s'est basé sur la nécessité de prendre en compte les données longitudinales, l'hypothèse de reconstitution immunitaire à long terme attendue sous traitement dans la problématique à l'étude et la compréhension des mécanismes biologiques sous-jacents. Ainsi, la réflexion menée autour du choix d'un modèle est une étape majeure pour prendre en compte à la fois la structure même des données et les mécanismes qui sous-tendent la variabilité de ces données. La fonction exponentielle proposée, en plus de s'appuyer sur des paramètres cliniquement et biologiquement interprétables, permet de faire des liens pour l'élaboration d'un modèle mécanistique de la reconstitution immunitaire des lymphocytes T CD4+. L'élaboration d'un tel modèle devrait en particulier prendre en compte la dynamique des lymphocytes T naïfs, liée à la thymopoïèse active chez l'enfant, qui sous-tend le profil de reconstitution immunitaire sous traitement. Puis, un modèle mathématique basé sur des équations différentielles pourrait être élaboré [Perelson, 2002].

En particulier, dans l'essai clinique ARROW, les données d'une sous-étude concernant 199 enfants, pour lesquels les lymphocytes T CD4+ naïfs et mémoires ont été mesurés, pourraient permettre des explorations des mécanismes immunologiques sous-tendant les profils de reconstitution observés.

6. DISCUSSION

L'épidémiologie doit évoluer dans un contexte de données variées, complexes et multiples [Ness, 2007 ; Palmer, 2004]. Dans le contexte appliqué aux troubles immunologiques des maladies du système immunitaire, les difficultés méthodologiques en termes d'analyse, de modélisation et d'inférence générées par les données d'immunologie doivent être prise en compte.

Cette thèse a ainsi porté sur les aspects méthodologiques liés à l'analyse des biomarqueurs dans les études épidémiologiques, avec une application concrète pour l'étude des troubles immunologiques des maladies du système immunitaire. Elle s'est positionnée à l'interface entre l'épidémiologie clinique, les biostatistiques et l'immunologie. Notre approche a consisté en l'utilisation de méthodes intégratives. Nous avons mis en œuvre des méthodes variées, qui ont été pertinentes pour répondre aux problématiques cliniques formulées.

En section 3, nous avons proposé des méthodes exploratoires multidimensionnelles pour prendre en compte des données immunologiques multiples de cytométrie en flux. L'application de ces méthodes, qui supposent l'absence d'hypothèses à priori au moment de l'analyse des données, nous a permis d'extraire les informations sur les marqueurs les plus pertinents parmi l'ensemble des marqueurs mesurés. Cette analyse a permis dans les suites de guider le rationnel méthodologique pour la mise en œuvre d'une méthode de régression multivariable et d'une méthode évaluant l'apport d'un marqueur pour améliorer une classification existante (index de reclassification), illustrant ainsi la complémentarité potentielle des méthodes non supervisées et supervisées. Dans le cadre de la problématique des déficits immunitaires humoraux primitifs de l'adulte, l'application de ces méthodes nous a permis de souligner l'importance de l'activation des lymphocytes T, et en particulier celle du marqueur CD8+HLA-DR+, qui pourrait permettre d'améliorer la discrimination des patients selon des profils cliniques distincts.

En section 4, nous avons proposé des modèles d'équations structurelles, comme alternative aux méthodes classiques de régression linéaire multivariable, pour étudier des relations complexes entre variables dans le cadre de processus multifactoriels. Ces modèles, basés sur l'élaboration de diagrammes de chemin et le test d'une série d'hypothèses quant aux associations supposées, permettent l'intégration de multiples variables décrivant différentes dimensions et la modélisation explicite de leurs relations. Les modèles d'équations structurelles constituent une méthode prometteuse pour les épidémiologistes, les modèles

utilisant des variables latentes étant relativement rares en épidémiologie. Les perspectives sont pourtant larges, l'extension de l'utilisation de ces modèles dans le cadre d'analyse de données longitudinales ouvrant notamment des perspectives vers des approches d'inférence causale. Néanmoins, la mise en œuvre de ces modèles n'est pas triviale et nécessite également de bien identifier les questions clinico-épidémiologiques pour lesquelles leur emploi peut être d'un apport considérable [Dahly et al., 2009 ; Tu, 2005]. Dans le cadre de la problématique de l'activation immunitaire chronique chez les patients adultes atteints de VIH traités efficacement, l'application des modèles d'équations structurelles a permis une meilleure compréhension des déterminants de cette activation. Elle a en particulier révélé un effet indirect de la réponse immunitaire induite par l'auto-immunité sur l'activation immunitaire chronique via la voie de l'interféron alpha. Ce signal mérite des recherches ultérieures.

En section 5, nous avons proposé une réflexion sur le modèle mathématique le plus adéquat pour modéliser la reconstitution du système immunitaire à long terme, dans l'exemple concret du VIH pédiatrique dans les pays à ressources limitées. Un modèle s'appuyant sur une fonction exponentielle a été retenu. La réflexion pour l'élaboration d'un tel modèle s'est largement basée sur les mécanismes biologiques sous-jacents qui pourraient concourir au profil de reconstitution attendu, sur la nécessité de résultats aisément interprétables cliniquement et biologiquement pour guider les politiques de santé et sur les perspectives d'approche ultérieure mécanistique plus explicative. L'application d'un tel modèle nous a permis de souligner un nombre prédit de lymphocytes T CD4+ à long terme plus élevé pour les enfants ayant initié un traitement antirétroviral à un âge plus jeune et avec un nombre de lymphocytes T CD4+ initial plus élevé. En parallèle, chez les enfants de plus de 10 ans en particulier, différer la mise sous traitement antirétroviral en raison d'un nombre de lymphocytes T CD4+ supérieurs aux seuils préconisés par l'OMS pour mise sous traitement pourrait résulter en une moins bonne reconstitution du système immunitaire à l'âge adulte. Ces résultats peuvent ainsi contribuer à guider les réflexions des experts du domaine [Gibb, 2015]. L'originalité de ce travail a résidé plus dans la réflexion préalable menée autour du choix du modèle plus que dans l'emploi des modèles non linéaires mixtes, soulignant ainsi que l'évolution de l'épidémiologie ne doit pas être associée, de manière systématique, à l'emploi de méthodes biostatistiques et épidémiologiques nouvelles.

Ainsi, ces travaux illustrent que l'augmentation du volume et de la complexité des données doit s'accompagner d'une réflexion plus poussée concernant les méthodes biostatistiques et épidémiologiques les plus adaptées permettant de les analyser, pour ainsi produire de nouvelles connaissances. Cette réflexion doit s'appuyer sur la structure même des données et

sur la compréhension des mécanismes potentiels qui sous-tendent la variabilité de ces données. L'approche multidisciplinaire est fondamentale et obligatoire. Il s'agit d'un travail quotidien de collaborateurs issus d'horizons très différents mais s'intéressant aux autres disciplines. Les épidémiologistes disposent de méthodes biostatistiques et épidémiologiques variées. Une connaissance approfondie des avantages et des limites des méthodes classiques est fondamentale pour identifier l'utilisation adéquate de ces méthodes et si nécessaire, viser à les enrichir par l'acquisition et la mise en œuvre de méthodes plus complexes. Parmi les méthodes intégratives proposées dans cette thèse, si les méthodes d'analyse en composantes principales, de classification hiérarchique ascendante et les modèles d'équations structurelles peuvent être appliquées dans le cadre de données multiples, la question clinique reste majeure pour guider le choix de la méthode à employer dans un tel contexte. Dans l'étude ALTADIH, nous étions confrontés à une problématique de classification et de prédiction de la survenue de complications d'où le choix d'une première étape de réduction de dimension dans un contexte de données multiples par analyse en composantes principales puis d'une deuxième étape par classification hiérarchique ascendante et méthode d'index de reclassification pour sélectionner le marqueur le plus pertinent pour améliorer le schéma de classification existant. Dans l'étude ACTHIV, nous étions confrontés à une approche plus physiopathologique, avec la formulation d'hypothèses potentiellement mécanistiques, d'où le choix des modèles d'équations structurelles pour explorer les relations supposées entre variables. Enfin, concernant les résultats, comme dans toute analyse statistique, il convient de distinguer la significativité statistique de la pertinence clinique des résultats.

Les aspects méthodologiques exposés dans cette thèse contribuent à améliorer les connaissances pour adresser des questions clinico-épidémiologiques complexes dans les troubles immunologiques des maladies du système immunitaire. Les méthodes proposées trouvent particulièrement leur application dans le domaine de l'Immunologie. Tout d'abord, les méthodes d'analyse intégrative telles que les méthodes exploratoires multidimensionnelles ou modèles d'équations structurelles permettent une approche du système immunitaire dans sa globalité. Un autre point est l'utilisation de variables latentes pour mieux caractériser un processus immunologique à partir de plusieurs marqueurs immunologiques supposés mesurer ce processus. Par exemple, la signature interféron alpha est indirectement estimée par l'analyse de l'expression des gènes de l'interféron alpha. Cette signature est majoritairement impliquée dans de nombreuses maladies auto-immunes, telles que le lupus érythémateux systémique, la sclérodermie systémique, la polyarthrite rhumatoïde, le syndrome de Goujerot-

Sjögren [Banchereau et Pascual., 2006 ; Pascual et Banchereau, 2012 ; Sozzani et al., 2010]. Des efforts continus sont ainsi faits pour mieux identifier les gènes de la voie interféron [Schneider et al., 2014]. L'utilisation de variables latentes, avec extension pour analyse de données longitudinales, pourrait être d'un apport considérable pour modéliser la signature interféron alpha via l'expression des nombreux gènes existants. Cette approche rejoint l'intérêt des approches par groupe de gènes, qui sont définies comme un ensemble de gènes étant à priori co-régulés ou liés à une même fonction biologique [Ashburner et al., 2000 ; Chaussabel et al., 2008 ; Hejblum et al., 2015 ; Kanehisa et Goto, 2000]. Enfin, l'intégration de données longitudinales avec un modèle basé sur des connaissances biologiques et une fonction exponentielle ouvre des perspectives vers des approches plus mécanistiques.

