STUDIES ON THE CLINICAL PHARMACOLOGY OF ANTIRETROVIRAL AGENTS

Filip Josephson M.D.

Stockholm 2009
All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet
Printed by Larserics Digital Print AB
© Filip Josephson, 2009
ABSTRACT

This thesis begins with an account of the development of the antiretroviral treatment paradigm for HIV infection, focusing on the emergence of present day principles of antiretroviral therapy (e.g., how to rationally combine drugs, the monitoring of treatment effect, the definition of treatment failure, and the use of “ritonavir-boosting” to augment drug exposure). The implications of the evolving treatment paradigm for drug development and the conduct of clinical trials are emphasized (e.g., the definition of study populations, the role of biomarkers and the definition of endpoints).

Following this, the promises and limitations of therapeutic drug monitoring (TDM) of antiretrovirals is described, as well as the pharmacogenetics of HIV therapy, and the mechanisms of the substantial drug-drug interaction potential of some major antiretroviral agents. The introduction concludes with a discussion of the conduct, findings and implications of the four papers comprising the thesis, insofar as there is a need for justification or further comments beyond what is stated in the articles, or where important new data have appeared since the time of writing the papers.

Paper 1 reports a study of the impact of pharmacogenetic variability at CYP3A on the metabolism of the HIV protease inhibitor (PI) saquinavir. This drug was shown to be a CYP3A5 substrate.

Paper 2 describes a pilot study of ritonavir-boosted atazanavir maintenance monotherapy, which was prematurely terminated due to an excess of protocol-defined treatment failure. Atazanavir plasma drug concentrations did not predict failure in this study, though on-treatment serum bilirubin (an increase of which is an exposure dependent side effect of atazanavir) did.

Paper 3 describes the effects of three different antiretroviral drug regimens (based on efavirenz, ritonavir-boosted atazanavir and ritonavir-boosted lopinavir) on the plasma concentration of 4β-hydroxycholesterol, a suggested endogenous marker of CYP3A activity. This marker performed according to hypothesis in the efavirenz and ritonavir/boosted atazanavir arms (increase and decrease, respectively, due to CYP3A induction and inhibition), but failed to clearly reflect the profound CYP3A inhibition in the ritonavir-boosted lopinavir arm.

Paper 4 reports the relation of the plasma concentrations of efavirenz, atazanavir and lopinavir, and treatment outcome in the NORTHIV clinical trial in treatment-naïve patients. No relation of drug concentration and effects was demonstrated thought there was a trend towards a relation of plasma bilirubin and outcome in patients treated with atazanavir.

It is concluded that: (a) Genetic variability in CYP3A5 is unlikely to have any measurable clinical impact on present day boosted PI regimens. (b) Though boosted PI maintenance monotherapy may be an option in selected patients, there are important concerns about, e.g., CNS penetration. Increasing PI doses beyond those presently recommended is unlikely to alter the efficacy of boosted PI monotherapy. (c) Plasma bilirubin is a useful marker for adherence in atazanavir-treated patients. (d) Routine TDM of antiretrovirals is not mandated.
LIST OF PUBLICATIONS


CONTENTS

1 Introduction ................................................................................................... 1

2 The evolution of the combination antiretroviral treatment paradigm ....... 3
  2.1 Proof of the principle and efficacy of antiretroviral therapy ............. 3
  2.2 The durability of antiretroviral effect ............................................. 4
  2.3 When to start treatment ............................................................... 4
  2.4 Zidovudine dosing reassessed ...................................................... 5
  2.5 The emergence of combination antiretroviral therapy .................... 6
  2.6 HIV “latency” characterised by a massive production of virions .... 6
  2.7 A theoretical rationale for drug resistance and maximally suppressive
     therapy ........................................................................................... 7
  2.8 The efficacy of combination therapy ......................................... 8
  2.9 Plasma HIV-RNA to monitor treatment effect ............................ 9
  2.10 A new paradigm for clinical trials .......................................... 9
  2.11 The standard triple combination: 2 NRTI + 1 PI or 1 NNRTI .... 10
  2.12 When to switch ....................................................................... 10
  2.13 Potency and barrier to resistance ............................................. 12
  2.14 Adherence .............................................................................. 14
  2.15 Drug interactions ..................................................................... 15
  2.16 Long-term adverse effects ...................................................... 16
  2.17 Intensification and simplification strategies ............................. 16
  2.18 Ritonavir boosting ................................................................. 17
  2.19 When to start reassessed ......................................................... 18
  2.20 The barrier to resistance of ritonavir-boosted PIs ....................... 19
  2.21 Three active drugs, chosen on the basis of resistance testing ..... 19
  2.22 Therapeutic Drug Monitoring (TDM) ...................................... 20

3 An introduction to HIV TDM ................................................................. 21
  3.1 The minimal effective concentration (MEC) .............................. 21
  3.2 The rationale for routine TDM .................................................. 23
  3.3 A technology still in its infancy ................................................. 24
  3.4 Intra-patient variability as an obstacle to effective TDM ......... 25
  3.5 The plasma matrix and the drug concentration at the site of action 25
  3.6 Which PK parameter is best associated with efficacy? .......... 26
  3.7 Drug exposure and the risk of resistance ................................ 27
  3.8 The barrier to resistance revisited ......................................... 27
  3.9 Drug exposure and mutant drug susceptibility ........................ 29
  3.10 Resistance and pharmacokinetics ............................................. 30
  3.11 Viral fitness and the barrier to resistance ................................ 30
  3.12 The MEC and in vitro susceptibility data ................................. 31
  3.13 The inhibitory quotient (IQ) .................................................... 33
  3.14 The genetic inhibitory quotient (GIQ) ..................................... 34
  3.15 Tolerance and the problem of the “toxicity” limit .................. 35
  3.16 The scope of TDM ............................................................... 36

4 On the evidence base of the therapeutic range .................................. 38
  4.1 The Lausanne cohort ............................................................ 38
4.2 Defining the therapeutic range of efavirenz .................................... 40
4.3 The Stockholm cohort ...................................................................... 40
4.4 Challenging the Marzolini conclusions ........................................... 41
4.5 The dose ranging studies of efavirenz ............................................. 42
4.6 The demands of atazanavir TDM .................................................... 43
4.7 TDM and the newer antiretroviral agents ........................................ 45
5 Pharmacogenetics and drug-drug interactions ..................................... 46
5.1 The pharmacogenetics of antiretroviral therapy ............................... 46
5.2 Antiretroviral therapy and drug-drug interactions............................ 47
6 The NORTHIV trial ................................................................................... 49
7 Paper I: CYP3A5 and saquinavir metabolism........................................... 51
7.1 CYP3A5 ............................................................................................ 51
7.2 The Tanzanian connection .................................................................. 52
7.3 Fortovase® ....................................................................................... 53
7.4 Saquinavir is a substrate of CYP3A5 .................................................. 53
8 Paper II: Atazanavir, bilirubin and boosted PI maintenance monotherapy 55
8.1 The design of boosted PI monotherapy trials .................................. 55
8.2 The evolving view of PI monotherapy ............................................ 55
8.3 The TDM approach in the atazanavir monotherapy study ............... 57
8.4 The unpublished manuscript ............................................................ 59
8.5 Failure due to the use of acid reducing agents? ............................... 62
8.6 Plasma bilirubin as a marker of virologic outcome in atazanavir therapy 62
9 Paper III: Testing a biomarker of CYP3A activity in the setting of NORTHIV 67
9.1 The discovery of a putative endogenous CYP3A biomarker .......... 67
9.2 Long half-life a merit or a problem? .............................................. 68
9.3 The NORTHIV trial as a vehicle to investigate 4β-hydroxycholesterol 69
9.4 CYP3A induction and inhibition ..................................................... 69
10 Paper IV: TDM of PI and NNRTI ............................................................. 73
10.1 Notes on the PK sampling strategy and trough imputation .......... 73
10.2 TDM practice as an obstacle to its study ..................................... 74
10.3 The case of efavirenz related CNS side effects ........................... 74
11 Conclusions ........................................................................................... 76
12 Acknowledgements ................................................................................ 77
13 References ............................................................................................. 79
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired Immune Deficiency Syndrome</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under concentration-time curve</td>
</tr>
<tr>
<td>b.i.d</td>
<td>Twice daily</td>
</tr>
<tr>
<td>Cmax</td>
<td>Maximal concentration</td>
</tr>
<tr>
<td>Cmin</td>
<td>Minimum concentration</td>
</tr>
<tr>
<td>Ctrough</td>
<td>Trough concentration</td>
</tr>
<tr>
<td>DAA</td>
<td>Directly acting antiviral</td>
</tr>
<tr>
<td>Emax</td>
<td>Maximal efficacy</td>
</tr>
<tr>
<td>EMEA</td>
<td>European Medicines Agency</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>GIQ</td>
<td>Genetic inhibitory quotient</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C Virus</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HPLC-UV</td>
<td>High performance liquid chromatography with UV detector</td>
</tr>
<tr>
<td>IC50/95</td>
<td>Inhibitory concentration 50/95%</td>
</tr>
<tr>
<td>IQ</td>
<td>Inhibitory quotient</td>
</tr>
<tr>
<td>ITT</td>
<td>Intention to treat</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography – mass spectrometry</td>
</tr>
<tr>
<td>MEC</td>
<td>Minimal effective concentration</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimal inhibitory concentration</td>
</tr>
<tr>
<td>NNRTI</td>
<td>Non-nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>NRTI</td>
<td>Nucleoside (or nucleotide) reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>OT</td>
<td>On treatment</td>
</tr>
<tr>
<td>PD</td>
<td>Pharmacodynamic</td>
</tr>
<tr>
<td>PI</td>
<td>HIV protease inhibitor</td>
</tr>
<tr>
<td>PK</td>
<td>Pharmacokinetic</td>
</tr>
<tr>
<td>q.d.</td>
<td>Once daily</td>
</tr>
<tr>
<td>TDM</td>
<td>Therapeutic drug monitoring</td>
</tr>
<tr>
<td>TLOVR</td>
<td>Time to loss of virologic response</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

The principal concern of the discipline of clinical pharmacology is to understand the determinants of drug exposure, response and the relation of exposure to response, and to translate this understanding into evidence-based advice on how to maximise drug effects and minimise side effects. Antiretroviral agents are drugs for the treatment of HIV infection (in the case of our studies, the treatment of HIV-1) that directly interfere with the replicative cycle of the virus. Thus, the title of this book refers to studies aimed at optimizing the treatment of HIV infection. More particularly, the primary, though not exclusive, topic of this thesis is the exposure-response relations of efavirenz, atazanavir and lopinavir.

I shall take the liberty of not beginning this account with a general overview of HIV and HIV-related disease, the mechanisms of action of the antiretroviral drugs or the fundamental principles of clinical pharmacology, the way that is common in doctorate theses. For this I refer to textbooks of the sort one usually quotes when following the traditional path. As regards the presently available antiretroviral agents and their recommended use, recent Swedish, British and US guidelines are available [1-3].

Instead, I shall begin by relating the history of the antiretroviral treatment paradigm, from the advent of zidovudine monotherapy to the time when the first of the studies reported here were initiated. The intent of this is to illustrate the emergence of present day practice, and thereby to describe the general background to our inquiries for readers not mainly concerned with HIV therapy. Hopefully, this narrative will be more captivating than a textbook approach might have been, while still serving as a proper introduction to the particular concerns characteristic of the field, due to the biology of HIV and its replication.

In my endeavor to describe this period prior to my entering the field, I will use the published consensus documents of the National Institute of Allergy and Infectious Diseases (NIAID) state of the art conferences 1990 and 1993 [4, 5], and the International AIDS Society-USA (IAS-USA) guidelines from 1996 to 2002 [6-10], as a point of departure. These particular sources of collective memory are, in principle, arbitrarily chosen; one might certainly have approached this topic in other ways. However, these guidelines of yesterday seem highly informative, in so far as they provide the evolving rationale for much of what is presently taken for granted.
The further outline will go on with an attempt to delineate the general view on therapeutic drug monitoring (TDM) of antiretroviral agents at the time of the inception of this project. In what follows, TDM refers to plasma drug concentration measurements in clinical care, with the intent of adjusting the dose (or making some other sort of intervention, such as adherence counseling or switching therapy) in case the result obtained is not thought as should be.

I will then consider the evolution of the HIV TDM practices in Stockholm during the period of my studies, mainly insofar as this is necessary to justify decisions within this thesis project, but also to provide the rationale for advice given to the HIV clinicians during the period in question. After this, I will briefly review the issues of HIV pharmacogenetics and antiretroviral drug interactions. Following this introduction to the field in general and in particular, I shall discuss the four papers of this thesis, and finish off with some general conclusions.
2 THE EVOLUTION OF THE COMBINATION ANTIRETROVIRAL TREATMENT PARADIGM

2.1 Proof of the principle and efficacy of antiretroviral therapy

In 1990, the NIAID sponsored a state of the art conference in order to “evaluate available scientific information and to resolve safety and efficacy issues related to the use of zidovudine in the treatment of HIV-infected patients with few or no symptoms of disease.” The nucleoside analogue zidovudine (at the time often referred to as azidothymidine or AZT), which was originally developed as a putative anticancer agent, is an inhibitor of the HIV reverse transcriptase. Three randomised clinical trials of zidovudine monotherapy are said to inform the consensus statement: the ACTG-016, -019 and -002 protocols:

The 016 study evaluated the efficacy and safety of zidovudine in patients with mildly symptomatic disease. Zidovudine at the dose of 200 mg x 6 (six times daily) was compared with placebo. Subjects were stratified according pretreatment CD4+ T-cell counts.

The 019 protocol studied the safety and efficacy of zidovudine therapy in asymptomatic patients with HIV, comparing higher (300 mg x 5) and lower (100 mg x 5) doses to placebo. Again subjects were stratified by CD4+ T-cells.

The 002 protocol investigated the efficacy and safety of a lower dose of zidovudine in the treatment of advanced disease (patients were randomised after a first episode of Pneumocystis Carinii pneumonia, an AIDS-defining diagnosis). The treatments were zidovudine 250 mg x 6, 200 mg x 6 for 4 weeks, then reduced to 100 mg x 6 (of note, a sort of an induction-maintenance strategy, a theme to which we will return), and placebo.

Disease progression, AIDS and death were the rock hard endpoints in these studies, but CD4+ T-cell count and the detectability of HIV antigen (of note, not HIV-RNA) in plasma were also investigated. There were significant effects on composite endpoints of progression to AIDS, advanced AIDS-related complex (ARC) or death in patients with baseline CD4+ T-cells < 500/μL, and on survival in patients with AIDS. Patients with detectable HIV antigenemia converted to undetectable, opaquely reflecting a direct antiviral effect. These studies delivered resounding proof of the principle as well as efficacy of antiretroviral therapy by direct inhibition of the
replicative cycle of the virus. An important finding in studies -019 and -002 was that the efficacy of the lower doses of zidovudine was at least on par with the higher, with considerably less side effects.

The path had been identified. It would be antiretroviral therapy rather than, e.g., immunotherapy (or therapeutic vaccines) that would “make a difference” in the decades to come. This will probably remain the case within the foreseeable future.

2.2 The durability of antiretroviral effect

Soon it would be clear that the antiretroviral effect of zidovudine monotherapy was generally not sustainable, as viral susceptibility to zidovudine decreased stepwise with the accrual of resistance mutations. This mechanism, however, was not immediately fully recognised [4, 5]. Further experience would demonstrate that this principal tendency of antiretroviral drug action could only be overcome if the sum antiretroviral potency and barrier to resistance of the antiretrovirals used concomitantly was sufficient to bring down and retain viral replication below a “threshold” value, which is reflected by a maximal steady state plasma HIV-RNA concentration compatible with durable effect. As by curious incident, this limit is presently thought to roughly coincide with the limit of quantification of routine tests for plasma HIV-RNA quantification, and to be somewhere in the region of 20-200 copies/mL (as a note, the great majority of untreated patients have levels of viremia between \(10^3\)-\(10^6\) copies/mL). Even with suppression below the target levels in treatment, however, viremia is still demonstrable, if sufficiently sensitive methods are used [11, 12]. Thus, there may not exist any true “threshold” for the selection of drug resistance mutation, but rather a level of HIV replication at which the accrual rate of resistance mutations becomes so low that the number of patient-years that would need to be followed to find a mutation would be in the order of decades [13].

2.3 When to start treatment

The conclusions drawn from these early experiences is that treatment is recommended for both symptomatic and asymptomatic HIV-infected patients whose CD4 T-cell counts are below 500/\(\mu\)L blood, based on the results from studies 016 and 019 that were summarised above. The optimal use of antiretroviral therapy, however, is recognised as unclear: “The demonstrated ability of zidovudine to delay the decline
of CD4 cells and to affect viral activity, as measured by serum levels of HIV antigens, show that there may be benefits of an even earlier treatment initiation. However, if the effects of zidovudine are indeed transient due to the development of resistance (...) it is considered possible that later initiation of the drug may be more efficacious.” This delicate balance will recur as a major strategic concern in subsequent guidelines, and when to start is still an issue not fully settled.

2.4 Zidovudine dosing reassessed

As the state-of-the-art conference gathers again two years later in 1993 [5], it is noted that antiretroviral therapy is starting to become complex. The group complains that it has often felt the need to extrapolate from whatever trials have been done, to patients in other situations that may not be properly accounted for by these trials. The treatment recommendation now breaks down into 12 different scenarios based on (a) symptoms, (b) CD4+ T-cell counts, (c) prior treatment history, and (d) response to ongoing (or past) therapy. Prior treatment history will evolve into a central determinant of study populations in antiretroviral clinical trials, with efficacy in the treatment naïve and the treatment experienced being investigated under different protocols.

As first line therapy, zidovudine at a dose of 600 mg x 3 was recommended in 1993. The new, less frequent dosing regimen was “based on clinical experience which suggests that patient compliance may be enhanced without obvious changes in safety or efficacy.”

The future would herald much in the understanding of nucleos(t)ide reverse transcriptase inhibitor (NRTI) pharmacology, from, e.g., the mechanisms of mitochondrial toxicity [14-17], to the intracellular pharmacokinetics not necessarily reflective of plasma pharmacokinetics (PK) [18-22]. The latter insight would drive down zidovudine dosing regimens to twice daily, and eventually inform the development of the presently used, co-formulated once daily NRTI regimens (tenofovir/emtricitabine and abacavir/lamivudine). The impact of clinical pharmacology was to be great in the field of antiretroviral drug development. Yet it is somewhat ironic that its first appearance at the centre of the stage, was with the misguided notion that zidovudine needed to be administered 4-6 times a day based on plasma PK.
2.5 The emergence of combination antiretroviral therapy

The most important herald of what is to come that we find in the 1993 guideline may be this statement: “Clinical judgment might dictate use of combination therapy in some situations. Definite data, however, are sorely lacking”. Such data would appear in the following years. Though dual therapy could not consistently be demonstrated to be more effective than monotherapy in pretreated patients (e.g., [23]), the principal superiority of combination therapy was soon an established and indubitable fact (e.g., [24-27]).

2.6 HIV “latency” characterised by a massive production of virions

Along with novel and truly revolutionary treatment options, the next US consensus guidelines, the first IAS-USA recommendations of 1996 [7], presents a new paradigm to inform treatment decisions, monitoring and goals, as well as the design, conduct, endpoint assessment and interpretation of clinical trials: that of virology (rather than that of immunology which, along with clinical staging, had dominated the definitions of study populations and endpoints in the early years). Three years had gone since the previous document. “Important advances in understanding the biology and treatment of HIV infection have occurred during the past 18 months”, we are told. “In light of this, the 1993 guidelines are no longer applicable”. There is no doubt that these were exiting days; indeed, we are arriving at the time when the prognosis in HIV infection was so radically altered due to therapeutic advances, that antiretroviral treatment effects were to become detectable in gross epidemiological data [28].

Contrary to early reports, hampered by low assay sensitivity, it had previously been shown that viremia is present in most antibody-positive patients (e.g., [29-31]). Now further data on viral kinetics had emerged, demonstrating that around $10^9$-$10^{10}$ virions are produced each day [32, 33]. The idea of HIV “latency” as a clinical stage characterized by low or absent HIV replication, was forever dispelled (though certainly there are latently infected cells, thought to be a formidable obstacle to cure [34-36]). Also, the loss of CD4+ T-cells characteristic of HIV pathogenesis was no longer a riddle to be explained, but clearly the consequence of productive infection.

Other important novelties included the emergence of assays for everyday use in the clinics to measure viral load in individual patients (plasma HIV-RNA), the availability of several new effective drugs, and the aforementioned demonstration that combination therapy is more effective than zidovudine monotherapy [7].
2.7 A theoretical rationale for drug resistance and maximally suppressive therapy

Emerging data on drug resistance were now interpreted in light of the full realization of the extent of HIV replication. The development of clinically relevant drug resistance due to point mutations in the genes coding for the viral drug targets, a matter of speculation in the 1993 guidelines, now appears like a foregone conclusion in the light of an immense production of virions by a sloppy reverse transcriptase steadily producing mutant variants:

“It is recognised that, given the vast replication rate of HIV, if viral suppression is incomplete, it would be expected that viral variants with reduced susceptibility to antiretroviral drugs will evolve over time. In that viral variation begins with the first cycles of replication during primary infection, as HIV-1 infection proceeds, a subpopulation of viral variants with reduced susceptibility to antiretroviral drugs evolves, even in the absence of selective pressure. The prevalence of such mutants in the population of virus prior to the initiation of antiretroviral therapy is a function of the number of prior rounds of viral replication, the mutation rate, and the selective advantage (fitness) possessed by wild-type virus over variants that have incorporated mutations conferring drug resistance.”

Thus, at this point the consensus group affords an account of the phenomenon of drug resistance based on emerging knowledge of the fundamental biology of HIV, and viral drug resistance is affirmed as a concern of central importance for the field of antiretroviral therapy. Indeed, mathematical modeling suggests that this massive viral turnover coupled with a “sloppy” reverse transcriptase leads to the generation of every possible single point mutation more than 10,000 times per day [37]. Mutants with reduced drug susceptibility are continuously generated in untreated individuals, but not selected for in the absence of drug pressure (antiretroviral drug resistance provides a striking example of the operation of Darwinistic principles).

As a consequence of this recognition, the general tendency of reasoning is: the more viral suppression the better, as this will delay the selection of drug resistant variants and consequent loss of antiretroviral efficacy - which is expected to translate into more sustained clinical benefit. Yet, we are cautioned that the full clinical benefits of complete viral suppression, as well as the durability of this effect, are still unknown.
The importance of high doses and good adherence are affirmed in order to avoid resistance. An important testimony of the emerging general concerns of the field is the expert recommendation that “it is better to stop protease inhibitor treatment than to lower the dose”. In this line of reasoning, we note a clear recognition of the notion that subtherapeutic dosing of antiretrovirals might not only be ineffective, but in fact harmful with respect to future treatment options. This would be an important principle informing the field of antiretroviral therapy in general, as well as the TDM practice of which I shall later account.

At the time of my writing this thesis, it is interesting to note that the paradigm of understanding that has developed around HIV drug resistance (to which I shall return several times in this account) is now informing the investigations of DAAs against HCV, where it is clear that monotherapy with HCV protease inhibitors, and also drugs from other classes, rapidly select for resistant strains, often leading to virologic breakthrough within a week of drug administration [38]. While there are great apparent similarities between this and, e.g., the development of cytidine analogue or NNRTI resistance in HIV therapy, the extent of the clinical consequences of HCV drug resistance have still not been clarified, as there are important differences between the biology, as well as the treatment paradigms of HIV and HCV.

2.8 The efficacy of combination therapy

Travelling back to 1996 again, the magnitude of HIV protease inhibitor (PI) activity had clearly been beyond expectations, and was generating great excitement about the future: “In one large randomised phase III trial, in patients with CD4 cell counts of 50-300/mL and at least 16 weeks of prior zidovudine therapy, the combination of saquinavir and zalcitabine was significantly better in terms of clinical outcome and survival, than was saquinavir or zalcitabine alone”, it is noted in the consensus document. The additive or synergistic effects of combining drugs from different classes had been demonstrated. “The combination of zidovudine, lamivudine and indinavir led to >90% of a (my comment: certainly selected) population having <50 copies/mL at 6 months (…) Addition of ritonavir to an existing regimen (or no regimen) reduced progression to AIDS and mortality by about 50%.” On a cautionary note, the lack of efficacy and safety data beyond 52 weeks is mentioned.
2.9 Plasma HIV-RNA to monitor treatment effect

The indication for HIV treatment, apart from symptomatic disease, was now considered to be CD4+ T-cells <500/mL (or <25%). Treatment was also recommended in all patients with plasma HIV-RNA >30,000 copies/mL and to be considered in all patients with HIV-RNA >5000 copies/mL. In fact, that covers the majority of patients. Despite this addition of virologic concerns to treatment decisions, it is noted that “no clinical data support treatment of asymptomatic patients with CD4+ >500/mL.” Such discrepancies tell us much about what is anticipated in the field.

Therapeutic failure is now defined as “virologic rebound” (e.g., a return towards or within 0.3-0.5 log10 of pre-treatment plasma HIV-RNA value). This has important implications for patient care: It is emphasized that patients should be monitored so that virologic rebound can be detected prior to clinical disease progression. This, of course, remains a principally important point still today, which has proved to be difficult to implement in many resource-poor settings.

2.10 A new paradigm for clinical trials

It is recognized that guidelines have to be anchored in data from double-blinded randomized controlled trials, but also in information from trials in progress, “assessing available virologic and immunological endpoints.” Take note: this statement indicates not only the great pace of progress of the field, but also that the hard clinical endpoints that had informed the first generation of antiretroviral therapy trials, under the sway of the immunological paradigm, were no longer considered necessary nor justifiable, given the operating characteristics of the new biomarker, plasma HIV-RNA.

It is noted that up till now entry criteria for the pivotal trials have centered on pre-treatment symptom status, prior treatment history, and CD4+ T-cell count at entry. However, it had recently been clearly demonstrated in combination trials that “reduction in plasma HIV-RNA is associated with increased survival and decreased progression to AIDS.” Therefore it was apparent that further insistence on hard clinical endpoints in clinical trials would be inhibitory to drug development, insofar as these demand clinical trials with more patients and longer periods of observation than do efficacy demonstrations in terms of inhibition of viral replication, now considered a valid surrogate for disease progression and death.

Plasma HIV-RNA was now considered superior to CD4+ T-cell count for determining the indication for and response to treatment. I emphasize this, because this
is arguably the single most important novelty of the recommendations of the 1996 document (which yet does not formally recognize triple therapy as a first line option), in terms of the further development of the field.

### 2.11 The standard triple combination: 2 NRTI + 1 PI or 1 NNRTI

The totality of evidence in 1997 seemed to favor an earlier treatment initiation; in fact, it is recommended that all patients with plasma HIV-RNA >500-1000 copies/mL should be treated [8]. The major therapeutic strategy is now a three-drug combination to attain sufficient antiretroviral potency. Undetectable viremia is the goal, and the preferred regimen is anyone that leads to plasma HIV-RNA levels below the limit of detection (e.g., <400-500 copies/mL).