D'autres questions méthodologiques, non abordées dans ce travail de thèse, sont loin d'être résolues et des travaux supplémentaires seront nécessaires. Sinha et al ont proposé une revue des problèmes analytiques rencontrés avec un grand jeu de données [Sinha et al., 2009]. En particulier, dans un contexte de données multiples, la problématique des données manquantes reste un vaste champ de recherche. Les méthodes exploratoires multidimensionnelles et les modèles d'équation structurelles ont été réalisés pour les patients pour lesquels l'ensemble des marqueurs considérés étaient disponibles. Les analyses cas complets doivent ainsi vérifier que les patients exclus ne diffèrent pas de l'échantillon d'origine, autrement leur exclusion peut générer des biais. Si nécessaire, la réflexion ultérieure doit porter sur la nature de ces données manquantes, la méthode d'imputation multiple la plus adaptée parmi les nombreuses existantes et prendre un compte des temps de calcul potentiellement considérables dans le cadre de données multiples. Enfin, dans le cadre des résultats formulés sur un grand nombre de marqueurs et potentiellement peu de patients, l'évaluation de la robustesse des résultats par des méthodes de ré-échantillonnage pourrait être envisagée.

Dans le cadre de ce travail de thèse, les applications ont été centrées sur les troubles immunologiques des maladies du système immunitaire. Néanmoins, les données multidimensionnelles sont des données de plus en plus fréquentes en santé. Ainsi, le raisonnement méthodologique présenté pourrait également être utile pour l'analyse d'autres données.

D'un point de vue biologique, la réflexion développée sur les déterminants qui concourent à une activation immunitaire chronique chez des patients atteints de VIH traités efficacement pourrait permettre d'améliorer les connaissances sur d'autres pathologies. Pour faire le lien avec la problématique développée en section 3, des articles récents rapportent le rôle de

l'interféron alpha [Park et al., 2013] et l'implication du cytomégalovirus dans les déficits immunitaires humoraux primitifs [Marashi et al., 2012]. De plus, le parallèle entre l'infection à VIH et le vieillissement est un sujet d'intérêt croissant. Les patients atteints de VIH souffrent de nombreuses altérations (athérosclérose, cancer, ostéoporose, troubles cognitifs) qui sont habituellement observées au cours du processus de vieillissement de l'homme [Effros et al., 2008]. Ces manifestations sont rapportées comme la conséquence de l'activation immunitaire et de l'inflammation, éléments majeurs de la pathogenèse du VIH [Deeks, 2011]. Ainsi, le VIH chronique constitue un modèle de vieillissement accéléré et d'inflammation. Une meilleure compréhension des déterminants qui concourent à une activation immunitaire chronique chez des patients atteints de VIH pourrait ainsi permettre d'élargir les connaissances à d'autres domaines.

7. CONCLUSION

Les maladies du système immunitaire constituent un bon exemple de pathologies où les questions clinico-épidémiologiques sont de plus en plus complexes, nécessitant une réflexion avancée sur les méthodes biostatistiques et épidémiologiques les plus adaptées pour répondre à ces questions. D'une part, ces maladies sont multifactorielles et mettent en jeu différents acteurs du système immunitaire et, d'autre part, l'étude des troubles immunologiques qui leur sont associés repose en particulier sur la mesure de marqueurs immunologiques multiples, générant des données multidimensionnelles. Dans ce contexte, l'approche multidisciplinaire et l'analyse de ces données multidimensionnelles de manière adéquate, en prenant en compte les difficultés méthodologiques en termes d'analyse, de modélisation et d'inférence ainsi que les mécanismes biologiques qui sous-tendent la variabilité des données, sont essentielles pour augmenter non seulement la production des connaissances en immunologie clinique mais aussi pour la construction d'un savoir-faire applicable à d'autres thématiques.

Au total, cette thèse d'Université a plusieurs retombées potentielles. D'un point de vue épidémiologique, les méthodes intégratives proposées sont transposables à de nombreux domaines de recherche utilisant des données complexes, multiples et s'intéressant à des processus multifactoriels. D'un point de vue biologique, la réflexion proposée pour les pathologies à l'étude et l'approche du système immunitaire dans sa globalité pourraient permettre d'améliorer les connaissances sur d'autres pathologies.

REFERENCES

- Abad-Fernandez M, Vallejo A, Hernandez-Novoa B, Diaz L, Gutierrez C, Madrid N, Muñoz MA, Moreno S. Correlation between different methods to measure microbial translocation and its association with immune activation in long-term suppressed HIV-1-infected individuals. *J Acquir Immune Defic Syndr* 2013 Oct 1;64(2):149-53.
- Acchioni C, Marsili G, Perrotti E, Remoli AL, Sgarbanti M, Battistini A. Type I IFN--a blunt spear in fighting HIV-1 infection. *Cytokine Growth Factor Rev* 2015 Apr;26(2):143-58.
- Achhra AC, Phanuphak P, Amin J. Long-term immunological outcomes in treated HIV-infected individuals in high-income and low-middle income countries. *Curr Opin HIV AIDS* 2011 Jul;6(4):258-65.
- Aghamohammadi A, Parvaneh N, Rezaei N. Common variable immunodeficiency: a heterogeneous group needs further subclassification. *Expert Rev Clin Immunol* 2009;5(6):629-31.
- Aguilera AM, Escabias M, Valderrama MJ. Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Comput Stat Data Anal* 2006;(50):1905-24.
- Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004Jul;4(7):499-511.
- Al-Herz W, Bousfiha A, Casanova JL, Chatila T, Conley ME, Cunningham-Rundles C, Etzioni A, Franco JL, Gaspar HB, Holland SM, Klein C, Nonoyama S, Ochs HD, Oksenhendler E, Picard C, Puck JM, Sullivan K, Tang ML. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol* 2014 Apr 22;5:162.
- Appay V, Fastenackels S, Katlama C, Ait-Mohand H, Schneider L, Guihot A, Keller M, Grubeck-Loebenstein B, Simon A, Lambotte O, Hunt PW, Deeks SG, Costagliola D, Autran B, Sauce D. Old age and anti-cytomegalovirus immunity are associated with altered T-cell reconstitution in HIV-1-infected patients. *AIDS* 2011 Sep 24;25(15):1813-22.
- Appay V, Sauce D. Immune activation and inflammation in HIV-1 infection: causes and consequences. *J Pathol* 2008 Jan;214(2):231-41.

Arlinghaus A, Lombardi DA, Willetts JL, Folkard S, Christiani DC. A structural equation modeling approach to fatigue-related risk factors for occupational injury. *Am J Epidemiol* 2012 Oct 1;176(7):597-607.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000 May;25(1):25-9.

Avalos M, Pouyes H, Grandvalet Y, Orriols L, Lagarde E. Sparse conditional logistic regression for analyzing large-scale matched data from epidemiological studies: a simple algorithm. *BMC Bioinformatics* 2015;16 Suppl 6:S1.

Babiker AG, Emery S, Fatkenheuer G, Gordin FM, Grund B, Lundgren JD, Neaton JD, Pett SL, Phillips A, Touloumi G, Vjecha MJ; INSIGHT START Study Group. Considerations in the rationale, design and methods of the Strategic Timing of AntiRetroviral Treatment (START) study. *Clin Trials* 2013;10(1 Suppl):5-36.

Bains I, Thiébaud R, Yates AJ, Callard R. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *J Immunol* 2009 Oct 1;183(7):4329-36.

Balestre E, Eholie SP, Lokossue A, Sow PS, Charurat M, Minga A, Drabo J, Dabis F, Ekouevi DK, Thiébaud R; International epidemiological Database to Evaluate AIDS (IeDEA) West Africa Collaboration. Effect of age on immunological response in the first year of antiretroviral therapy in HIV-1-infected adults in West Africa. *AIDS* 2012 May 15;26(8):951-7.

Banchereau J, Pascual V. Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity* 2006 Sep;25(3):383-92.

Barton JC, Bertoli LF, Barton JC. Comparisons of CVID and IgGSD: referring physicians, autoimmune conditions, pneumovax reactivity, immunoglobulin levels, blood lymphocyte subsets, and HLA-A and -B typing in 432 adult index patients. *J Immunol Res* 2014;2014:542706.

Beran TN, Violato C. Structural equation modeling in medical research: a primer. *BMC Res Notes* 2010;3:267.

Blanco P, Palucka AK, Pascual V, Banchereau J. Dendritic cells and cytokines in human inflammatory and autoimmune diseases. *Cytokine Growth Factor Rev* 2008 Feb;19(1):41-52.