The paradigm of the “standard” regimen of 2 NRTI + 1 PI or 1 non-nucleoside reverse transcriptase inhibitor (NNRTI), which has survived into present day guidelines (with the important modification of ritonavir-boosting, further discussed below), was “considered likely to be appropriate in most cases”. Also, these are the first recommendations that take substantial account of the emerging NNRTI drug class. An “important principle” in the use of NNRTIs has been established through the development process of nevirapine - that the activity of NNRTIs is maximized when combined with other drugs to which the patient is naïve (an issue that will be further discussed below).

### 2.12 When to switch

There are important deliberations on when to change therapy. As the treatment goal in 1997 is undetectable viremia, detectable viral load is formally considered an indication to change therapy. In practice, though, it is recognized that, due to the limited number of drugs, it may be better to wait until plasma HIV-RNA has reached 2000-5000 copies/mL.

As I am writing, the exact viral load threshold that should be considered to define viral failure, taking into account the generally recognized but quantitatively ill defined risk of resistance development with low grade viremia (50-500 copies/mL), cannot be considered subject to real consensus, as evidenced by the still variable definition of virologic treatment failure in clinical trials.

The question is of principal importance because it is desirable to switch a failing regimen that is not expected to re-suppress viral replication, at a time when the
accrual of drug resistance mutations has minimally compromised future treatment options, even though there may still be some remaining immunological and clinical benefit of the failing treatment regimen. Of interest, the putative further benefit of non-suppressive regimens, in the unlucky event that there are no treatment options available, may depend on the fitness cost of resistance mutations, as well as the residual activity of the regimen, and may therefore be higher with PI-based than NNRTI based regimens [39, 40]. These issues are of considerable importance for the rational management of HIV in resource-poor settings [41].

As a background, studies have suggested that the incidence rate of new HIV drug resistance mutations for patients with ongoing, non-suppressive, mainly unboosted PI-based regimens, and plasma HIV-RNA levels between 1000-10,000 copies/mL, may be somewhere around 1-2 per person/year [13, 41]. The precise replication rate associated with the highest accrual rate of drug resistance is not established – as indeed would be expected given, for instance, that the replication rate on non-suppressive therapy is, in its turn, a function of drug resistance. Though one study reported the highest rates between 1000-10,000 [42], it is notable that standard population sequencing assays often fail when HIV-RNA is below 1000 copies/mL, whereas in individuals with ongoing viremia >10,000 despite therapy, adherence will either be compromised, or the virus will be approaching wild-type fitness and the selection rate for new mutations will consequently be low. Indeed, more than one study has failed to correlate plasma HIV-RNA within the interval of 1000 copies/mL and upward with the accrual rate of resistance mutations (e.g., [43-45]).

Several investigations have demonstrated the development of drug resistance mutations in patients with low grade viremia (<1000 copies/mL) (e.g., [46-50]). On the basis of very scarce data, a resistance mutation accrual rate of 0.16 mutations per person year has been inferred in patients with non-suppressive therapy and ongoing viremia in the range of 50-400 copies/mL, which would be about 10-fold less than in the interval 1000-10,000 copies/mL [13].

There are important limitations to these assumptions. The studies underlying these assessments of the risk of resistance as a function of time with a non-suppressive regimen were conducted in populations where most patients had resistance mutations at baseline (indeed a prerequisite for the studies was continuation of non-suppressive therapy, indicative of a lack of treatment options), and the patients were mainly treated with unboosted PI + NRTIs. Drug pressure with boosted PIs (see below) is much higher, and extrapolation of data to this situation may not be valid. Also, in contrast to
PI, single NNRTI resistance mutations yield high grade resistance at a low fitness cost [51], which may affect the further accrual rate of mutations.

In general, the rate of accumulation of drug resistance mutations with non-suppressive therapy is expected to vary considerably depending on drug regimen, prior drug resistance, level of adherence and plasma HIV-RNA level. It is also expected not to remain the same over time, as the fitness of the virus at the given drug pressure improves with the accumulation of drug resistance mutations.

Of interest, later studies demonstrate that the frequencies and extent of acquired resistance after the failure of first-line drug therapy in resource-poor settings is much greater than in the developed world or in randomized trials, as recently reviewed and meta-analyzed by Gupta et al [52]. This is certainly due to insufficient monitoring of viral load. Indeed, in resource poor settings, patients are in most cases monitored by CD4+ T-cell counts (if undergoing laboratory monitoring of therapeutic response at all) rather than viral load [41]. This biomarker should be used to assess immunological status and risk of disease progression; it is very insensitive for detecting loss of virologic suppression [53-55]. This practice is highly inoptimal from the point of view of drug resistance.

In summary, as stated earlier, the accrual rate of drug resistance mutations is approaching zero when plasma HIV-RNA is <50-200 copies/mL; the general assumption is that, if indeed resistance is selected for at all in this range with combination therapy, the likelihood is such that it is practically negligible.

### 2.13 Potency and barrier to resistance

In the 1997 discussion of the rationale for the first line recommendation, we find a very interesting statement that is somewhat illuminating of the conceptions of antiretroviral therapy and drug resistance at the time: “It is the potency of a therapeutic regimen that is important, and not the number of drugs per se”.

Today it is considered that the number of active drugs per se may in fact, in some cases, be crucial to a regimen being maximally effective, regardless of actual potency. This is because of the need for a barrier to resistance, as distinct from potency. This barrier can be thought of as consisting of the total number of mutations necessary to reduce the potency of a treatment combination so as to permit viral replication above a certain limit, with consequent further accrual of resistance (as discussed above) and thereby further loss of efficacy.
I do not feel competent to carry on a higher-level discussion of the stochastic principles that explain this circumstance, which concerns the likelihood of the generation of virions combining specific numbers of point mutations conferring resistance. The general lesson, however, confirmed by the lack of efficacy of certain drug combinations, is that a barrier to resistance of at least three mutations appears to be needed for an antiretroviral regimen to exhibit optimal durability of response.

A prime example of this principle in operation is provided by the high failure rate when using a combination of tenofovir, abacavir and efavirenz in treatment-naïve patients, which is not readily explainable in terms of lacking potency [56, 57]. Other experiences made in trials of different drug combinations (which were initially chosen on the basis of a very limited virologic or pharmacological rationale [58]), also demonstrate the need to think beyond potency to understand why some antiretroviral drug combinations work and others don’t. I refer not only to the differential need for support from co-treating agents of efavirenz or raltegravir on the one hand, and boosted PIs on the other (a crucial difference which will be revisited in this account) but also to issues such as how to combine NRTI, the emerging lesson being that analogues of the same nucleoside are liable to be non-additive, antagonistic or toxic when used together [59-65]. Yet further examples would be the later insight not to combine two NNRTI [66] or indeed, as I would personally think, not to use dual protease inhibitors (except for ritonavir-boosting, see below). In my view, this practice has not been convincingly shown to confer any advantage over one optimally chosen, boosted PI.

Thus, two drugs that have similar or greatly overlapping patterns of resistance \textit{in vivo} or \textit{in vitro} will probably not be rational to combine. Resistance profile is, as it were, a diffracted image of the receptor, and if a point mutation significantly affects viral susceptibility to more than one agent in the treatment combination, their combination tends not to raise the barrier to resistance of the regimen. Such insights from HIV drug development are recognized by those presently involved in the development of DAAs against HCV.

Later I will elaborate more extensively on the notions of potency and barrier to resistance, which most workers in the field today would regard as distinct. In the consensus document of 1997 however, they appear to be thought of as one and the same thing, as reflected by the insistence that sufficient potency would lead to viral suppression, and viral suppression would prevent the emergence of drug resistance [8].
### 2.14 Adherence

In the 1997 guidelines document, adherence is emphasized as an emerging concern of central importance, and it is reaffirmed that “*dose reductions of PIs are to be avoided at all costs*”. Less than excellent adherence will rapidly lead to viral breakthrough and resistance. Indeed, the recommendation on “when to treat” has now evolved into: all patients with detectable HIV-RNA in plasma that are “*committed to lifelong adherence*”. Adherence appears as an existential choice, a personal “commitment”.

The study of the relation of adherence to treatment outcome is of great interest, not only for the conduct of antiretroviral therapy in general, but also for the particular issues of the relation of exposure and effect investigated in this thesis. As a methodological note, electronic recording of medication bottle opening, pill count and self-assessment questionnaires are the most common methods for adherence measurement, with none of these clearly considered the golden standard [67].

Early concerns were fuelled by reports showing that near-perfect adherence was necessary to achieve maximal likelihood of virologic suppression with combination antiretroviral therapy [68-70]. These studies were conducted in predominantly treatment-experienced patients using unboosted PI + 2 NRTI. In this context, it is interesting to revisit the paper by Paterson et al, which gave rise to the so-called “>95% adherence rule” [70]. It reports a mostly treatment-experienced population recruited between 1997-99, mainly using nelfinavir or unboosted indinavir-based regimens. Adherence was measured with a microelectronic monitoring system (MEMS). Virologic response rated (HIV-RNA <400 copies/mL) dropped with decreasing adherence, with failure rates of 22% in patients with 95% or higher adherence, 61% in those with 80-95% adherence and 80% in those with below 80% adherence (median follow-up duration 6 months). These data, however, are more reflective of the inoptimal efficacy of first generation, unboosted PI regimens, than of the adherence requirements for present day antiretroviral regimens.

Subsequent studies indicate that near-perfect adherence is, in fact, probably not necessary for a satisfying treatment outcome in most circumstances (though bluntly stating this somehow feels heretic, or at least un-edifying). Several datasets in patients with variable treatment experience have shown that the adherence requirements for near maximal efficacy using NNRTI-based regimens is considerably lower than with unboosted PIs [71, 72]. In fact, in a small sample, NNRTI adherence rates of
approximately 55-75% were compatible with maximal efficacy [71]. Thus, NNRTI regimens appear more “forgiving” of non-adherence than are unboosted PIs.

When lopinavir (a boosted PI, see below) was used in a highly treatment experienced population, no drop in virologic response was seen with adherence rates (again measured with MEMS) down to approximately 75%. Even within the subgroup of patients with estimated adherence of approximately 25-50% adherence, the proportion of patients with virologic suppression at 24 weeks (plasma HIV-RNA <75 copies/mL) was over 50% [73]. Of interest, in these trials, average adherence has been around 75% [74]. Thus, on the relation of adherence to the antiretroviral efficacy and likelihood of virologic suppression, it could be that adherence rates of about this magnitude are sufficient for maximal efficacy when treating non-resistant virus with present day regimens. It is important to note, however, that adherence needs not only be related to the likelihood of durable virologic suppression, but also to the risk of emerging drug resistance in case of failure. In this context, not only the proportion of prescribed doses must be considered, but also the pattern of non-adherence. For instance, temporary treatment discontinuations rather than randomly missed doses are likely to have different impacts on different antiretroviral regimens. These issues will be further discussed below.

Also, while adherence certainly remains a crucial concern in antiretroviral therapy, and lack of it the most important cause of failure, the improved “forgiveness” of non-adherence characteristic of second generation (this most probably applies also the newest agents, though I am not aware of any published data) triple combination antiretroviral treatment regimens clearly demonstrate the increased efficacy of antiretroviral drugs and dosing regimens. I ask the reader to take note of this, because, as will be argued, several aspects of the studies reported in this thesis, which may seem unwarranted today, may be understood (and hopefully justified) with reference to early experiences of inoptimal combination antiretroviral regimens, at a time when the drastically increased efficacy of, e.g., boosted PI over first generation unboosted PI, was perhaps still to be fully recognized.

2.15 Drug interactions

In the IAS-USA 1998 guidelines [9], there is a novel, extensive table listing the PK drug interactions between different antiretrovirals, bearing witness of the perceived magnitude of this emerging problem (which is mainly, though not exclusively, due to
the effects exerted by the PI and NNRTI on cytochrome P450 mediated drug metabolism, and mostly so due to ritonavir-boosting, as further elaborated below). There is no doubt that drug interactions were by now considered a major problem. This will certainly remain the case as long as CYP3A inhibition is used to enhance exposure to antiretrovirals.

2.16 Long-term adverse effects

The 1998 expert panel is clearly worried by the emergence of new and long-term adverse effects, particularly with PI containing regimens. The precise incidence, mechanism and long-term implications of these events are felt still in need to be defined. The reference is, of course, to the metabolic side effects of many antiretrovirals (insulin resistance, dyslipidemia) and the various effects on body fat distribution sometimes grouped together under the label “lipodystrophy syndrome”. In the years to follow, this latter category would be essentially subdivided into lipoatrophy (nowadays considered principally to be a side effect of thymidine analogues, which are no longer recommended) and lipohypertrophy (still thought to be associated with the PIs) [75]. Indeed, in 2009 there is reasonable hope that the development of lipoatrophy may be a past chapter in the history of antiretroviral therapy, at least in the developed world. In resource-poor settings, the further use of stavudine necessitated by the non-availability of less toxic agents is somewhat of a scandal [76].

2.17 Intensification and simplification strategies

The 1998 IAS-USA document reflects a novel concern in the field, namely the search for treatment strategies beyond the paradigm of 2 NRTI + 1 PI or 1 NNRTI. It is noted that simplification strategies, such as induction-maintenance, are being tried (e.g., unboosted PI maintenance monotherapy) after triple therapy induction. These early simplification studies led to the conclusion that “longer term induction or more potent maintenance regimens are needed”. Here we recognize the embryo of the questions addressed in our paper II. Another strategic theme is intensification when undetectable viremia is not primarily reached, but – it is duly noted – not too late to avoid functional monotherapy.

The suggested primary strategy for switching in case of virologic failure at first seems somewhat remarkable: “Efforts should be made to change the regimen in its entirety” – that is, all three drugs. Present-day knowledge on viral drug resistance
informs us that, even in 1998, this would in very many cases have been inoptimal. In this we recognize the limited understanding of drug resistance and rational drug combination at the time.

2.18 Ritonavir boosting

In the consensus of the year 2000 [6], “ritonavir-boosted” lopinavir is recognised to be in the pipeline, offering hope of activity against resistant virus, increased potency, lower pill burden, dose frequency, cost and less stringent food restrictions. Lopinavir, which is cleared by CYP3A-mediated metabolism, is co-formulated with ritonavir, a strong inhibitor of CYP3A [77, 78].

Ritonavir-boosting is a technique similar to, e.g., the enhancement of penicillin, imipenem and ciclosporin exposure by concomitant administration of probenecid, cilastatin and ketoconazole, respectively. In the presence of ritonavir, the AUC of indinavir, saquinavir and lopinavir increase 5-80 fold [79-81].

The history of ritonavir-boosting begins with the PK study of the inhibition of CYP3A mediated saquinavir metabolism by ritonavir, and the clinical study of the co-administration of these agents – the two first PIs approved by the FDA [80, 82-84]. Soon, however, ritonavir at sub-therapeutic doses was investigated in combination with PIs such as indinavir and lopinavir, with the sole aim of increasing exposure to a single “therapeutic” PI [79, 81, 85]. Pure ritonavir-boosting had emerged out of an original concept that initially included both pharmacokinetic enhancement and putative additive/synergistic effects of dual PI therapy, as evidenced by the rationale presented in the trial reports [80, 82-84].

Ritonavir at a therapeutic dose (400-600 mg b.i.d.) is “arguably the least well tolerated PI” [86], with primarily metabolic and gastrointestinal adverse effects. At substantially better-tolerated (though by no means side-effect free) ritonavir doses of 50-200 mg/day [87, 88], (note that I exclude the ritonavir-boosted tipranavir regimen from this argument, indeed due to its low tolerability and – at least in Sweden – very infrequent use), an almost 80% inhibition of hepatic CYP3A is achieved, with even greater effects on pre-systemic activity, as evidenced by its impact on oral and i.v. midazolam clearance [89]. Interestingly, there are also data implying that ritonavir may increase the availability of PIs at their intracellular site of action [90]. This would likely displace the plasma pharmacokinetic/pharmacodynamic (PK/PD) relation. The clinical importance of this phenomenon, if any, is unknown.
Ritonavir-boosting has been crucial in improving PI treatment outcomes, preventing resistance, and making boosted PIs the cornerstone of most antiretroviral regimens in patients that have significant drug resistance. This technique is arguably one of the most important innovations in the conduct of antiretroviral therapy. Indeed, presently there are CYP3A inhibitors without intrinsic antiretroviral activity (which in some scenarios could promote the emergence of drug resistance) under development [91, 92]. Hopefully these agents will also lack other drawbacks of ritonavir, such as the abovementioned side-effects, as well as the complication that ritonavir is also an inducer of hepatic drug metabolizing enzymes [93] (see below, discussion of paper III).

2.19 When to start reassessed
The IAS-USA guidelines of 2002 [10] mark the full extent of a great change in the strategic conceptions surrounding the use of antiretroviral therapy. From the notion of “hit hard and early” [94], emerging as a consequence of the triumph of the virological paradigm of understanding of HIV infection and the availability of potent combination antiretroviral therapy, the recommendation is now, essentially, to wait as long as is considered safe. Treatment initiation in asymptomatic patients is now not directly recommended until the CD4+ T-cell counts dips below 200/uL. Why is this?

Somehow the text of the 2002 guidelines is not nearly as eloquent of the reasons for this as the fact that, to the list of drug interactions of the previous consensus document, has now been added another big table. This lists the known toxicities of antiretroviral drugs. It is considered worrisome that it is still unclear which drugs are to be blamed for the deeply disturbing abnormalities of body fat distribution seen with antiretroviral therapy. In the general spirit of the document, a new strategy, structured treatment interruptions, is suggested to afford “drug holidays”. First generation PIs, in most cases combined with a thymidine analogue and other drugs, are certainly regimens with many short- and long term adverse effects. Of note, at the time of writing this thesis, the pendulum is swinging back again towards an earlier start, due to the use of safer regimens with a higher barrier to resistance, as well as to an increasing awareness of HIV-related pathology not directly related to the overt immunodeficiency associated with low CD4+ T-cell counts [1-3].
2.20 The barrier to resistance of ritonavir-boosted PIs

On the more hopeful side, it is noted that ritonavir-boosted lopinavir is now available, though resistance data are still pending. As they emerged, it was revealed that 0/51 available genotypes from treatment-naïve patients failing on boosted lopinavir + 2 NRTI in the pivotal phase III trial (with nelfinavir, an unboosted PI, as reference treatment), showed de novo PI resistance [95]. Indeed, as already mentioned in the discussion of the varying adherence requirements for maximal efficacy, the new treatment regimens, ritonavir-boosted PI based triple therapy, or efavirenz together with two fully active NRTI, would turn out to be considerably more pharmacologically solid than were the first generation regimens.

2.21 Three active drugs, chosen on the basis of resistance testing

As a general rule, it is stated that new regimens should contain at least 2, and if possible 3 drugs deemed to be active. This recommendation feels fully modern (compare with my comments on the notions of potency and resistance expressed in the 1997 consensus); indeed, its percept has recently been corroborated by results from the developmental programs of etravirine and raltegravir [96, 97]. Also, the outcomes of strategic trials in treatment-naïve patients would soon lead to the further conclusion that four drugs add side effects but offer no relevant additional effect over three [66, 98]. Also, in contrast to the previous recommendation to, if possible, change all drugs in the regimen at failure, it is recognized that “virus that replicates during treatment failure may not be resistant to all of the drugs in the failing regimen”. This marks a further insight in the complexities of treatment failure, and, importantly, manifests the emerging reliance on drug resistance testing as a technology to decide which drugs to use, on the basis of their expected activity given the individual viral drug resistance profile, present and previous. The need to take also earlier resistance tests into account is due to the fact that previously acquired resistance mutations may not be detectable in plasma samples in the absence of continuous selective pressure (e.g., if the drugs that selected for the particular mutations are no longer used, or if treatment has been interrupted), as they are outcompeted by wild-type strains, but are liable to recur and negatively affect the efficacy of treatment when again selected for [99].
2.22 Therapeutic Drug Monitoring (TDM)

As stated in the introduction, TDM refers to the use of drug plasma concentration measurements in the clinic, to guide optimal dosing. No IAS-USA document before or after the 2002 edition spends as many words on TDM as does this expert statement. “Adjusting doses of a drug to maintain a desired plasma level is common with some drugs, and has been suggested for antiretroviral drugs”, it is stated. A number of notions on the relation of different pharmacokinetic parameters to efficacy are evoked. The apparent aim is to hint at a rationale for trough concentration (C\text{trough}) sampling (that is, measuring drug concentration at the end of a dosing interval, just prior to the new dose). The value of TDM, it is suggested, might be the greatest in cases of treatment failure, drug interactions or when salvage therapy has recently been started. It is, however, noted that it is difficult to give clear recommendations, because the therapeutic range of plasma concentrations is not known for most drugs. Thus, decisions about adjusting doses of antiretroviral drugs based on level determination “need to be individualized until more data is available”

This narrative has now reached approximately the point in time near when the first studies reported in this thesis were designed. As most of the fundamental concepts of antiretroviral therapy that are recognized today had by this time emerged, I will go on to discuss the topic of TDM of antiretrovirals.
3 AN INTRODUCTION TO HIV TDM

In order to present the problems of HIV TDM, so as to make sense of some important decisions made during the research projects here related, I shall give an account of the perception of this technology in the first half of this decade, as expressed in two position papers published in 2002: one by van Heeswijk, and one by Acosta et al, on behalf of the Adult Pharmacology Committee of the AIDS Clinical Trials Group [100, 101]. I will elaborate particularly on the relation of drug exposure to resistance, before finishing this section with some notes on the present state of HIV TDM.

3.1 The minimal effective concentration (MEC)

In the abstract of the van Heeswijk paper, it is stated that the interest in TDM is growing rapidly, and that for PIs and to a lesser extent NNRTI, relationships between plasma drug concentration and efficacy/toxicity have been identified. Furthermore, there is great variability in drug exposure in patients, which may serve to motivate routine TDM. Clinical trials evaluating the role of TDM are showing promising results (e.g., [102, 103]). On a note of caution, the reader is then reminded that there are still many questions that need to be answered before large-scale introduction of TDM can be justified. These include which PK parameter to optimize as well as what is the minimal effective concentration (MEC).

The MEC represents the lowest plasma drug concentration thought compatible with a maximal or near maximal clinical efficacy, in terms of reaching the therapeutic target in a particular group of patients (e.g. proportion of patients with undetectable viremia). It indicates the lower end of the therapeutic range. The therapeutic range, in its turn, is taken to mean an interval of drug concentrations with the MEC as a lower limit, and an upper limit defined in relation to the frequency and/or severity of some exposure-limiting toxicity.

The term Emax is used to denote the maximal effect obtainable of a drug on a certain variable in a given system. As stated above, the MEC would coincide or near-coincide with the concentration necessary to reach Emax in the relevant clinical situation. Of note, when describing the activity of an antiretroviral drug in vivo, the term potency is often used (as previously in this account). Potency in this context is somewhat ill defined, but is generally taken to represent the degree to which a drug is
capable of inhibiting viral replication, as reflected by the maximal impact on plasma HIV-RNA obtainable in short term monotherapy trials. Of note, the concept of “potency” is here used differently than in basic pharmacology, which would understand this term to represent the relation of drug concentration to, e.g., IC50 (see the following), whereas the effect on plasma HIV-RNA would rather be understood as indicating the activity or efficacy of the drug.

The MEC is certainly not unrelated to the drug concentration needed for maximal potency (as this concept is used in the field of antiretroviral therapy), but it does not necessarily coincide with it. Indeed, in principle, the MEC may be either higher or lower than the concentration yielding maximal potency. Due to the need for a sufficient barrier to resistance in order to obtain sustained virological effects, as further elaborated below, the MEC might be higher. It could also be lower, due to the presence of co-treating agents. In practice, doses higher than the lowest one yielding maximal decrease of plasma HIV-RNA in short term monotherapy trials are generally chosen for further drug development, if tolerability is not significantly compromised (e.g., [104-106]).

Taking the further step from the clinical to the microbiological world, from the MEC of TDM, via the notion of antiretroviral potency as demonstrated during short term monotherapy, and its corresponding drug exposures, to in vitro systems for determining viral drug susceptibility, the term IC50 denotes the drug concentration at which 50% of maximal viral inhibition is reached, and IC95% that were 95% is reached.

In this context it is notable that, though from a biological perspective IC95 might be thought a more appropriate parameter, IC50 values are often primarily referred to. This is due to the methodological fact that this parameter is easier to determine exactly in an experimental system (which will be clear on perusal of the idealized concentration-effect curve presented on page 23). Of further importance, it is often tacitly assumed that the relation of IC95 and IC50 would be similar for all drugs; that is, that the slope of the sigmoid curve (as that depicted below) is roughly the same. The practical validity of this idealized assumption has been challenged [107]. This problem, however, is beyond the scope of the present account.

The figure on page 23 provides a principal idea on the relation of concentration (arbitrary units) and antiretroviral effect (% of maximal), conceptualised after the Hill equation commonly used to describe ligand-receptor interactions [108]. This formal relation may be applicable to the relation of plasma drug concentrations to effect when co-treating with other agents in the clinic, or in monotherapy studies, or to the relation
of drug concentration and \textit{in vitro} viral inhibition to determine viral drug susceptibility. With increasing drug exposure increasing effects on viral replication is seen, until the curve tapers at a drug-specific maximal inhibition. In principle, for each drug the curve would have a certain slope, relating increasing effect to drug concentration, and a certain maximal efficacy in terms of the ability to inhibit viral replication.

\begin{center}
\begin{tikzpicture}
\begin{axis}[
    xlabel=drug concentration,
    ylabel=\text{Effect, \% of maximal},
    xmin=0.1, xmax=10000,
    ymin=0, ymax=100
]
\addplot[smooth, thick, black] coordinates {
    (0.1,0)
    (1,0.01)
    (10,0.1)
    (100,0.5)
    (1000,0.9)
    (10000,1)
};
\end{axis}
\end{tikzpicture}
\end{center}

\subsection{The rationale for routine TDM}

The abstract of the Acosta paper begins by noting that TDM is a strategy that has successfully been applied to many different classes of drugs. Against the background of "considerable concentration and response data", PI and NNRTI are proposed as suitable objects of TDM, since limited clinical trial data suggest that monitoring plasma concentrations of these agents may indeed clinically benefit patients with HIV infection. Then, on the cautionary side, it is noted that antiretroviral drugs are given in regimens with multiple agents, which may complicate the evaluation of drug concentration data.

Further on in the abstract, it is emphasized that procedures for sample collections, cross-validation of analytical procedures and interpretation of assay results need to be standardized. Alongside the aforementioned concerns of van Heeswijk about the PK parameter best measured and MEC values to guide evaluation, this in retrospect informs us of the extent to which TDM of antiretrovirals was still in its visionary stage.
More clinical data are needed to confirm and improve the TDM approach. Yet the position paper aims to serve as a guideline to aid clinicians who already at this time-point “chose to incorporate TDM into the routine care of their patients”.