Bollen KA. Latent variables in psychology and the social sciences. *Annu Rev Psychol* 2002;53:605-34.

Bollen KA, Noble MD. Structural equation models and the quantification of behavior. *Proc Natl Acad Sci U S A* 2011 Sep 13;108 Suppl 3:15639-46.

Bollen KA, Noble MD, Adair LS. Are gestational age, birth weight, and birth length indicators of favorable fetal growth conditions? A structural equation analysis of Filipino infants. *Stat Med* 2013 Jul 30;32(17):2950-61.

Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, Kazzaz Z, Bornstein E, Lambotte O, Altmann D, Blazar BR, Rodriguez B, Teixeira-Johnson L, Landay A, Martin JN, Hecht FM, Picker LJ, Lederman MM, Deeks SG, Douek DC. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med* 2006 Dec;12(12):1365-71.

Budtz-Jorgensen E, Keiding N, Grandjean P, Weihe P. Estimation of health effects of prenatal methylmercury exposure using structural equation models. *Environ Health* 2002 Oct 14;1(1):2.

Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. *Neuroimage* 2011 Apr 15;55(4):1519-27.

Casin P. *Analyse des données et des panels de données*. Belgique: De Boeck; 1999.

CEREDIH: The French PID study group. The French national registry of primary immunodeficiency diseases. *Clin Immunol* 2010 May;135(2):264-72.

Cha L, Berry CM, Nolan D, Castley A, Fernandez S, French MA. Interferon-alpha, immune activation and immune dysfunction in treated HIV infection. *Clin Transl Immunology* 2014 Feb;3(2):e10.

Chapel H, Lucas M, Lee M, Bjorkander J, Webster D, Grimbacher B, Fieschi C, Thon V, Abedi MR, Hammarstrom L. Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* 2008 Jul 15;112(2):277-86.

Chatenoud L, Bach JF. Immunologie : de la biologie à la clinique. 6e ed. Paris: Médecine Sciences Publications-Lavoisier; 2012.

Chattopadhyay PK, Roederer M. Cytometry: today's technology and tomorrow's horizons. *Methods* 2012 Jul;57(3):251-8.

Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H, Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008 Jul 18; 29(1):150–64.

Chavance M, Escolano S, Romon M, Basdevant A, de Lauzon-Guillain B, Charles MA. Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology. *BMC Med Res Methodol* 2010;10:37.

Chinen J, Shearer WT. Secondary immunodeficiencies, including HIV infection. *J Allergy Clin Immunol* 2010 Feb;125(2 Suppl 2):195-203.

Cunningham-Rundles C. The many faces of common variable immunodeficiency. *Hematology Am Soc Hematol Educ Program* 2012;2012:301-5.

Cunningham-Rundles C. How I treat common variable immune deficiency. *Blood* 2010 Jul 8;116(1):7-15.

Dahly DL, Adair LS, Bollen KA. A structural equation model of the developmental origins of blood pressure. *Int J Epidemiol* 2009 Apr;38(2):538-48.

De Aquino MT, Malhotra A, Mishra MK, Shanker A. Challenges and future perspectives of T cell immunotherapy in cancer. *Immunol Lett* 2015 Aug;166(2):117-33.

De Beudrap P, Rouet F, Fassinou P, Kouakoussui A, Mercier S, Ecochard R, Msellati P. CD4 cell response before and after HAART initiation according to viral load and growth indicators in HIV-1-infected children in Abidjan, Cote d'Ivoire. *J Acquir Immune Defic Syndr* 2008 Sep 1;49(1):70-6.

Deeks SG. HIV infection, inflammation, immunosenescence, and aging. *Annu Rev Med* 2011;62:141-55.

Deeks SG, Kitchen CM, Liu L, Guo H, Gascon R, Narvaez AB, Hunt P, Martin JN, Kahn JO, Levy J, McGrath MS, Hecht FM. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. *Blood* 2004 Aug 15;104(4):942-7.

De Rossi A, Walker AS, Klein N, De Forni D, King D, Gibb DM. Increased thymic output after initiation of antiretroviral therapy in human immunodeficiency virus type 1-infected children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 Trial. *J Infect Dis* 2002 Aug 1;186(3):312-20.

Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press; 2002.

Direction Générale de l'Offre de Soins et Agence Nationale de la Recherche. Programme de Recherche Translationnelle en Santé. [Consulté le 7 août 2015]. URL: <http://www.agence-nationale-recherche.fr/suivi-bilan/editions-2013-et-anterieures/biologie-sante/programme-de-recherche-translationnelle-en-sante-prts/>

Ditlevsen S, Christensen U, Lynch J, Damsgaard MT, Keiding N. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. *Epidemiology* 2005 Jan;16(1):114-20.

Drylewicz J, Matheron S, Lazaro E, Damond F, Bonnet F, Simon F, Dabis F, Brun-Vezinet F, Chêne G, Thiébaud R. Comparison of viro-immunological marker changes between HIV-1 and HIV-2-infected patients in France. *AIDS* 2008 Feb 19;22(4):457-68.

Dumas O, Siroux V, Le Moual N, Varraso R. Causal analysis approaches in epidemiology. *Rev Epidemiol Sante Publique* 2014 Feb;62(1):53-63.

Dunson DB. Bayesian methods for latent trait modelling of longitudinal data. *Stat Methods Med Res* 2007 Oct;16(5):399-415.

Effros RB, Fletcher CV, Gebo K, Halter JB, Hazzard WR, Horne FM, Huebner RE, Janoff EN, Justice AC, Kuritzkes D, Nayfield SG, Plaeger SF, Schmader KE, Ashworth JR, Campanelli C, Clayton CP, Rada B, Woolard NF, High KP. Aging and infectious diseases: workshop on HIV infection and aging: what is known and future research directions. *Clin Infect Dis* 2008 Aug 15;47(4):542-53.

Elmstahl S, Gullberg B. Bias in diet assessment methods--consequences of collinearity and measurement errors on power and observed relative risks. *Int J Epidemiol* 1997 Oct;26(5):1071-9.

Everitt BS. Commentary: classification and cluster analysis. *BMJ* 1995 Aug 26;311(7004):535-6.

Fevang B, Yndestad A, Sandberg WJ, Holm AM, Muller F, Aukrust P, Frøland SS. Low numbers of regulatory T cells in common variable immunodeficiency: association with chronic inflammation in vivo. *Clin Exp Immunol* 2007 Mar;147(3):521-5.

Fiscella K, Bennett NM, Szilagyi PG. Nomenclature in translational research. *JAMA* 2008 May 14;299(18):2148-9.

Focosi D, Bestagno M, Burrone O, Petrini M. CD57+ T lymphocytes and functional immune deficiency. *J Leukoc Biol* 2010 Jan;87(1):107-16.

Fulcher DA, Avery DT, Fewings NL, Berglund LJ, Wong S, Riminton DS, Adelstein S, Tangye SG. Invariant natural killer (iNK) T cell deficiency in patients with common variable immunodeficiency. *Clin Exp Immunol* 2009 Sep;157(3):365-9.

Gathmann B, Mahlaoui N, Gerard L, Oksenhendler E, Warnatz K, Schulze I, Kindle G, Kuijpers TW, Dutch WID, Van Beem RT, Guzman D, Workman S, Soler-Palacin P, De Gracia J, Witte T, Schmidt RE, Litzman J, Hlavackova E, Thon V, Borte M, Borte S, Kumararatne D, Feighery C, Longhurst H, Helbert M, Szaflarska A, Sediva A, Belohradsky BH, Jones A, Baumann U, Meyts I, Kutukculer N, Wagström P, Galal NM, Roesler J, Farmaki E, Zinovieva N, Ciznar P, Papadopoulou-Alataki E, Bienemann K, Velbri S, Panahloo Z, Grimbacher B; European Society for Immunodeficiencies Registry Working Party. Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *J Allergy Clin Immunol* 2014 Jul;134(1):116-26.

Garg V, Lipka S, Rizvon K, Singh J, Rashid S, Mustacchia P. Diffuse nodular lymphoid hyperplasia of intestine in selective IgG 2 subclass deficiency, autoimmune thyroiditis, and autoimmune hemolytic anemia: case report and literature review. *J Gastrointestinal Liver Dis* 2012 Dec;21(4):431-4).

Geneletti S, Gallo V, Porta M, Khoury MJ, Vineis P. Assessing causal relationships in genomics: From Bradford-Hill criteria to complex gene-environment interactions and directed acyclic graphs. *Emerg Themes Epidemiol* 2011 Jun 9;8(1):5.

Gibb DM. Preventing Pediatric HIV and Managing HIV-Infected Children: Where Are We Now, and Where Are We Going? 22th Conference on Retroviruses and Opportunistic Infections, Seattle, United States, 23-26 February 2015.

Gibb DM, Newberry A, Klein N, de Rossi A, Grosch-Woerner I, Babiker A. Immune repopulation after HAART in previously untreated HIV-1-infected children. Paediatric European Network for Treatment of AIDS (PENTA) Steering Committee. *Lancet* 2000 Apr 15;355(9212):1331-2.

Giovannetti A, Pierdominici M, Aiuti F. T-cell homeostasis: the dark(ened) side of common variable immunodeficiency. *Blood* 2008 Jul 15;112(2):446.