In retrospect, it is difficult to avoid a feeling of ambiguity in this, as it were, primer of a technology that is recognized as probably not yet developed to functionality. Still, the thoughts expressed in these abstracts are rather typical of our own perceptions of TDM at the start of this thesis project. The underlying rationale for routine HIV TDM seemed rather strong, given the inter-individual variability of drug exposure, and - mainly based on first generation combination therapy – the perception that antiretrovirals were characterized by relatively narrow therapeutic intervals (of note, there was still a considerable use of nelfinavir). On top of this, there was the threat of an irrevocable loss of treatment options in case of under-dosing, and a scarcity of salvage agents. Few therapeutic areas seemed as suitable for a TDM approach as did antiretroviral therapy. In fact, it is not at all strange that many clinicians indeed did chose to incorporate TDM into the routine care of their patients.

3.3 A technology still in its infancy

In both the Van Heeswijk and the Acosta papers, outcomes of the ATHENA trial [102] - a randomised controlled trial of TDM versus no TDM, where the TDM arm saw increased response rates at 1 year of therapy - are described in some detail. I shall not summarize this oft quoted study, as it is to a large extent irrelevant for today’s concerns; this since its evidence for the efficacy of TDM in increasing response rates as well as decreasing toxicity refer to first generation unboosted PIs, such as the abovementioned nelfinavir.

The experience of ritonavir-boosting has clearly imprinted that unboosted PIs were not (and in most cases could not be) dosed to reach the minimal drug exposures necessary for maximal efficacy. This is self evident for a drug like saquinavir, but can perhaps also be inferred from the dose ranging process of atazanavir which may, arguably, be the most pharmacologically solid unboosted PI regimen ever to reach regulatory approval [109-111]. The two dose-ranging studies comparing the efficacy of doses from 200-600 mg once daily, are not uniformly suggestive that 400 mg q.d. is optimal (arguably, had Bristol-Myers Squibb fully investigated a twice daily posology for this PI, we might have had an unboosted PI as a putative first-line regimen
alternative in treatment-naive patients, albeit at the expense of the convenience and symbolism of once-daily administration).

From these deliberations, I draw the general conclusion that the efficacy of TDM in increasing virologic response in populations treated with drugs not dosed to Emax, given that the strategy serves to de facto increase drug exposure (that is, assuming retained tolerability and maintained adherence), may be a forgone conclusion. Failure to demonstrate positive effects of TDM under such circumstances would only be due to insufficient size or quality of the trial. I will return to this consideration in my discussion of paper II.

The enthusiasm of van Heeswijk is palpable when he states, apropos the outcome of the ATHENA trial, that “these results add to the arguments for the immediate implementation of routine TDM, at least for naïve patients treated with indinavir or nelfinavir”. The very next sentence, however, expresses misgivings. The application of TDM for antiretroviral drugs is “still in its infancy”, we read, and several critical issues need to be addressed.

3.4 Intra-patient variability as an obstacle to effective TDM
Both papers list a number of factors, attempting to delineate the rationale for TDM. The drug must be easy to measure in a suitable matrix, there must be a clear relationship between the exposure to the drug and its effects/toxicity, the drug must have a considerable inter-patient pharmacokinetic variability, and the effects of the drug must be difficult to assess clinically. Also, the intra-patient variability of exposure over time should not be so great as to make a single or a few measurements non-predictive of average exposure. Indeed, in one small study with intensive repeat sampling at the same time in the dosing interval in patients treated with boosted or unboosted PI, the median intra-individual coefficient of variation for repeat sampling was 43%, and in some patients considerably greater [112]. In practice, this may be a most considerable problem for the utility of TDM, demanding repeat sampling prior to any well-founded suggestions for dose adjustments.

3.5 The plasma matrix and the drug concentration at the site of action
It is noted in both papers that drug measurement in the plasma matrix makes NRTI unsuitable targets for TDM, as they are pro-drugs whose intracellular di- or triphosphorylated active forms may be present at concentrations that are not correlated
to the plasma concentrations in a simple way. This statement is informed by
experiences with zidovudine and lamivudine, where enzymatic phosphorylation is
saturated at clinical drug exposures, making the relation between plasma and
intracellular concentration highly non-linear [18, 22, 113].

The presently (at the time of my writing) emerging picture of differentially
active membrane transport pumps that may affect drug disposition to different tissues
and cell types (and thereby the entire relation between drug exposure measured in
plasma and drug response), makes this cautionary note on pro-drugs slightly ominous
not only concerning the predictive values of plasma concentrations for target exposure
to the active moiety of the NRTIs, but also for agents that are in fact not pro-drugs, but
rather substrates for transmembrane pumps, where, e.g., genetic polymorphisms may be
associated with differential activity. Also, I have previously mentioned that ritonavir-
boosting may affect the distribution of PIs.

Tissue drug distribution in humans is, however, notoriously difficult to study.
Intracellular drug concentrations of non-NRTI antiretrovirals in circulating blood cells
have been investigated, and has been reviewed by Ford et al [114]. Though intuitively
more appropriate for drugs with intracellular sites of action, this is in several ways more
complicated than measuring plasma concentrations, and investigations in the field has
remained limited, as has clinical application. Drug measurements in compartments such
as seminal fluid and CSF have also been performed [115-117] These technologies are
far from routine, and though certainly of great importance for our understanding of
pharmacology (given that the results obtained are reproducible), putative clinical use is
presently unclear.

3.6 Which PK parameter is best associated with efficacy?

The next issue dealt with by van Heeswijk is what PK parameter should be monitored.
It is stated that Cmin, the lowest concentration during the dosing interval (generally
assumed to be identical with the Ctrough but perhaps sometimes appearing an hour or
so after the intake of a scheduled dose), is likely to be the best parameter, at least for PI
and NNRTI. Clinical data from the use of amprenavir and nelfinavir are cited, showing
a greater correlation of Cmin than AUC (the area under the concentration/effect curve)
or Cmax (the maximal concentration during the dosing interval) with efficacy [118,
119]. This is supported by hollow-fibre model data on atazanavir, where time above a
threshold concentration, rather than AUC or Cmax, was best linked to the viral
response [120]. I would say that it is still generally considered that, for PI and NNRTI, Cmin may be the most important PK parameter. This is reflected, I believe, in the skepticism against once daily administration of Kaletra® in treatment-naïve patients voiced in some quarters, despite the formal efficacy demonstration of non-inferiority to bi-daily administration in the treatment-naïve [121].

For, e.g., raltegravir, however, the situation may be different [97]. It was recently reported that for this the first approved integrase inhibitor, hollow fibre studies indicate that AUC rather than Cmin is primarily linked to virologic response [122].

3.7 Drug exposure and the risk of resistance

In this context, Van Heeswijk touches on a very important issue. That is the relation between drug exposure and the risk of drug resistance in cases of therapeutic failure.

A large body of evidence from the use of PIs demonstrates that the pattern of drug exposure is crucial for the risk of resistance in case of failure [95, 109, 123]. To summarize, the emergence of major PI mutations is common when failing on non-boosted PI, in treatment naïve- as well as experienced patients. Contrariwise, the emergence of PI resistance is distinctly rare in treatment-naïve patients failing boosted PI therapy, as previously mentioned concerning lopinavir [124]. To add to the pharmacological complexity of this issue, exposure to one agent (e.g. a boosted PI) may profoundly affect the risk of resistance to the other agents in the drug regimen in the event of virologic failure [124]. This issue is of great importance, as emerging resistance in case of failure is increasingly seen as a crucial aspect when evaluating the relative merits of different antiretroviral treatment options. After all, today around 80% of treatment naïve patients in trials are generally responders at 48 weeks (ITT) to any of the recommended therapeutic alternatives, and the vast majority, indeed virtually all, treatment naïve patients without transmitted resistance that tolerate and adhere to a recommended drug regimen will reach an adequate virologic response.

3.8 The barrier to resistance revisited

As opposed to the case in 1997, when the experience of antiretroviral therapy was still relatively limited, it is nowadays mainstream to differentiate between the potency of an antiretroviral agent and its barrier to resistance. As previously described, potency is somewhat ill defined, but is often taken to be indicated by the ability of an antiretroviral agent to lower viral load in short term monotherapy studies (e.g. 7-14 days). It is
assumed to refer to the magnitude of impact on viral replication of a given antiviral when dosed in such a way that the drug exposure necessary for Emax is reached.

Barrier to resistance, on the other hand, refers to the relative likelihood of emerging resistance in case of virologic failure, and to the activity of the agent against viruses with one or more resistance mutations. The latter suggests that barrier to resistance, just like activity, will be a non-linear function of dose/exposure. Of note, the Emax for activity and for barrier to resistance may not be the same (Emax for potency perhaps being lower, since it relates to the susceptibility of the dominating population, while Emax for barrier to resistance may relate to the susceptibility of putative emerging mutants, as suggested below).

Experiences from clinical trials clearly demonstrate that the concepts of potency and barrier to resistance must be differentiated, and that this is crucial to the strategic decisions as to how to use antiretroviral drugs in combination. The ACTG 5142 study that compared efavirenz- and lopinavir based therapies in treatment-naïve patients showed that the virologic efficacy of efavirenz is perhaps greater than that of lopinavir, while the risk of resistance in case of failure is much higher when treated with efavirenz than with lopinavir [125]. This difference is also reflected in the very high frequency of emerging high grade resistance within two weeks of efavirenz monotherapy, compared to that of boosted lopinavir monotherapy [126, 127].

The principal aspect of this lesson has reappeared in the guise of the recently reported outcomes of the SWITCHMRK studies, in which patients on suppressive therapy with at least 2 NRTIs in combination with lopinavir were randomized to continue or to substitute the integrase inhibitor raltegravir for lopinavir [128]. Of note, patients that had failed prior therapies were by no means excluded, given that they had been virologically suppressed for ≥3 months prior to inclusion. Raltegravir failed to show non-inferiority at 24 weeks, and it may very well be that the lower point-estimates indicate a true lower efficacy in this population. Though, to my knowledge, baseline resistance data have not been presented, the outcome of the SWITCHMRK studies has been widely interpreted to imply that preexisting mutations (subsequent to previous virologic failure) compromising the efficacy of the co-treating agents but not raltegravir (to which the patients were naïve), may have caused patients to fail after the switch to raltegravir (indeed, the pattern of resistance present in some of the patients at failure could hardly have been incurred during the 24 weeks of the study). This would be due to the total resistance barrier of the regimen being insufficient, despite the potency of raltegravir (which in terms of antiviral effects in short-term monotherapy
might perhaps be considered more potent than lopinavir [97, 129]). Thus, from a clinical perspective, the barrier to resistance of raltegravir appears more like that of efavirenz or nevirapine than of a boosted PI. This interpretation is supported by emerging data on raltegravir resistance [128, 130, 131].

This sort of information is of great importance for the proper use of antiretroviral agents, and crucial to inform the design of strategic clinical trials on, e.g., the optimal drug combinations and regimens in antiviral therapy (as noted previously, these experiences are likely to be of relevance also to the emerging field of DAA therapy against HCV). In fact, as implied above, resistance barrier is probably a concept that is functionally best applied to the whole of the antiretroviral regimen, rather than to each component.

3.9 Drug exposure and mutant drug susceptibility

As demonstrated by the studies mentioned above [124, 125, 128], it is evident that antiretroviral agents are characterized by very different barriers to resistance. For some agents (e.g. efavirenz, nevirapine and the cytidine analogues) a single point mutation is sufficient to greatly increase the IC50 (the drug concentration yielding 50% of maximal viral inhibition in vitro) and make impossible drug delivery and/or tolerance of the doses that might be necessary for a direct antiviral effect. For others, paradigmatically the boosted PIs, the IC50 will increase gradually with an increasing number of mutations, and significant loss of response is not seen until, say, 3-5 resistance mutations have been accumulated [105, 132-134], while a full or near-full abrogation of response requires yet more (for a further discussion of the phenomenon of PI resistance, I refer to my discussion of the so called genetic inhibitory quotient, below). This difference has been suggested to be attributable, in part, to the fact that PIs act at an enzyme’s active site, whereas NNRTI act at a distance from the active site. Drug resistance mutations within an active site are likely to reduce enzymatic activity and viral fitness [51].

The theoretical understanding I until recently had of the relation of drug exposure and the risk of drug resistance was focused on the relation of drug exposure, not to the level of susceptibility of the presently dominant strain in the patient (which would be more of a determinant of potency), but rather to the (reduced) susceptibility of putative emergent strains with one or more mutations. This is what prompted me earlier to note that the activity against strains with resistance mutations and the risk of emerging
resistance on failure are likely to be expressions of the same quality. Thus, the genetic
barrier of a new compound could be reflected in the relation of Cmin (or average
exposure) and the IC50 of key single (most likely pre-existing at low frequencies) or
multiple mutants seen to emerge in vitro or in clinical trials. An important implication
of this argument is that the most adequate denominator in the inhibitory quotient (IQ)
equation to be discussed below, might not be wild-type or dominant strain
susceptibility, but the susceptibility of the most resistant commonly emerging mutant –
that is, a viral strain that as yet is not detectable, and hopefully will never be.

3.10 Resistance and pharmacokinetics

The aforementioned, simplistic approach to the problem, however, is clearly not the full
story. Evidence imply that, apart from the relation between drug exposure and the drug
susceptibility of putative emerging mutant variants, the PK profile of a drug can further
affect the barrier to resistance, at least under special circumstances such as when
discontinuing antiretroviral therapy [135]. Of note, efavirenz and nevirapine both have
long half-lives and key single resistance mutations that increase the IC50 so as to
tirely abrogate efficacy. The boosted PIs, on the other hand, have relatively short
half-lives, and the incremental increase in IC50 of each single mutation is relatively
small. It is thought that these two factors (long half-life + key single mutations
conferring full resistance) underlie the propensity for resistance to develop, e.g., after
the use of nevirapine as a single agent for the prevention of mother-to-child
transmission, or secondary to efavirenz-based treatment discontinuation [135-137]. On
the other hand, as mentioned above, virtually no emergent PI resistance at all was seen
in patients failing on 2 NRTI + boosted lopinavir in the pivotal trial showing the
advantage of a boosted PI regimen over an unboosted in treatment-naïve patients [95].
This finding has since been repeatedly corroborated with boosted PI [124].