Giovannetti A, Pierdominici M, Mazzetta F, Marziali M, Renzi C, Mileo AM, De Felice M, Mora B, Esposito A, Carello R, Pizzuti A, Paggi MG, Paganelli R, Malorni W, Aiuti F. Unravelling the complexity of T cell abnormalities in common variable immunodeficiency. *J Immunol* 2007 Mar 15;178(6):3932-43.

Guiguet M, Boue F, Cadranel J, Lang J-M, Rosenthal E, Costagliola D; Clinical Epidemiology Group of the FHDH-ANRS CO4 cohort. Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. *Lancet Oncol* 2009;10:1152–1159.

Grace JB, Schoolmaster DR, Guntenspergen GR, Little AM, Mitchell BR, Miller KM, Schweiger W. Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* 2012;3(8):73.

Grammatikos AP, Tsokos GC. Immunodeficiency and autoimmunity: lessons from systemic lupus erythematosus. *Trends Mol Med* 2012 Feb;18(2):101-8.

Hainaut M, Ducarme M, Schandene L, Peltier CA, Marissens D, Mascart F, Levy J. Age-related immune reconstitution during highly active antiretroviral therapy in human immunodeficiency virus type 1-infected children. *Pediatr Infect Dis J* 2003;22:62–69.

Hejblum BP, Skinner J, Thiébaud R. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data. *PLoS Comput Biol* 2015 Jun 25;11(6):e1004310.

Hellriegel B. Immunoepidemiology--bridging the gap between immunology and epidemiology. *Trends Parasitol* 2001 Feb;17(2):102-06.

Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965 May;58:295-300.

Horn J, Manguiat A, Berglund LJ, Knerr V, Tahami F, Grimbacher B, Fulcher DA. Decrease in phenotypic regulatory T cells in subsets of patients with common variable immunodeficiency. *Clin Exp Immunol* 2009 Jun;156(3):446-54.

Janes KA, Yaffe MB. Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* 2006 Nov;7(11):820-8.

Holst KK JE. Linear latent variable models: the lava-package. *Comput Stat* 2013;28:1385–452.

Hoyle RH. *Handbook of Structural Equation Modeling*. New York: Guildford Press; 2012.

Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6(1):1-55.

Huenecke S, Behl M, Fadler C, Zimmermann SY, Bochennek K, Esser R, Klarmann D, Kamper M, Sattler A, von Laer D, Klingebiel T, Lehrnbecher T, Koehl U. Age-matched lymphocyte subpopulation reference values in childhood and adolescence: application of exponential regression analysis. *Eur J Haematol* 2008;80:532–539.

Hulstaert F, Hannel I, Deneys V, Munhyeshuli V, Reichert T, Strauss K. Age related changes in human blood lymphocyte subpopulations. *Clin Immunol Immunopathol* 1994;70:152–158.

Hunt PW, Martin JN, Sinclair E, Epling L, Teague J, Jacobson MA, Tracy RP, Corey L, Deeks SG. Valganciclovir reduces T cell activation in HIV-infected individuals with incomplete CD4+ T cell recovery on antiretroviral therapy. *J Infect Dis* 2011 May 15;203(10):1474-83.

Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010;340:2289.

Hsue PY, Deeks SG, Hunt PW. Immunologic basis of cardiovascular disease in HIV-infected adults. *J Infect Dis* 2012 Jun;205 Suppl 3:375-82.

Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nature reviews Immunology* 2014 Jan;14(1):36-49.

Jacqmin-Gadda H, Joly P, Commenges D, Biquet C, Chêne G. Penalized likelihood approach to estimate a smooth mean curve on longitudinal data. *Stat Med* 2002 Aug 30;21(16):2391-402.

Jiang W, Lederman MM, Hunt P, Sieg SF, Haley K, Rodriguez B, Landay A, Martin J, Sinclair E, Asher AI, Deeks SG, Douek DC, Brenchley JM. Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in persons with antiretroviral-treated HIV infection. *J Infect Dis* 2009 Apr 15;199(8):1177-85.

Joint United Nations Programme on HIV/AIDS (UNAIDS). How AIDS changed everything - MDG6: 15 years, 15 lessons of hope from the AIDS response. Geneva: UNAIDS 2015.

Kader M, Smith AP, Guiducci C, Wonderlich ER, Normolle D, Watkins SC, Barrat FJ, Barratt-Boyes SM. Blocking TLR7- and TLR9-mediated IFN- α production by plasmacytoid dendritic cells does not diminish immune activation in early SIV infection. *PLoS Pathog* 2013;9(7):e1003530.

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000 Jan 1;28(1):27-30.

Kekitiinwa A, Cook A, Nathoo K, Mugenyi P, Nahirya-Ntege P, Bakeera-Kitaka S, Thomason M, Bwakura-Dangarembizi M, Musiime V, Munderi P, Naidoo-James B, Vhembo T, Tumusiime C, Katuramu R, Crawley J, Prendergast AJ, Musoke P, Walker AS, Gibb DM. Routine versus clinically driven laboratory monitoring and first-line antiretroviral therapy strategies in African children with HIV (ARROW): a 5-year open-label randomised factorial trial. *Lancet* 2013 Apr 20;381(9875):1391-403.

Kelesidis T, Yang O. Good's syndrome remains a mystery after 55 years: A systematic review of the scientific evidence. *Clin Immunol* 2010 Jun;135(3):347-63.

Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. *Am J Epidemiol* 2010 Sep 1;172(5):517-24.

Kinloch AJ, Chang A, Ko K, Henry Dunand CJ, Henderson S, Maienschein-Cline M, Kaverina N, Rovin BH, Salgado Ferrer M, Wolfgeher D, Liarski V, Haddon DJ, Utz PJ, Wilson PC, Clark MR. Vimentin is a dominant target of in situ humoral immunity in human lupus tubulointerstitial nephritis. *Arthritis Rheumatol* 2014 Dec;66(12):3359-70.

Klatt NR, Funderburg NT, Brenchley JM. Microbial translocation, immune activation, and HIV disease. *Trends Microbiol* 2013 Jan;21(1):6-13.

Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982 Dec;38(4):963-74.

Lebart L, Morineau A, Piron M. Statistique exploratoire multidimensionnelle. 3e ed. Paris: Dunod; 2000.

Le Moing V, Thiébaud R, Chêne G, Leport C, Cailleton V, Michelet C, Fleury H, Herson S, Raffi F; APROCO Study Group. Predictors of long-term increase in CD4(+) cell counts in human immunodeficiency virus-infected patients receiving a protease inhibitor-containing antiretroviral regimen. *J Infect Dis* 2002 Feb 15;185(4):471-80.

Lewis J, Walker AS, Castro H, De Rossi A, Gibb DM, Giaquinto C, Klein N, Callard R. Age and CD4 count at initiation of antiretroviral therapy in HIV-infected children: effects on long-term T-cell reconstitution. *J Infect Dis* 2012 Feb 15;205(4):548-56.

Lichtenstein KA, Armon C, Buchacz K, Chmiel JS, Buckner K, Wood K, Holmberg SD, Brooks JT; HIV Outpatient Study (HOPS) Investigators. Low CD4+ T cell count is a risk factor for cardiovascular disease events in the HIV Outpatient Study. *Clin Infect Dis* 2010;51:435–447.

Litzman J, Nechvatalova J, Xu J, Ticha O, Vlkova M, Hel Z. Chronic immune activation in common variable immunodeficiency (CVID) is associated with elevated serum levels of soluble CD14 and CD25 but not endotoxaemia. *Clin Exp Immunol* 2012 Dec;170(3):321-32.

Liu K. Measurement error and its impact on partial correlation and multiple linear regression analyses. *Am J Epidemiol* 1988 Apr;127(4):864-74.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 2001 Dec;25(4):402-8.

Lugli E, Pinti M, Nasi M, Troiano L, Ferraresi R, Mussi C, Salvioli G, Patsekina V, Robinson JP, Durante C, Cocchi M, Cossarizza A. Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data. *Cytometry A* 2007 May;71(5):334-44.

Lugli E, Roederer M, Cossarizza A. Data analysis in flow cytometry: the future just started. *Cytometry A* 2010 Jul;77(7):705-13.

Malphettes M, Gerard L, Galicier L, Boutboul D, Asli B, Szalat R, Perlat A, Masseur A, Schleinitz N, Le Guenno G, Viallard JF, Bonnotte B, Thiercelin-Legrand MF, Sanhes L, Borie R, Georgin-Lavialle S, Fieschi C, Oksenhendler E; DEFicit Immunitaire de l'adulte Study Group. Good Syndrome: An Adult-Onset Immunodeficiency Remarkable for Its High

Incidence of Invasive Infections and Autoimmune Complications. *Clin Infect Dis* 2015 Jul 15;61(2):13-9.

Mansiaux Y, Salez N, Lapidus N, Setbon M, Andreoletti L, Leruez-Ville M, Cauchemez S, Gougeon ML, Vély F, Schwarzing M, Abel L, Delabre RM, Flahault A, de Lamballerie X, Carrat F. Causal analysis of H1N1pdm09 influenza infection risk in a household cohort. *J Epidemiol Community Health* 2015 Mar;69(3):272-7.

Marashi SM, Raeiszadeh M, Enright V, Tahami F, Workman S, Chee R, Webster AD, Milne RS, Emery VC. Influence of cytomegalovirus infection on immune cell phenotypes in patients with common variable immunodeficiency. *J Allergy Clin Immunol* 2012 May;129(5):1349-56 e3.

Martinson JA, Montoya CJ, Usuga X, Ronquillo R, Landay AL, Desai SN. Chloroquine modulates HIV-1-induced plasmacytoid dendritic cell alpha interferon: implication for T-cell activation. *Antimicrob Agents Chemother* 2010 Feb;54(2):871-81.

Mason CE, Porter SG, Smith TM. Characterizing multi-omic data in systems biology. *Adv Exp Med Biol* 2014;799:15-38.

Meier A, Alter G, Frahm N, Sidhu H, Li B, Bagchi A, Teigen N, Streeck H, Stellbrink HJ, Hellman J, van Lunzen J, Altfeld M. MyD88-dependent immune activation mediated by human immunodeficiency virus type 1-encoded Toll-like receptor ligands. *J Virol* 2007 Aug;81(15):8180-91.

Ministère des Affaires sociales, de la Santé et des Droits des femmes. Les maladies rares. [Consulté le 7 août 2015]. URL: <http://www.sante.gouv.fr/les-maladies-rares.html>

Mouillot G, Carmagnat M, Gerard L, Garnier JL, Fieschi C, Vince N, Karlin L, Viallard JF, Jaussaud R, Boileau J, Donadieu J, Gardembas M, Schleinitz N, Suarez F, Hachulla E, Delavigne K, Morisset M, Jacquot S, Just N, Galicier L, Charron D, Debré P, Oksenhendler E, Rabian C; DEFI Study Group. B-Cell and T-Cell Phenotypes in COVID Patients Correlate with the Clinical Phenotype of the Disease. *J Clin Immunol* 2010 Sep;30(5):746-55.

Mugenyi P, Walker AS, Hakim J, Munderi P, Gibb DM, Kityo C; DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* 2010 Jan 9;375(9709):123-31.

Ness RB. "Big" science and the little guy. *Epidemiology* 2007 Jan;18(1):9-12.

O’Gorman MR, Zollett J, Bensen N. Flow cytometry assays in primary immunodeficiency diseases. *Methods Mol Biol* 2011;699:317-35.

Oksenhendler E, Gerard L, Fieschi C, Malphettes M, Mouillot G, Jaussaud R, Viallard JF, Gardembas M, Galicier L, Schleinitz N, Suarez F, Soulas-Sprauel P, Hachulla E, Jaccard A, Gardeur A, Théodorou I, Rabian C, Debré P; DEFI study group. Infections in 252 patients with common variable immunodeficiency. *Clin Infect Dis* 2008 May 15;46(10):1547-54.

Pakker NG, Notermans DW, de Boer RJ, Roos MT, de Wolf F, Hill A, Leonard JM, Danner SA, Miedema F, Schellekens PT. Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nat Med* 1998 Feb;4(2):208-14.

Palmer LJ. The New Epidemiology: putting the pieces together in complex disease aetiology. *Int J Epidemiol* 2004 Oct;33(5):925-8.

Park J, Munagala I, Xu H, Blankenship D, Maffucci P, Chaussabel D, Banchereau J, Pascual V, Cunningham-Rundles C. Interferon signature in the blood in inflammatory common variable immune deficiency. *PLoS One* 2013;8(9):e74893.

Park MA, Li JT, Hagan JB, Maddox DE, Abraham RS. Common variable immunodeficiency: a new look at an old disease. *Lancet* 2008 Aug 9;372(9637):489-502.

Pascual V, Banchereau J. Tracking interferon in autoimmunity. *Immunity* 2012 Jan 27;36(1):7-9.

Peakman M. Broadening the translational immunology landscape. *Clin Exp Immunol* 2012 Dec;170(3):249-53.

Pearl J. Causal inference in the health sciences: a conceptual introduction. *Health Serv Outcomes Res Methodol* 2001;2:189–220.

Pencina MJ, D’Agostino RB Sr, D’Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008 Jan 30;27(2):157-72.

Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol* 2002 Jan;2(1):28-36.

Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996 Mar 15;271(5255):1582-6.

Piqueras B, Lavenu-Bombled C, Galicier L, Bergeron-van der Cruyssen F, Mouthon L, Chevret S, Debré P, Schmitt C, Oksenhendler E. Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects. *J Clin Immunol* 2003 Sep;23(5):385-400.

Plaeger SF, Collins BS, Musib R, Deeks SG, Read S, Embry A. Immune activation in the pathogenesis of treated chronic HIV disease: a workshop summary. *AIDS Res Hum Retroviruses* 2012 May;28(5):469-77.

Preud'homme JL, Hanson LA. IgG subclass deficiency. *Immunodeficiency Rev* 1990;2(2):129-49.

Rajasuriar R, Khoury G, Kamarulzaman A, French MA, Cameron PU, Lewin SR. Persistent immune activation in chronic HIV infection: do any interventions work? *AIDS* 2013 May 15;27(8):1199-208.

Rawson PM, Molette C, Videtta M, Altieri L, Franceschini D, Donato T, Finocchi L, Propato A, Paroli M, Meloni F, Mastroianni CM, d'Ettorre G, Sidney J, Sette A, Barnaba V. Cross-presentation of caspase-cleaved apoptotic self antigens in HIV infection. *Nat Med* 2007 Dec;13(12):1431-9.

Regent A, Bussone G, Kaveri SV, Mouthon L. Humoral and cellular autoimmunity: from physiology to pathology. *Rev Med Interne* 2009 Dec;30(12 Suppl):H1-8.

Reinhardt JD, Fellinghauer BA, Strobl R, Stucki G. Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis. *Disabil Rehabil* 2010;32(12):1000-10.

Resino S, Correa R, Bellon JM, Sanchez-Ramon S, Munoz-Fernandez MA. Characterizing immune reconstitution after long-term highly active antiretroviral therapy in pediatric AIDS. *AIDS Res Hum Retroviruses* 2002 ;18:1395–1406.

Resino S, Seoane E, Perez A, Ruiz-Mateos E, Leal M, Munoz-Fernandez MA. Different profiles of immune reconstitution in children and adults with HIV-infection after highly active antiretroviral therapy. *BMC Infect Dis* 2006;6:112.

Rezaei N, Wing JB, Aghamohammadi A, Carling J, Lees A, Asgarian-Omran H, Pourpak Z, Sarrafnejad A, Kardar GA, Shahrestani T, Masoumi F, Zare A, Saghaifi S, Sarrafzadeh S,

Foster RA, Heath AW, Read RC. B-cell-T-cell activation and interaction in common variable immunodeficiency. *Hum Immunol* 2010 Apr;71(4):355-62.

Ries M, Pritschet K, Schmidt B. Blocking type I interferon production: a new therapeutic option to reduce the HIV-1-induced immune activation. *Clin Dev Immunol* 2012;2012:534929.

Ringner M. What is principal component analysis? *Nat Biotechnol* 2008 Mar;26(3):303-4.

Rocchetti M, Germani M, Del Bene F, Poggesi I, Magni P, Pesenti E, De Nicolao G. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth after administration of an anti-angiogenic agent, bevacizumab, as single-agent and combination therapy in tumor xenografts. *Cancer Chemother Pharmacol* 2013 May;71(5):1147-57.

Rouet F, Fassinou P, Inwoley A, Anaky M-F, Kouakoussui A, Blanche S, Msellati P; ANRS 1244/1278 Programme Enfants Yopougon. Long term survival and immuno-virological response of African HIV-1-infected children to highly active antiretroviral therapy regimens. *AIDS* 2006;20: 2315–2319.

Rutter M. Epidemiological methods to tackle causal questions. *Int J Epidemiol* 2009 Feb;38(1):3-6.

Sabin CA, Smith CJ, d'Arminio Monforte A, Battegay M, Gabiano C, Galli L, Geelen S, Gibb D, Guiguet M, Judd A, Leport C, Dabis F, Pantazis N, Porter K, Raffi F, Thorne C, Torti C, Walker S, Warszawski J, Wintergerst U, Chene G, Lundgren J; Collaboration of Observational HIV Epidemiological Research Europe (COHERE) Study Group. Response to combination antiretroviral therapy: variation by age. *AIDS* 2008 Jul 31;22(12):1463-73.

Sanchez BN, Budtz-Jørgensen E, Ryan LM, and Hu H. Structural Equation Models: A Review With Applications to Environmental Epidemiology. *JASA* 2005;100(472):1443-5.

Sauvageot D, Schaefer M, Olson D, Pujades-Rodriguez M, O'Brien DP. Antiretroviral therapy outcomes in resource-limited settings for HIV-infected children <5 years of age. *Pediatrics* 2010 May;125(5):e1039-47.

Schneider WM, Chevillotte MD, Rice CM. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* 2014;32:513-45.

Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol* 2011 Dec;1(6):519-25.

Scott-Taylor TH, Green MR, Raeiszadeh M, Workman S, Webster AD. Defective maturation of dendritic cells in common variable immunodeficiency. *Clin Exp Immunol* 2006 Sep;145(3):420-7.

Simeoni M, Magni P, Cammia C, De Nicolao G, Croci V, Pesenti E, Germani M, Poggese I, Rocchetti M. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res* 2004 Feb 1;64(3):1094-101.