3.11 Viral fitness and the barrier to resistance

Recently I was alerted to a paper that has furthered my views on resistance beyond that
which I have accounted for above [138]. The authors interpret the concept of “barrier to
resistance” in terms of the relative fitness of mutant and wild-type as a function of
increasing drug pressure. The argument is that resistance would only be selected within
a window of drug concentrations where, at the lower limiting level, drug pressure is
sufficient for the resistant variant to appear more fit than the wild-type, and at the upper
limiting level, total drug pressure is large enough to sufficiently reduce replication in order to prevent the emergence of resistance. This relation is evidenced by the findings of Bangsberg et al [123]. In this paper it is shown that the NNRTI efavirenz and nevirapine, which have (a) long half-lives and (b) whose key mutations do not lower viral fitness, are prone to show an increased rate of drug resistance upon failure with decreasing reported adherence. On the other hand, with the unboosted PI, that have (a) short half lives, and (b) where primary mutations lower viral fitness, the relation is the opposite, the likelihood of resistance in case of failure being greater the higher the reported adherence rate.

So this difference in the relation of adherence rate and drug pressure comes about since with the NNRTI, when adherence is too low to achieve virologic suppression, long periods of time will be spent with non-inhibitory drug concentrations within a drug concentration interval where mutants are more fit than wild type, whereas with unboosted PI at low adherence, drug concentration rapidly falls towards zero, and a relatively high adherence (and subsequent drug concentration) is necessary for drug pressure to be sufficient for mutants to outcompete wild-type.

Finally, though not studied in the aforementioned paper, by inference, the great barrier to resistance of boosted PIs, particularly in treatment-naïve patients, would arise since the time spent with drug concentration high enough to select for mutants over wild-type, but not sufficiently high to achieve virologic suppression, is minimal, regardless of adherence rate. Under such circumstances it would be expected that virologic failure would generally not be accompanied by acquired resistance, as indeed is the case with the use of boosted PI in the treatment-naïve.

3.12 The MEC and in vitro susceptibility data

Following this digression (antiretroviral PK/PD simply cannot be summarized without a discussion of drug resistance), I shall now return to the concept papers.

Van Heeswijk states one of the crucial questions in TDM (regardless of whether it is performed routinely, or due to adverse effect or unpredictable pharmacokinetics), namely, what is the MEC? It is acknowledged that “there is as yet no consensus”. The question arises: could the MEC be inferred from in vitro susceptibility data?

“Extrapolation of in vitro inhibitory concentrations to MECs in vivo is complex because many factors that are important for the efficacy of a drug in vivo, are difficult to mimic in vitro (e.g., protein binding, penetration of sanctuary sites, formation of
active metabolites, development of resistance), states the paper by van Heeswijk. “The combined use of at least three different antiretroviral drugs simultaneously complicates the definition of threshold concentrations because of anticipated additive or synergistic effects between drugs in the regimen”.

The approach of comparing in vitro sensitivity data and actual in vivo concentrations has a great intuitive attractiveness. The IC50/95 (for definition, see above) is of great importance for dose selection, as this parameter, or a certain multiple of it, may be chosen as the initial exposure target. However, the relation between in vitro sensibility and drug concentration may subsequently undergo re-interpretation during phase Ib/II studies, based on the generated in vivo data, in terms of what multiples of the IC50/95 (usually adjusted to account for protein binding) might be considered necessary for optimal efficacy.

The notion of drug concentration targets as multiples of, say, the IC50, is nothing but that of an inhibitory quotient (IQ), which will be further discussed below. First, however, I shall note the position of the Acosta paper on this issue. It differs in tone from that of van Heeswijk, and by being more elaborate further discloses what may be fundamental weaknesses of this approach.

In their call for the standardization of antiretroviral drug level interpretation, Acosta et al seem ready to substantially rely on in vitro findings to inform about the MEC. The problem of adjusting for protein binding is recognized. “Phenotypic tests (my addition: that is to determine IC50 and IC95), currently do not take into consideration the extent of drug protein binding (...) A simple (but possibly not the most accurate) way to account for protein binding is to divide the IC50 by the free fraction of a drug.”

It is noted that this conversion method is not exact, but that it provides results close to the range expected on the basis of in vitro protein-binding experiments. Actually, several PIs are highly protein bound (>90%), and thus any inaccuracies in this determination may have a great impact when comparing total plasma drug concentrations to IC50 values. It is also worth mentioning that IC50 values are not determined in the presence of 100% human plasma, as most tissue culture cells do not survive in higher serum concentrations, and in vitro viral replication is inhibited by the presence of high levels of serum [139, 140]. Furthermore, estimated IC50 values may vary substantially depending on assay, cell type and virus strain.

The Acosta paper contains a formula for determining the EC90 (the in vivo concentration compatible with 90% of maximal effect), which, insofar as I understand,
is assumed to represent the plateau of the concentration-effect curve (that is, by implication, the MEC). This is taken to approximately equal the IC50 times 3, adjusted for protein binding.

There follows a review of the literature on reported wild-type IC50, IC95, α1-acid glycoprotein corrected IC50, and reported degree of protein binding for these agents. It is notable that the calculated protein binding adjusted IC50 of these authors range, e.g., from 150-2700 ng/mL for saquinavir, 150-5500 ng/mL for ritonavir and from 200-5500 for nelfinavir. This approach does not appear to generate very precise suggestions on the level of the MEC, with 18-37 fold differences between the lower and upper ranges of the quoted estimates.

In summary, IC50 values certainly relate to the MEC, but are not expected to coincide with it, nor define it; and clearly this relation is yet more complicated for agents such as raltegravir, where AUC rather than Cmin is likely to be the most important PK parameter.

3.13 The inhibitory quotient (IQ)

I shall not further discuss the merits and problems, or the technicalities of IC50 determinations with or without serum proteins (of which I have no great knowledge). I shall, however, relate that the aforementioned concept of the inhibitory quotient (IQ) has been coined to refer to the ratio of the Ctrough and some measurement of the viral susceptibility to the drug, such as (protein-binding adjusted) IC50/95. Apparently this was first elaborated by Ellner and Neu [141] as way of describing the relation between antibiotic drug exposure and MIC. It relates drug exposure to multiples (or theoretically, fractions) of some inhibitory concentration or other, thereby recognising, as mentioned above, that the MEC is more likely to be correlated to some multiple of, say, the IC50/95 rather than to its crude estimate. The concept of IQ as well as its derivatives, such as “virtual inhibitory quotient” (where the IC50/95 is inferred on the basis of the fold-change in susceptibility expected given a certain resistance genotype), and “normalized inhibitory quotient” (where the measured patient IQ is expressed as fractions or multiples of a reference IQ, derived from efficacy data in a reference population), has been reviewed by Morse et al [142]. In Sweden we never used a virtual or real phenotypic IQ approach to TDM, and actual phenotyping for drug resistance was seldom, if at all performed (discounting tropism tests for maraviroc).
3.14 The genetic inhibitory quotient (GIQ)

The GIQ refers to the ratio of a pharmacokinetic parameter (normally Ctrough) and the number of drug resistance mutations (listed in a defined panel such as, e.g., IAS-USA) present in the virus. The practical rationale behind using a genotypic denominator in the exposure/susceptibility equation is that resistance genotype rather than phenotype is generally available from clinical isolates in patient care, as stated above. The GIQ was, just like the abovementioned variants of the phenotypic IQ, devised in response to the challenge of interpreting PI drug concentrations in patients with different degrees of PI resistance, in recognition of the fact that the MEC is likely to be pressed upwards with the accumulation of PI resistance mutations.

The idea underlying the GIQ concept is thus an assumption that the MEC would be linearly increased by the multiple of the number of mutations - an assumption as contrafactual as any. For instance, it seems somewhat less than self-evident that the MEC for patients with two rather than one PI mutation would be doubled, as implied by the GIQ concept.

My full view on the GIQ is well described in the following letter, intended for the editor of AIDS (but never sent due to my indolence), that I devised a few years ago with the help of professor Anders Sönnerborg:

“Kempf et al described a log-linear relation between the total number of accumulated HIV protease mutations at 11 defined amino acid positions, and the increase in lopinavir in vitro IC50 of a given viral isolate [132]. This formed the basis for the Kempf panel of HIV protease mutations, which were then prospectively validated, insofar as an increasing number of Kempf panel mutations were associated with a progressively lower likelihood of successful outcome in lopinavir-based antiretroviral treatment [143]. The predictive value of the Kempf panel of mutations was subsequently corroborated by the findings of Masquelier et al [134]. Thus, the Kempf panel came to form the basis for the concept of a GIQ, which is utilized in order to define the optimal plasma drug exposure in a treatment cohort in terms of the ratio of the trough plasma concentration and the number of HIV protease resistance mutations.

Ideally, the empirically determined optimal GIQ should then be applied to calculate the optimal target trough concentration of an individual patient as the product of the GIQ and the number of HIV protease mutations in the virus of the individual patient. Several attempts have been made to determine the optimal GIQ threshold for
lopinavir [144-146] as well as for other protease inhibitors, on the basis of the Kempf panel and of other such sets of resistance mutations. The wide difference in optimal GIQs described in the respective publications is striking, despite the utilization of the same mutation panel. Quite apart from the obvious fact that the GIQ erroneously gives equal weight to each mutation in the panel used as its basis, there appears to be yet another dubious assumption bolstering the concept of the GIQ, which may be an important reason for this divergence in findings concerning optimal GIQs in different cohorts. Kempf et al described a log-linear relation between the number of HIV protease mutations and viral drug susceptibility, whereas the GIQ assumes a linear relation between the number of mutations and the optimal drug exposure according to the following equation: Target drug concentration = GIQ * number of PI resistance mutations.

This means that the GIQ presents a linear approximation, within a certain interval, of a log-linear or exponential relation. As a consequence, whatever GIQ will be found as optimal or most predictive of a good treatment outcome in a given patient cohort will be a function of the average number of HIV protease mutations in the cohort. In three different studies attempting to define the lopinavir GIQ, the number of baseline Kempf panel mutations in the respective populations is incompletely described, but the presented data imply that they are not equivalent [144-146].

In conclusion, the GIQ concept as used in the aforementioned studies is questionable. More adequate models, taking into account the exponential relation between the number of HIV protease mutations and the necessary drug exposure, or preferably, taking into account the differential impact of different mutations, should be used for investigating the important relationship between viral resistance genotype and the drug exposure necessary for a satisfactory treatment outcome.”

3.15 Tolerance and the problem of the “toxicity” limit

Apart from the problem of defining the MEC, “toxicity” limits might also be moving targets, as instanced by the study by Haas et al [147] on the PK/PD consequences of the CYP2B6 G516T mutation (see below), in terms of the risk and severity of efavirenz related CNS side effects. This paper shows that individuals homozygous for the TT genotype had higher efavirenz exposure at all measuring points of the study, up to 24 weeks. At week 1, TT homozygotes had significantly more intense CNS side effects, but not at any time point thereafter. It is interesting to note that, while one might think
that the disappearance of this initial, concentration dependent effect despite continuously elevated drug exposure could be due to selection bias (with a higher dropout rate in individuals with higher drug concentrations), in fact, the dropout rate was higher among wild-type homozygotes. 14/14 patients with the low clearance TT genotype continued through 24 weeks, as opposed to 69/83 of wild-type GG. This paper indicates that the CNS side effects of these agent (today often cited as an indication for drug concentration monitoring) shows a classic pattern of pharmacological tolerance otherwise characteristic of CNS-acting drugs such as opiates and benzodiazepines. Though the frequency of treatment-limiting neurological side effects are no doubt higher in “real life”, it is worthwhile noting that the discontinuation rate due to CNS toxicity over 48 weeks in the efavirenz arm of the recent STARTMRK study was reported to 1/281, demonstrating what little impact this may have on overall efavirenz tolerability in a selected population [131].

3.16 The scope of TDM

Now to the conclusions of these position statements. The Acosta paper finishes with some practically minded bullet points regarding the timing of sampling and how to intervene, which would confirm with general TDM practice today. The conclusion of Van Heeswijk is more conceptual. He notes that “to date, TDM of the PIs and NNRTIs may be of value in selected situations, such as patients with liver and kidney impairment, pregnant women, patients at risk for drug-drug interactions, and children.” Furthermore he states that “the aim of future studies should be to provide a solid basis for TDM of antiretroviral drugs, either alone or in combination with viral resistance monitoring, by performing prospective, randomized clinical trials.”

Given the wide therapeutic intervals of present day antiretrovirals, it has by now become very clear (and, I believe, generally recognised in the field) that TDM will not be needed for the vast majority of patients. A Cochrane systematic review of the randomized controlled trials of HIV TDM implied what has been stated in relation to the ATHENA trial above [102], that only when treating with unboosted PI is there evidence in support of its routine use [148]. Remains selected patients with unpredictable PK (that is, special populations were PK may not have been formally investigated pre/post marketing), and perhaps patients with certain adverse effects.

Concerning these special populations, we have been somewhat troubled by the issue of how to dose lopinavir in late pregnancy, where apparent clearance is often
substantially increased, with consequent relatively low exposures [149]. Though the PK effect is indubitable, it remains questionable whether a third trimester dose increase is generally mandated in the absence of viral drug resistance. TDM is often performed (in Sweden), but the evaluation of results is highly tentative, as the oft-quoted MEC of 1000 ng/mL in treatment-naïve has virtually no evidence base (see paper IV).