Sinha A, Hripesak G, Markatou M. Large datasets in biomedicine: a discussion of salient analytic issues. *J Am Med Inform Assoc* 2009 Nov-Dec;16(6):759-67.

Sleasman JW, Nelson RP, Goodenow MM, Wilfret D, Hutson A, Zuckerman J, Pizzo PA, Mueller BU. Immunoreconstitution after ritonavir therapy in children with human immunodeficiency virus infection involves multiple lymphocyte lineages. *J Pediatr* 1999;134:597-606.

Sodora DL, Silvestri G. Immune activation and AIDS pathogenesis. *AIDS* 2008 Feb 19;22(4):439-46.

Song JJ, Ren Y, Yan F. Classification for high-throughput data with an optimal subset of principal components. *Comput Biol Chem* 2009 Oct;33(5):408-13.

Sozzani S, Bosisio D, Scarsi M, Tincani A. Type I interferons in systemic autoimmunity. *Autoimmunity* 2010 Apr;43(3):196-203.

Steinmann GG, Klaus B, Muller-Hermelink HK. The involution of the ageing human thymic epithelium is independent of puberty. A morphometric study. *Scand J Immunol* 1985 Nov;22(5):563-75.

Stone SF, Price P, Khan N, Moss PA, French MA. HIV patients on antiretroviral therapy have high frequencies of CD8 T cells specific for Immediate Early protein-1 of cytomegalovirus. *AIDS* 2005 Mar 24;19(6):555-62.

Stratton R, Slapak G, Mahungu T, Kinloch-de Loes S. Autoimmunity and HIV. *Curr Opin Infect Dis* 2009 Feb;22(1):49-56.

Sutcliffe CG, van Dijk JH, Bolton C, Persaud D, Moss WJ. Effectiveness of antiretroviral therapy among HIV-infected children in sub-Saharan Africa. *Lancet Infect Dis* 2008 Aug;8(8):477-89.

Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM. Virus-Specific CD4(+) Memory-Phenotype T Cells Are Abundant in Unexposed Adults. *Immunity* 2013;21;38(2):373-83.

Tebas P, Henry WK, Matining R, Weng-Cherng D, Schmitz J, Valdez H, Jahed N, Myers L, Powderly WG, Katzenstein D. Metabolic and immune activation effects of treatment interruption in chronic HIV-1 infection: implications for cardiovascular risk. *PLoS One* 2008;3(4):2021.

Teijaro JR, Ng C, Lee AM, Sullivan BM, Sheehan KC, Welch M, Schreiber RD, de la Torre JC, Oldstone MB. Persistent LCMV infection is controlled by blockade of type I interferon signaling. *Science* 2013 Apr 12;340(6129):207-11.

Thiébaud R, Alioum A. Modèles de régression en épidémiologie. In: Dabis F, Desenclos JF, editors. *Epidémiologie de terrain : méthodes et application*. Montrouge: John Libbey Eurotext; 2012 p. 375-89.

Thiébaud R, Hejblum B, Richert L. The analysis of "Big Data" in clinical research. *Rev Epidemiol Sante Publique* 2014 Feb;62(1):1-4.

Thiébaud R, Jacqmin-Gadda H, Chêne G, Leport C, Commenges D. Bivariate linear mixed models using SAS proc MIXED. *Comput Methods Programs Biomed* 2002 Nov;69(3):249-56.

Thiébaud R, Jacqmin-Gadda H, Leport C, Katlama C, Costagliola D, Le Moing V, Morlat P, Chêne G; APROCO Study Group. Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left censoring of HIV RNA measures. *J Biopharm Stat* 2003;13: 271–282.

Thiébaud R, Jacqmin-Gadda H, Walker S, Sabin C, Prins M, Del Amo J, Porter K, Dabis F, Chêne G; CASCADE Collaboration. Determinants of response to first HAART regimen in antiretroviral-naïve patients with an estimated time since HIV seroconversion. *HIV Med* 2006 Jan;7(1):1-9.

Thiébaud R, Morlat P, Jacqmin-Gadda H, Neau D, Mercie P, Dabis F, Chêne G. Clinical progression of HIV-1 infection according to the viral response during the first year of antiretroviral treatment. Groupe d'Epidémiologie du SIDA en Aquitaine (GECSA). *AIDS* 2000 May 26;14(8):971-8.

Thiébaud R, Pellegrin I, Chêne G, Viallard JF, Fleury H, Moreau JF, Pellegrin JL, Blanco P. Immunological markers after long-term treatment interruption in chronically HIV-1 infected patients with CD4 cell count above 400 x 10⁶ cells/l. *AIDS* 2005 Jan 3;19(1):53-61.

Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Stat Methodol* 1996;58:267-88.

Tornøe CW, Agerø H, Jonsson EN, Madsen H, Nielsen HA. Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations. *Comput Methods Programs Biomed* 2004 Oct;76(1):31-40.

Triant VA, Regan S, Lee H, Sax PE, Meigs JB, Grinspoon SK. Association of immunologic and virologic factors with myocardial infarction rates in a U.S. health care system. *J Acquir Immune Defic Syndr* 2010;55: 615–619.

Trinchieri G. Cancer Immunity: Lessons From Infectious Diseases. *J Infect Dis* 2015 Jul 15;212 Suppl 1:67-73.

Tu YK. Commentary: Is structural equation modelling a step forward for epidemiologists? *Int J Epidemiol* 2009 Apr;38(2):549-51.

Tu YK, West R, Ellison GT, Gilthorpe MS. Why evidence for the fetal origins of adult disease might be a statistical artifact: the "reversal paradox" for the relation between birth weight and blood pressure in later life. *Am J Epidemiol* 2005 Jan 1;161(1):27-32.

Van Helden P. Data-driven hypotheses. *EMBO Rep* 2013 Feb;14(2):104.

VanderWeele TJ. Mediation and mechanism. *Eur J Epidemiol* 2009;24(5):217-24.

VanderWeele TJ. Invited commentary: structural equation models and epidemiologic analysis. *Am J Epidemiol* 2012 Oct 1;176(7):608-12.

Viallard JF, Blanco P, Andre M, Etienne G, Liferman F, Neau D, Vidal E, Moreau JF, Pellegrin JL. CD8+HLA-DR+ T lymphocytes are increased in common variable immunodeficiency patients with impaired memory B-cell differentiation. *Clin Immunol* 2006 Apr;119(1):51-8.

Viallard JF, Bloch-Michel C, Caubet O, Parrens M, Texier-Maugein J, Neau-Cransac M, Taupin JL, Moreau JF, Pellegrin JL. Gammadelta T lymphocytosis associated with granulomatous disease in a patient with common variable immunodeficiency. *Clin Infect Dis* 2002;35(12):134-7.

Vlková M, Thon V, Sárffyová M, Bláha L, Svobodník A, Lokaj J, Litzman J. Age dependency and mutual relations in T and B lymphocyte abnormalities in common variable immunodeficiency patients. *Clin Exp Immunol* 2006 Feb;143(2):373-9.

Wade AM, Ades AE. Age related reference ranges: significance tests for models and confidence intervals for centiles. *Stat Med* 1994;13:2359–2367.

Wahren-Herlenius M, Dorner T. Immunopathogenic mechanisms of systemic autoimmune disease. *Lancet* 2013 Aug 31;382(9894):819-31.

Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. *Eur J Epidemiol* 2009;24(12):733-6.

Walker AS, Doerholt K, Sharland M, Gibb DM, for the Collaborative HIV Paediatric Study (CHIPS) Steering Committee. Response to highly active antiretroviral therapy varies with age: the UK and Ireland Collaborative HIV Paediatric Study. *AIDS* 2004;18:1915–1924.

Ward J. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;58(301):236-44.

Warnatz K, Denz A, Drager R, Braun M, Groth C, Wolff-Vorbeck G, Eibel H, Schlesier M, Peter HH. Severe deficiency of switched memory B cells (CD27(+)IgM(-)IgD(-)) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* 2002 Mar 1;99(5):1544-51.

Wehr C, Kivioja T, Schmitt C, Ferry B, Witte T, Eren E, Vlkova M, Hernandez M, Detkova D, Bos PR, Poerksen G, von Bernuth H, Baumann U, Goldacker S, Gutenberger S, Schlesier M, Bergeron-van der Cruyssen F, Le Garff M, Debré P, Jacobs R, Jones J, Bateman E, Litzman J, van Hagen PM, Plebani A, Schmidt RE, Thon V, Quinti I, Espanol T, Webster AD, Chapel H, Vihinen M, Oksenhendler E, Peter HH, Warnatz K. The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* 2008 Jan 1;111(1):77-85.

Wilson EB, Yamada DH, Elsaesser H, Herskovitz J, Deng J, Cheng G, Aronow BJ, Karp CL, Brooks DG. Blockade of chronic type I interferon signaling to control persistent LCMV infection. *Science* 2013 Apr 12;340(6129):202-7.

Wittkop L, Bitard J, Lazaro E, Neau D, Bonnet F, Mercie P, Dupon M, Hessamfar M, Ventura M, Malvy D, Dabis F, Pellegrin JL, Moreau JF, Thiébaud R, Pellegrin I; Groupe d'Epidémiologie Clinique du SIDA en Aquitaine. Effect of cytomegalovirus-induced immune response, self antigen-induced immune response, and microbial translocation on chronic

immune activation in successfully treated HIV type 1-infected patients: the ANRS CO3 Aquitaine Cohort. *J Infect Dis* 2013 Feb 15;207(4):622-7.