In practice, we have often considered TDM of use in patients with atazanavir-induced symptomatic hyperbilirubinemia (which does not attenuate during therapy) or efavirenz-related CNS side effects though the utility of this has not been formally proven (the latter issue is further discussed below in the section on paper IV). To end this introduction to the principles of HIV TDM on a personal note, one study considered for the project here reported, but not yet performed, was a retrospective assessment in terms of outcome (tolerability and efficacy) in patients on atazanavir or efavirenz for whom doses were lowered (in routine care) due to symptomatic hyperbilirubinemia or CNS side effects, when TDM data suggested that this would be feasible without a loss of virologic efficacy (a study similar in kind to that recently reported by van Luin et al [150] and further discussed in the context of paper IV). Indeed, things being as they are, we are in the somewhat unsatisfactory state of not knowing to what extent, if at all, our TDM practice over these years has been of any value in optimizing outcomes of antiretroviral therapy in Sweden.

As an end to this general introduction to TDM, I would like to note a statement from the paper by van Heeswijk: “Premature implementation of TDM may hamper the conduct of proper, randomised prospective trials to confirm its anticipated clinical benefit.” Looking at my own papers II and IV, these statements about the problems for the scientific evaluation that might follow a premature introduction of (more or less) routine TDM seem prophetic.
4 ON THE EVIDENCE BASE OF THE THERAPEUTIC RANGE

I shall now describe in further detail how our evolving thoughts on the so-called therapeutic range came to inform the conduct of the research reported in paper II. Indeed, I consider this small pilot study of boosted atazanavir monotherapy pivotal to the general account of this thesis, since the experiences there reported were formative of an important turn in our views on TDM. That is not to imply that this paper itself forms evidence for these views, but rather that these experiences served to greatly alert us to the weight of evidence emerging in the field in general. After all, if one recognizes a problem, one is considerably more likely to listen carefully to others that also do so, than if one experiences that everything is working very well.

As a starting point, I will review two papers of major importance for our TDM-based advice at the time of the inception of the research projects under consideration. These are by Marzolini et al 2001, and Ståhle et al 2004 [151, 152]. Both papers aim to define the therapeutic range of efavirenz, which is today in most cases part of the reference regimen when investigating new agents in treatment-naïve patients. The Marzolini paper is still often referred to as an important part of the evidence base of the oft quoted therapeutic range for efavirenz (see paper IV), and may be one of the most frequently cited references in the field.

4.1 The Lausanne cohort

Marzolini et al reports a cohort of 130 patients, of whom 85 contributed more than one drug concentration sample. These patients were recruited at the Lausanne University Hospital in the period 1999-2000. They must have been on efavirenz therapy for at least 3 months in order to participate. In fact, sampling for efavirenz drug concentrations was undertaken after between 3-18 months of therapy, with an average of 8 months. It is thus notable that the study consisted of an enriched set of patients that had tolerated efavirenz for at least twelve weeks (CNS side effects are most severe during the first month or so [147]) and that, presumably, in most cases had shown a satisfying early virologic response. It is by no means clear what proportion of the total number of patients starting efavirenz during this period in Lausanne were eligible for...
inclusion, and whatever consequences the enriched population might have for our understanding of the concentration-effect/toxicity relation of efavirenz.

Patients were treated with combination antiretroviral therapy, including, apart from efavirenz, either 2 NRTI or a PI +/- NRTIs. This is not further defined, though, apart from that 40 patients were treated with PIs, and that the most common agents used in combination were zidovudine, lamivudine and nelfinavir.

As further elaborated below and in paper IV, the dose-ranging trials leads me to believe that the MEC for efavirenz will vary depending on the support given by the other drugs (notwithstanding the fact that we would today be reluctant to use efavirenz, like raltegravir (see above), at all without the support of two fully active NRTIs or a boosted PI). However, there is not one word in the paper by Marzolini et al dedicated to the treatment history of the patients included. I would not dare to assume that they were all treatment naïve, given the time when this study was conducted. If not, one might suspect that the efficacy of the drugs used for co-treatment was very variable, due to varying degrees of pre-treatment with, e.g., mono- or dual NRTI therapy, or previous non-suppressive triple therapy.

Efficacy was assessed by measurement of plasma HIV RNA. In accordance with standards at the time of the study, patients with a viral load > 400 copies/mL were considered to experience virologic failure. It is interesting that nothing at all is stated about the time at which the primary outcomes were to be measured, or the relation of this time to that of initiation of therapy or of drug concentration measurement; nor to what extent this was pre-specified in the protocol.

Neither is it stated whether TDM data were regularly communicated with the clinicians, and if so, whether they could have prompted any changes in dosage (though this point may be somewhat irrelevant if exposure and outcome were in fact measured simultaneously). CNS side effects were assessed by a “standardized evaluation”, not further defined, which was repeated at a 3-month interval in an undefined subset of patients. It is interesting to note that there is no mention of adherence assessment in the study, though the issue of adherence, and the possibility of monitoring adherence through TDM, is briefly mentioned in the discussion section.

In this cohort, sampling took place between 8-20 hours post dose (average 14 hours). The sampling schedule is considered unproblematic and there is no trough imputation; in fact the authors claim that sampling time explained only 3% of the total variance. It would have been more interesting, though, to have the variability due to timing of the sample related to the individual patient rather than to the population
40

(which is expected to be highly variable due to genetic polymorphisms in CYP2B6). The intra-patient variability is reported to be 30% on repeat measures, which is lower than has been reported with PIs [112]. In the end, the appropriateness of the approach to data handling is motivated by the long half-life of efavirenz.

4.2 Defining the therapeutic range of efavirenz

The predictive value of efavirenz concentrations for viral suppression or CNS effect was evaluated by logistic regression. It is not stated that this model included any other independent variables but efavirenz concentration. The outcome of the PK/PD analysis was that 50% (5/10) patients with concentrations below 1000 ng/mL failed, compared to 22% (23/103) with 1000-4000 ng/mL and 18% (3/17) with concentration above 4000 ng/mL. It is not mentioned what indeed were the drug levels of these 10 patients with measurements below 1000 ng/mL, nor, as stated previously, how they related to adherence. CNS-toxicity was deemed (unclear by what criteria) to have occurred in 4/17 patients with concentrations >4000 ng/mL, but only in 9 out of the remaining 113 patients. The authors therefore proposed the interval 1000-4000 ng/mL as therapeutic, and indeed this was, at least for some time, accepted by the greater part of the HIV TDM community [153]. Of note, some sort of support for the lower efficacy margin is found in the PK/PD sub-study of the 2NN trial [154], though this publication indicates that the operative characteristics of efavirenz drug concentrations as an predictor of outcome treatment-naïve patients is likely to be very poor.

4.3 The Stockholm cohort

In 2004, however, this interpretation of the results from the Marzolini study did not inform the recommendations of our TDM service in Stockholm. This was perhaps not mainly because of any argument against the methodological approach of Marzolini, nor because the external validity, in terms of “to whom these truths pertain”, was differently interpreted, as might have been conceived given the different inclusion criteria and outcome measures of these studies. Rather, I believe it was because data from the Stockholm cohort indicated that a somewhat higher MEC might be more discriminative, and perhaps because, at this time in the history of antiretroviral therapy, there was a certain feeling that increasing drug exposure might never be considered entirely the wrong thing to do (that this attitude was in concordance with generally held
assumptions about antiretroviral therapy should be clear from the previous summary of the US guidelines evolution).

The Stockholm efavirenz dataset was published by Ståhle et al [152]. This paper reports outcomes in 58 patients that were prospectively recruited at the HIV policlinic at the Karolinska University Hospital (formerly Huddinge Hospital). The publication does not fully state which antiretroviral drugs the patients were treated with apart from efavirenz. Similar to the paper by Marzolini et al, nothing is said about the treatment history of the patients. In fact, the Stockholm cohort was highly pre-treated, and in many patients the resistance pattern demonstrates that efavirenz is unlikely to have had much support from the co-treating agents.

Thus, both the Lausanne and the Stockholm cohorts were reported in peer-reviewed journals without any information indicating the viral susceptibility to the drugs used in combination. This forms the basis of my tentative inference that the complexities of the definition of the MEC were not fully recognized at this time by many workers in the field.

4.4 Challenging the Marzolini conclusions

The drug concentrations reported by Ståhle et al were “normally taken any time between 10-24 hours post dose”. Similar to the Marzolini paper reported above, no Ctrough imputation is made; in fact, Marzolini et al are quoted as support that this is not necessary. There was an attempt to retrospectively assess CNS adverse effects through patient charts, but this was not successful.

Then the PK/PD results are reported in a way that is unfortunately all-too frequent in the TDM literature; indeed we are guilty of it ourselves in paper II. Geometric mean concentrations are shown for two groups, responders and non-responders – that is, drug concentrations are, as it were, described as a function of response/non-response, rather than response as a function of drug concentrations. I admit that it took me a while to come to terms with how non-informative this in fact is, as regards the assessment of the relation between concentration and effect, and the putative clinical predictive values of drug concentrations. Also, this mode of data presentation can imply a concentration-effect relation within the range of exposures seen at a given dose when in fact there is none, e.g., if there are a number of very non-adherent patients among non-responders, serving to lower the average concentration in this group. A more informative way to handle this is to use logistic regression or some
other curve-fitting approach (e.g., a sigmoid Emax model after Hill, as discussed above) to describe outcome as a function of drug concentration, or simply to view drug concentration data as any diagnostic test by constructing a so called receiver operating characteristics (ROC) curve, and then explore the data for the most predictive concentration cut-offs in terms of the proportion of responders or non-responders (the statistical significance of which can then be analysed with, e.g., a chi-square test). Indeed, in the paper by Ståhle et al follows a data-fitting approach, on the basis of which it is reported that the cut-off value with the best operating characteristics is 6.9 µmol/L (approximately 2200 ng/mL).

What can be understood from this dataset, presented in a somewhat unusual way, is that 30/33 (91%) of patients above 6.9 µmol/L were in fact primary responders (reached undetectable viremia), as against only 15/25 (60%) among those below. That having been said, it is argued that increasing the lower target concentration from approximately 3 µmol/L to 7 µmol/L (approximately 1000 to 2200 ng/mL), the response rate might be increased from 70 to 80%, with reference to the paper by Marzolini et al. These deliberations informed our evaluation of TDM samples, and our recommendation that mid-dose interval efavirenz concentrations should exceed 7 µmol/L (which corresponds to Ctrough values between 4.6-5.9 µmol/L, using linear extrapolation, assuming (relatively short) efavirenz half-lives between 20-50 hours [155]).

4.5 The dose ranging studies of efavirenz

It was my friend and colleague, doctor Olle Karlström, who first made it clear to me what now seems obvious - that this efavirenz MEC was too high. Olle had noted the discrepancy between our suggestion of > 7 µmol/L as a target, and the fact that the average Cmin at 600 mg q.d, as reported in the Swedish Physician’s Desk Reference, was 5.6 µmol/L (approximately 1800 ng/mL) [156]. Did we really mean that a very considerable fraction of patients treated with efavirenz according to labeling were actually underdosed, despite the fact that virtually all treatment-naïve patients without transmitted resistance that adhere and tolerate a regimen of this drug + 2 NRTI would have durable virologic suppression?

This prompted a review of the dose-ranging process for efavirenz, which was performed in two highly distinct sets of patients: treatment naïve patients and patients that were failing on mono- or dual NRTI therapy [157, 158]. The PK/PD relation seem
to be different in naïve patients and in patients where efavirenz was added to failing monotherapy, despite the fact that the virus was assumed to be fully sensitive to efavirenz. Obviously, the Emax of the regimens in these different situations is not the same, but neither was, seemingly, the MEC for efavirenz. In fact, in treatment-naïve patients the dose ranging process failed to indicate any efficacy advantage of 600 mg q.d. over 200 mg q.d. [157]. Only when efavirenz was added to failing dual NRTI therapy (which is contrary to present day recommendations, as it results in the rapid development of resistance) was 600 mg more efficacious than 400 mg [158].

So (a) the average Cmin at 600 mg was lower than 7 µmol/L (in other datasets median efavirenz exposures have been somewhat higher, but not so as to impact our general argument or conclusions), and (b) there was no indication of an advantage of 600 mg over 200 mg in the dose ranging trial in treatment-naïve patients (though admittedly there might, of course, still be an efficacy difference if you study a sufficiently large sample). The selection of 600 mg q.d. as the recommended dose for efavirenz was informed by the notion of a maximal tolerable dose, and not by any particular efficacy advantage over the lower doses tested [104]. Certainly, due to polymorphisms in CYP2B6 (see below), there would be an overlap in exposure between doses, but all in all, PK data, the outcome of the dose ranging process and the virologic suppression rates in adherent, treatment naïve patients strongly implied that mid-dose interval concentrations of 7 µmol/L would be an inflated MEC value in treatment naïve patients treated with efavirenz and two fully active NRTI, prompting all-too frequent suggestions of dose increases that retrospectively seem wholly unnecessary.

4.6 The demands of atazanavir TDM

In 2004, a new PI was approved and introduced. This was soon to reach an important position in the therapeutic armamentarium. I am, of course, referring to atazanavir. There was a perceived need of atazanavir TDM in the prescribing community, and we were much concerned with building a necessary knowledge base. We were at this time becoming aware of the difficult problems when studying and interpreting the relation of exposure and response in antiretroviral therapy. Thus we did not with certainty foresee that a solidly motivated, generally accepted MEC value for atazanavir would emerge. On the other hand, it seemed like the upper level would never have to be defined. Each patient has one, and that is visible jaundice, as atazanavir inhibits UGT1A1 and thereby
causes unconjugated hyperbilirubinemia, the magnitude of which is affected by UGT1A1 genotype (see below).