Wittkop L, Commenges D, Pellegrin I, Breilh D, Neau D, Lacoste D, Pellegrin JL, Chêne G, Dabis F, Thiébaud R. Alternative methods to analyse the impact of HIV mutations on virological response to antiviral therapy. *BMC Med Res Methodol* 2008 Oct 22;8:68.

Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008 Jan 9;299(2):211-3.

World Health Organization. Antiretroviral therapy of HIV infection in infants and children: recommendations for a public health approach-2010 revision. Geneva: World Health Organization 2010.

Xiao X, Miao Q, Chang C, Gershwin ME, Ma X. Common variable immunodeficiency and autoimmunity--an inconvenient truth. *Autoimmun Rev* 2014 Aug;13(8):858-64.

Yang S, Netterwald J, Wang W, Zhu H. Characterization of the elements and proteins responsible for interferon-stimulated gene induction by human cytomegalovirus. *J Virol* 2005 Apr;79(8):5027-34.

Yong PF, Thaventhiran JE, Grimbacher B. "A rose is a rose is a rose," but CVID is Not CVID common variable immune deficiency (CVID), what do we know in 2011? *Adv Immunol* 2011;111:47-107.

Yoo W, Mayberry R, Bae S, Singh K, Peter He Q, Lillard JW Jr. A Study of Effects of MultiCollinearity in the Multivariable Analysis. *Int J Appl Sci Technol* 2014 Oct;4(5):9-19.

Zhou Y, Zhang Y, Moorman JP, Yao ZQ, Jia ZS. Viral (hepatitis C virus, hepatitis B virus, HIV) persistence and immune homeostasis. *Immunology* 2014 Nov;143(3):319-30.

ANNEXES

Annexe 1. Projets pris en charge dans le champ de la médecine interne, des maladies infectieuses et de l'immunologie.

Programmes Hospitaliers de Recherche Clinique (PHRC) nationaux pris en charge en tant que méthodologiste à l'Unité de Soutien Méthodologique à la Recherche.

Investigateur	Titre du projet
Pr Yves PEREL	Cohorte prospective nationale de suivi des enfants atteints de cytopénies auto-immunes sévères
Pr Yves LEVY	Essai randomisé multicentrique de phase II d'évaluation immunologique d'une stratégie vaccinale de type « prime-boost » associant une administration du vaccin conjugué anti-pneumococcique à S0 suivie de l'injection du vaccin polysaccharidique à S4 comparé à l'administration du vaccin polysaccharidique seul à S4 chez des patients ayant un déficit immunitaire commun variable
Pr Odile LAUNAY	Essai pilote multicentrique évaluant l'immunogénicité d'une stratégie vaccinale anti-pneumococcique innovante chez des adultes splénectomisés
Dr Nathalie ALADJIDI	Etude nationale adossée à une cohorte maladie rare de la balance bénéfice/risque des traitements immunomodulateurs prescrits chez l'enfant et l'adolescent pour une cytopénie auto-immune
Dr Jérôme HARAMBAT	Essai multicentrique évaluant la réponse immunologique de la vaccination contre l'infection par les papillomavirus humains chez des jeunes filles recevant un traitement immunosuppresseur

Projets préparés en tant que méthodologiste dans le cadre des appels d'offres 2015 de la Direction Générale de l'Offre de Soins

PHRC national Cancer (Onco-immunologie)

Lettre d'intention acceptée

PHRC national (Maladies infectieuses)

Lettre d'intention acceptée

PHRC inter-régional (Médecine interne)

Lettre d'intention acceptée

Soutien méthodologique pour étude de service de Médecine Interne et Maladies Infectieuses

Activation immunitaire chronique chez les patients atteints d'anémies hémolytiques auto-immunes (Pr Estibaliz LAZARO, Dr Emmanuel RIBEIRO)

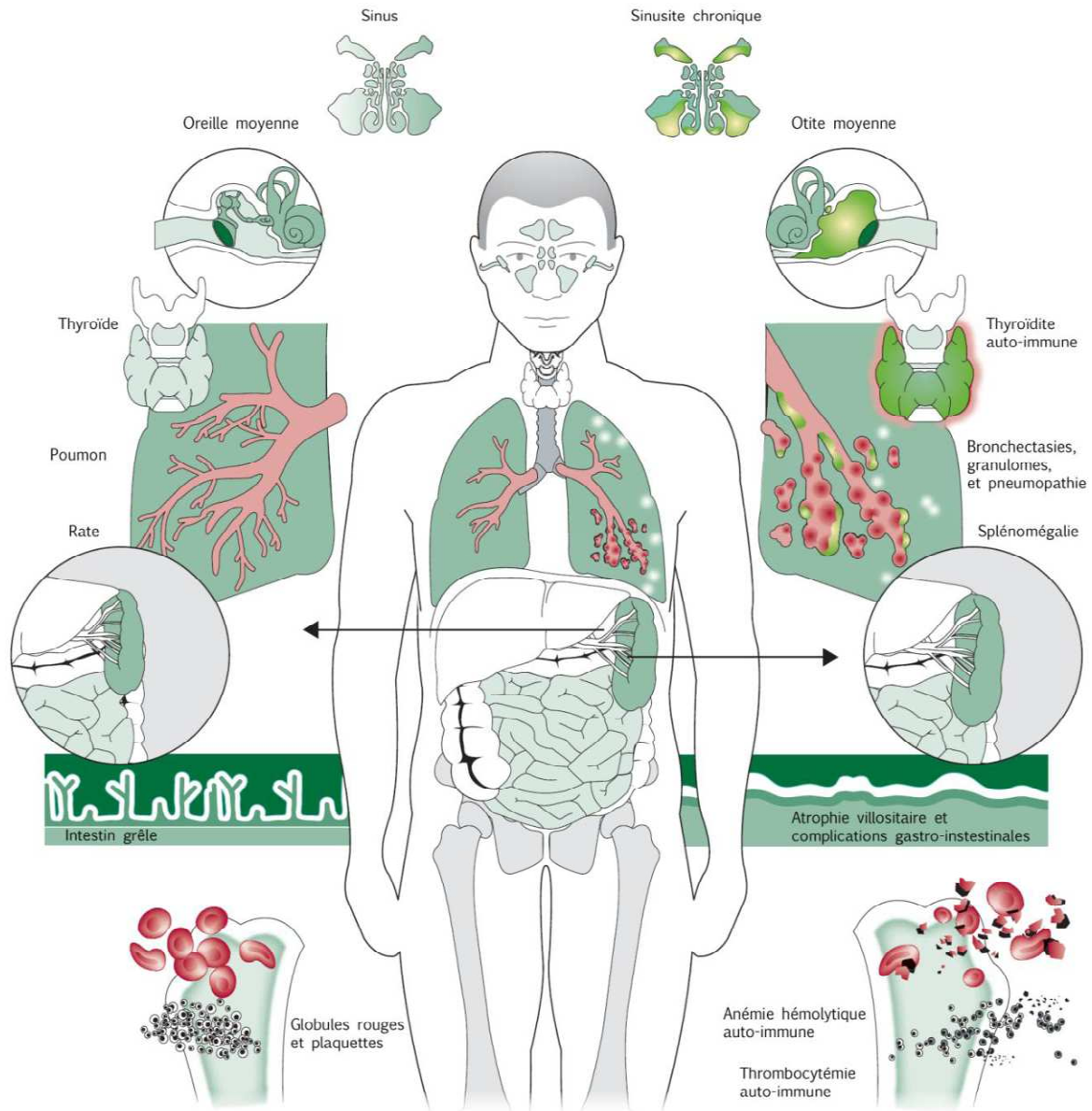
Soutien méthodologique pour thèses d'exercice de médecine

Etude rétrospective au CHU de Bordeaux, à propos de 43 cas atteints de syndrome de Goujerot-Sjögren. Directeur de thèse : Pr Patrick MERCIE. Présentée et soutenue publiquement le 24 avril 2014 par Chandra ADJODAH. Membre du Jury.

Identification de facteurs prédictifs d'une réponse au rituximab en traitement de deuxième ligne du purpura thrombopénique immunologique. Directeur de thèse : Pr Jean-François VIALARD. Présentée et soutenue publiquement le 20 décembre 2013 par Etienne RIVIERE.

Elévation du ratio CD4+/CD8+, étude rétrospective à propos de 36 cas. Directeurs de thèse : Pr Jean-François MOREAU / Pr Jean-Luc TAUPIN. Présentée et soutenue publiquement le 18 septembre 2015 par Alexandre GUY.

Annexe 2. Représentation des atteintes organiques retrouvées dans le déficit immunitaire commun variable de l'adulte.



D'après Park MA et al. Lancet 2008;372(9637):489-502.

A gauche, figurent les organes sains. A droite, figurent les atteintes organiques dont le déficit immunitaire commun variable est responsable.

Annexe 3. Analyse en composantes principales : valeurs propres, variances expliquées par les composantes principales et contribution factorielle des marqueurs immunologiques. Etude ALTADIIH.

Dans l'analyse transversale des données d'inclusion de la cohorte ALTADIIH, l'analyse en composantes principales présentait l'avantage de pouvoir analyser la corrélation des 23 marqueurs immunologiques d'intérêt. Elle permettait d'identifier la structure immunologique de notre population d'étude et d'obtenir de nouvelles variables non corrélées (composantes principales) résumant les données.

Les valeurs propres de l'analyse en composantes principales et la part de variance des données, expliquée par chaque composante principale, sont présentées dans le tableau 1 et la figure 1 ci-après.

Tableau 1 (Annexe 3). Valeurs propres et variances expliquées par les composantes principales. Etude ALTADIIH (n = 79/81 patients).

Composantes principales (CP)	Valeur propre	Variance	Variance cumulée
CP1	6,087	26,566	26,566
CP2	3,778	16,327	42,894
CP3	1,970	8,566	51,460
CP4	1,638	7,120	58,580
CP5	1,596	6,938	65,518
CP6	1,135	4,934	70,453
CP7	1,013	4,405	74,858
CP8	0,883	3,841	78,699
CP9	0,798	3,472	82,171
CP10	0,731	3,178	85,348
CP11	0,653	2,838	88,186
CP12	0,590	2,563	90,749
CP13	0,445	1,933	92,683
CP14	0,370	1,607	94,289
CP15	0,292	1,268	95,557
CP16	0,269	1,172	96,728
CP17	0,242	1,054	97,782
CP18	0,192	0,835	98,616
CP19	0,143	0,621	99,237
CP20	0,093	0,404	99,642
CP21	0,053	0,231	99,872
CP22	0,020	0,085	99,958
CP23	0,010	0,042	100,000

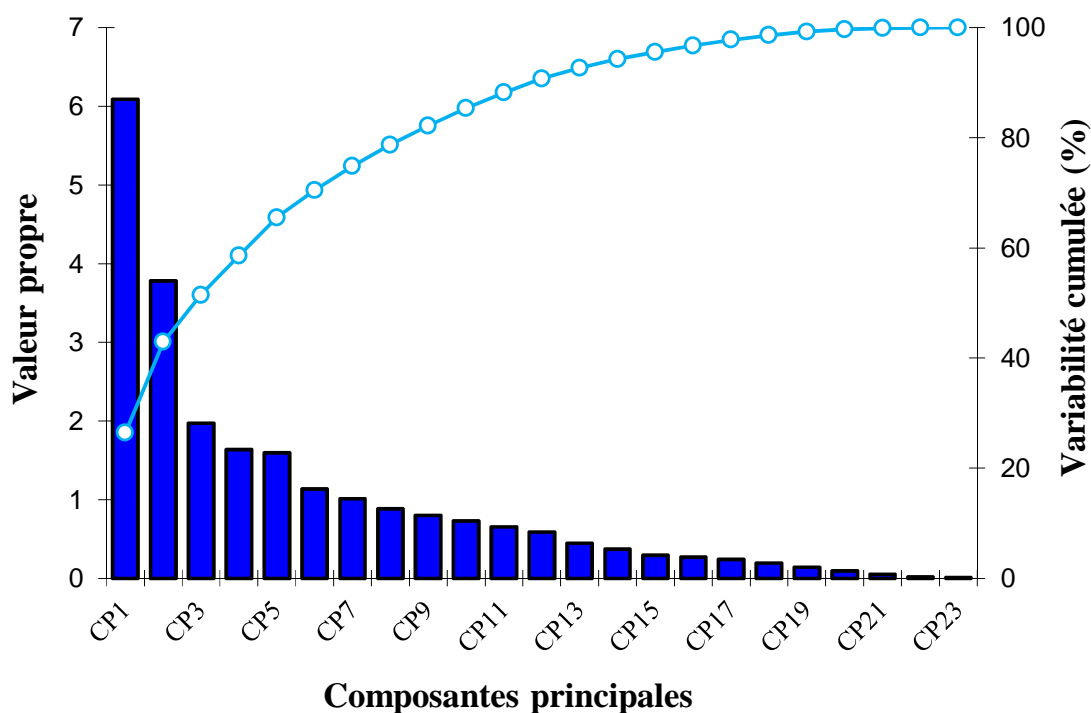


Figure 1 (Annexe 3). Valeurs propres et variabilités expliquées par les composantes principales. Etude ALTADIIH (n = 79/81 patients).

On retrouve ainsi que la première composante principale expliquait 27 % de la variabilité des données et que la deuxième composante principale expliquait 16 % de la variabilité des données.

Les coordonnées des marqueurs immunologiques (ou points variables) sur les deux premières composantes principales sont présentées en tableau 2.

Tableau 2 (Annexe 3). Contribution factorielle des marqueurs immunologiques. Etude ALTADIH (n = 79/81 patients).

	CP1 (27 %)	CP2 (16 %)
CD3+	0,746	0,559
CD4+	0,805	0,311
CD8+	0,518	0,725
RA+CCR7-	-0,127	0,585
RA+CCR7+	0,417	-0,452
RA-CCR7-	-0,259	0,007
RA-CCR7+	0,244	-0,356
CD4+/CD8+ ratio	0,258	-0,478
CD3+HLA-DR+	-0,554	0,639
CD4+HLA-DR+	-0,612	0,425
CD8+HLA-DR+	-0,455	0,640
CD8+CD57+	-0,328	0,581
CD19+	0,598	0,187
CD27+IgD-	0,632	0,120
CD27-IgD-	0,202	0,088
CD27+IgD+	0,544	0,032
CD27-IgD+	-0,596	-0,015
DCp	0,651	-0,025
DCm	0,518	0,016
NK	0,462	0,308
Reg T	0,764	0,106
GD	0,378	0,569
GD2	0,392	-0,344

Ces coordonnées permettaient d'évaluer la contribution factorielle de chaque marqueur immunologique à la composante principale et permettaient de représenter graphiquement, par projection, les marqueurs immunologiques impliqués dans les déficits immunitaires humoraux primitifs, sur un plan à deux dimensions, constitué par les axes des deux premières composantes principales. La représentation graphique des marqueurs immunologiques dans le plan formé par CP1 x CP2 a donné lieu à la figure 1 de l'article présenté en section 3.4.1 et a permis d'interpréter les deux premières composantes principales. Pour valider cette interprétation, il a également été nécessaire d'explorer la projection des marqueurs immunologiques sur les autres composantes principales (contributions non présentées ici).

Annexe 4. Principes de l'index de reclassification.

Nous avons utilisé la méthode de l'index de reclassification proposée par Pencina MJ et al, pour évaluer l'intérêt du marqueur CD8+HLA-DR+ pour la complétude de la classification EUROclass chez les patients atteints de déficit immunitaire commun variable.

Cette méthode était une alternative à l'utilisation de l'aire sous la courbe pour évaluer la qualité d'un modèle pour prédire la survenue de complications. Elle permettait de calculer le gain apporté par l'utilisation du nouveau marqueur par rapport au modèle de référence. L'intérêt de cette méthode résidait dans le fait qu'elle amenait des informations sur la pertinence clinique de l'utilisation du nouveau marqueur proposé.

Dans notre cas, la classification EUROclass était considérée comme la référence pour prédire la survenue de complications du déficit immunitaire commun variable (DICV). Suite à des discussions étroites avec les cliniciens, nous avons considéré un seuil exploratoire de probabilité de survenue de complications de 50 %.

Le modèle confronté à la référence était celui qui comprenait l'introduction du nouveau marqueur. Pour chaque patient, le modèle de régression logistique était capable de calculer la probabilité de présenter une complication du déficit immunitaire commun variable (variable à expliquer), via les coefficients de régression estimés β et les variables explicatives incluses dans le modèle. Cela a permis de compléter le tableau de contingence présenté ci-après (tableau 1, annexe 4).

Tableau 1 (Annexe 4). Classement des patients atteints ou non de complications du DICV selon l'apport du nouveau marqueur.

Modèle EUROclass	Modèle EUROclass+nouveau marqueur		Total
	< 50 %	≥ 50 %	
Probabilités prédites (%)			
Complications du DICV			
< 50 %	a	b	a + b
≥ 50 %	c	d	c + d
Total	a + c	b + d	a + b + c + d
Pas de complications du DICV			
< 50 %	e	f	e + f
≥ 50 %	g	h	g + h
Total	e + g	f + h	e + f + g + h

En gris, figurent les concordances des deux modèles pour classer les patients.

L'index de reclassification est basé sur les discordances de reclassement entre les deux modèles. Il est calculé de la manière suivante :

$$\left[\left(\frac{b}{a+b+c+d} \right) - \left(\frac{c}{a+b+c+d} \right) \right] - \left[\left(\frac{f}{e+f+g+h} \right) - \left(\frac{g}{e+f+g+h} \right) \right]$$

Si l'introduction du nouveau marqueur dans le modèle permet de mieux prédire la survenue de complications, l'index de reclassification est positif. Au-delà de la valeur même de cet index, l'intérêt de cette méthode réside dans l'intérêt clinique de pouvoir identifier le nombre de patients mieux reclassés ou moins bien reclassés, en prenant en compte le nouveau marqueur, et de pouvoir étudier plus spécifiquement leurs profils cliniques.

Les résultats de l'application de cette méthode dans l'étude ALTADIIH sont présentés en tableau 6 de l'article présenté en section 3.4.1.