In this situation we attempted to address TDM from a somewhat different angle (different at least from the general approach to TDM in Stockholm at the time); that is, via the dose-ranging procedure, which, as just mentioned, had seemed so informative concerning the problems of the efavirenz MEC. The basic idea is simple enough. If the overlap in exposures between dose levels in dose-ranging trials is not too great, doses could be linked to the exposures they generated, and these latter then compared with the outcomes in the respective dosing arms. Thus a rough estimate of the MEC might be inferred without the resort to the naturalistic cohorts that we no longer trusted would generate any useful data.

I earlier expressed our views on the genetic inhibitory quotient (GIQ). Yet, though we dismissed this approach, we clearly somehow had to take into account that patients treated with atazanavir were likely to have varying degrees of PI resistance, some being previously naïve to therapy and/or having fully susceptible virus, whereas others would have accumulated varying numbers of major and/or minor PI mutations. As it were, we never attempted to recommend any particular MEC for atazanavir in treatment-naïve patients in our TDM practice, but only for experienced patients. This was done by the indirect approach of comparing outcomes of the BMS–044 and –045 studies relating them to the range of exposures generated by unboosted and boosted atazanavir [159, 160]. Neither the BMS-089 trial nor CASTLE were available at the time [109, 161]. This approach to the therapeutic range will be further elaborated in the discussion of paper II.

Also, we fully recognized that there was no great differentiation in our recommendation, given the likelihood of a substantial “real” MEC variability within the treatment-experienced population; we mainly wanted to avoid exposures that might be too low (this genuine uncertainty about the appropriateness of recommended doses in fact is a strong feature in our general approach to antiretroviral therapy at the time, which I believe is reflected in the willingness to utilize TDM data, despite a very shaky evidence base, to motivate increased doses, even within ongoing trials, as witnessed by papers II and IV).
4.7 TDM and the newer antiretroviral agents

It is notable that with the newer agents approved for HIV therapy, such as darunavir and raltegravir, the dose ranging procedure and the general investigations at the time of approval have been more systematic than with some of the older agents. Indeed, the clinical pharmacological studies performed prior to marketing, not only in the field of antiretrovirals, but also for DAAs against HCV, are considerably more extensive than was the case, say, 10 years ago. For many of these drugs, more informative concentration-effect datasets than those generated pre-marketing may not emerge post marketing. Data for darunavir and etravirine from the POWER and DUET studies imply some sort of concentration-effect relation over the whole cohorts studied, but given the variability in resistance pattern and activity of concomitant drugs in patients treated with these agents, it does not seem probable that any operative MECs be defined [96, 105]. Furthermore, darunavir is not only highly protein-bound, but this binding is also saturable, making the predictive value of total plasma concentrations for the active moiety dubitable [105]. As for raltegravir, we have a drug with very low short-medium term toxicity and a dose that, as can be inferred from the dose-ranging procedure, generates exposures far exceeding the MEC [106]. In fact, as mentioned above, the raltegravir experience seriously challenges another of the common assumptions of HIV TDM, that of the trough concentration (or its relation to IC50/95) being the most predictive PK determinant.
5 PHARMACOGENETICS AND DRUG-DRUG INTERACTIONS

5.1 The pharmacogenetics of antiretroviral therapy

The study of genes that determine antiretroviral pharmacokinetics and toxicity have added very valuable information to our understanding and management of HIV therapy. The field of HIV pharmacogenetics is by now rather large, though many of the results obtained must still be considered preliminary. I will not attempt to fully describe this area of investigation, as TDM and concentration-effect relations are the main focus of this account (and indeed I shall not pretend to be an expert on pharmacogenetics). Still there are some findings of HIV pharmacogenetics that are of general importance, and I will attempt a brief description of the field. For those interested in a more extensive general outline of this field of research, I refer to, e.g., the University of Liverpool pharmacogenomics website [162].

The quite elegant research program leading to the identification and confirmation of HLA-B*5701 as a marker for abacavir hypersensitivity has been noted far beyond the HIV treatment community, and has set a paradigm for the investigation of type 2, or “idiosyncratic”, drug reactions in general [163-167]. On a note of reservation, though, most such drug reaction have several-fold lower incidence that does abacavir hypersensitivity, which may, in many cases, limit the predictive value of pharmacogenetic tests.

The most important findings of the pharmacogenetics of drug metabolizing enzymes, as related to antiretroviral therapy, is the elucidation of the relation of the variability of efavirenz pharmacokinetics, and genotype at CYP2B6 [147, 168-170]. The most prevalent single nucleotide polymorphism of importance is the G516T mutation, which together with A785G constitutes the *6 allele [171]. The frequency of this has been reported to 26% in Caucasians, 32% in African-Americans and 47% in Ghanaians [172]. At the recommended dose, efavirenz AUC is approximately threefold greater in patients with homozygous for the TT genotype, compared to those homozygous for CC [147, 168]. Polymorphisms in CYP2B6 also have implications on the efavirenz dose requirement when co-treating with rifampicin [173, 174], as well as on the risk for drug resistance when discontinuing an efavirenz-based regimens, due to the prolongation of an already long half-life (see further below) [135, 155].
On the other hand, the known CYP3A polymorphisms (the topic of the first paper of this thesis), are not considered likely by many investigators, to be of practical relevance in this era when antiretroviral drugs are generally much more tolerable than was the first generation, and where the crucial agents are dosed so that exposures in the vast majority of adherent patients are far above the MEC. Given the wide therapeutic ranges of many present-day antiretroviral drug regimens, the kinetic impact of a given pharmacogenetic variant in a drug metabolizing enzyme would have to be large in order to seriously impact tolerability or virologic efficacy. Indeed, despite the fact that the aforementioned CYP2B6 polymorphisms such as the highly prevalent G516T (estimated frequency of homozygous TT individuals being, e.g., 3% in Caucasians and 20% in African-Americans [147, 168]) greatly affects efavirenz PK, as mentioned above, several studies have failed to demonstrate its general impact on efavirenz treatment outcomes [175, 176].

The literature on MDR-1 polymorphisms is contradictory and not easy to briefly summarize [177]. Suffice to say that their importance for the outcome of antiretroviral therapy has not convincingly been demonstrated. The role of other genetic polymorphisms in drug transporters presently appears unclear. Data presented on the role of polymorphisms in ABCC2, the gene coding for MRP2, on the risk of renal adverse effects of tenofovir [178, 179] are somewhat confusing, as in vitro studies imply a role for MRP4 but not MRP2 in the renal handling of tenofovir [180, 181].

In the light of what is presently known about the PK/PD and side effects profile of raltegravir, the genetic polymorphism at UGT1A1 (where the common “Gilbert” genotype *28*28 yields a 40% higher exposure compared with wild-type) is highly unlikely to have any clinical consequences [182]. The impact of UGT1A1 genotype on the extent of atazanavir-induced unconjugated hyperbilirubinemia (see below, discussion of paper II) is well described [183-185]. Yet, the overall clinical impact is not overwhelming, as the discontinuation rates due to hyperbilirbinemia have been reported at 0-3% in major clinical trials with ritonavir-boosted atazanavir [109, 155, 160, 161], despite the high prevalence of mutations resulting in decreased UGT1A1 activity [186].

5.2 Antiretroviral therapy and drug-drug interactions

The final issue I want to mention in this introduction is that of drug-drug interactions in HIV therapy. Most presently recommended regimens either include a ritonavir-boosted
PI or an NNRTI. As discussed more extensively above, ritonavir is an HIV protease inhibitor of the first generation which is now only used in sub-therapeutic doses to strongly inhibit CYP3A (and P-glycoprotein) and thereby increase exposure to another drug (so far other PIs, but in phase II trials also an integrase inhibitor and, interestingly, at least one investigational HCV protease inhibitor is ritonavir-boosted [187, 188]).

To somewhat complicate things, underlying this inhibition, ritonavir (and in several cases also the boosted PI itself) is a substantial inducer of drug metabolizing enzymes, leading to a very complex spectrum of clinically relevant and putatively relevant drug interactions, often increasing exposure to the co-administered agent, but in some cases lowering it. The list of drugs that interact with ritonavir is very long and encompasses important agents for co-treatment in the relevant patient group, such as, e.g., methadone, rifampicin, rifabutin, voriconazole and simvastatin [189], as well as, most probably, many or all of the emerging HCV protease inhibitors.

Also, all presently available NNRTI are inducers of hepatic drug metabolism. This last prevalent fact underlies our recognition that the Scandinavian NORTHIV clinical trial (see below) might be highly informative as a setting for the evaluation of the operating characteristics of 4β-hydroxycholesterol, a purported endogenous marker of CYP3A activity, as one would expect differential effect on this biomarker in the different treatment arms of NORTHIV. Indeed, the underlying mechanisms of drug-drug interactions with ritonavir is further discussed in my paper III.

For those interested in a more extensive general overview of the area of antiretroviral drug interactions, I refer to the respective Summaries of Product Characteristics (SPCs), available at the EMEA home page [190], and particularly to that of Norvir® (ritonavir) [189]. Also, there are several online databases devoted to this topic (e.g.[191]).
6 THE NORTHIV TRIAL

About half of the work reported here has been performed under the NORTHIV umbrella, as can be inferred by perusing the scientific papers included in this volume. NORTHIV is a randomized, open label, multi-center efficacy trial carried out in treatment-naïve patients in Sweden and Norway. The primary outcomes of the NORTHIV trial have yet to be reported.

The NORTHIV trial recruited a total of 242 patients in Sweden and Norway, 239 of whom received at least one dose of the study drugs. The patients were included from 2004 to 2007 and randomised to either one of three different treatment regimens: (a) efavirenz 600 mg q.d. + 2 NRTI q.d., (b) atazanavir 300 mg q.d. + ritonavir 100 mg q.d + 2 NRTI q.d., or (c) lopinavir 400 mg b.i.d + ritonavir 100 mg b.i.d. + 2 NRTI b.i.d. Most patients on lopinavir started with the Kaletra® soft gel capsules (SGC). However, during the NORTHIV trial, the tablet formulation of Kaletra® was introduced. Some patients started with tablets and many switched from SGC to tablets during the course of the study. Subjects are to be followed for a protocol-stated 144 weeks.

This brief outline of NORTHIV gives rise to some observations that serve well to illuminate the rapid development of the field during the years when the research here reported was conducted. One feature of the trial design which seems very problematic in retrospect is that lopinavir, which is dosed bi-daily in this study, was supposed to be given together with 2 NRTIs of the individual physicians choice given twice daily, whereas the other arms, where the test medication (efavirenz and atazanavir) was dosed once daily, would be combined with once-daily NRTIs. This no doubt seemed quite reasonable at the time, given that this was a naturalistic, open-label trial set up to investigate the performance of efavirenz, lopinavir or atazanavir based antiretroviral therapy in “real life”. But as has probably been made clear by this review, real life when it comes to antiretroviral therapy has been a rapidly changing phenomenon. Apart from the problem mentioned above, that a new formulation of one of the study drugs appeared, more importantly the view of the scientific community on the relative merit of the NRTI combination generally given twice daily (zidovudine/lamivudine), and therefore most often co-prescribed in patients randomised to lopinavir-based therapy, compared to abacavir/lamivudine or tenofovir/emtricitabine, has changed profoundly during the period in question.
Between the inception of the NORTIV trial and today emerged a number of studies that linked zidovudine to the dreaded lipoatrophy syndrome (e.g.,[192-194]) mentioned previously in my description of the development of the field in general, whereas there is still no compelling evidence that abacavir or tenofovir would be causative of this disturbing side effect. Also, results from the Gilead 934 study were presented, demonstrating superiority of tenofovir over zidovudine in terms of overall efficacy and tolerability [195].

I believe that this experience is illustrative of the problems of designing long-term strategic trials when standard-of-care is under virtually constant change, as has been the case in the field of HIV therapy.
In 2004, a letter to the editor by Fröhlich et al was published in *British Journal of Clinical Pharmacology* [196]. This described the pharmacokinetics of saquinavir in 19 participants given Invirase® or Fortovase® (two different saquinavir formulations) in drug-drug interaction studies that had been analyzed in relation to expression of CYP3A5. 14 /19 were homozygous for CYP3A5*3 (the most common allele among Caucasians, coding for a splice defect, resulting in a lack of functional CYP3A5) and five were heterozygous for wild-type CYP3A5*1. While this sample was too small to detect differences in plasma pharmacokinetics between non-expressors of CYP3A5 and heterozygotes for the wild-type, data on the 24h urinary metabolic ratio of saquinavir over its M2 and M3 metabolites clearly indicated that CYP3A5 was involved in the metabolism of this PI (which might be suspected since all available PIs are substrates of CYP3A4, and the substrate specificity between these members of the CYP3A subfamily is assumed to be roughly overlapping, as will be elaborated below).

This observation prompted the study reported in paper I. At the time of planning the study, there were also data on the effects of CYP3A5 polymorphisms on indinavir metabolism, presented by Anderson et al at the 11th CROI, San Francisco 2004, and later published [197]. In a retrospective cohort of 33 patients, CYP3A5 expressors (hetero- and homozygotes for *1) were calculated to have a 46% higher oral clearance than did non-expressors. We were not aware of any other in vivo studies of the impact of CYP3A5 genotype on the pharmacokinetics of PIs.

### 7.1 CYP3A5

CYP3A is the major drug metabolizing enzyme family in human adults and is involved in the metabolism of more than 50% of clinically used medications [198]. Cytochrome P450 3A4 and 3A5 are both members of the CYP3A subfamily of enzymes [199]. CYP3A4 is in all cases the major contributor to CYP3A-activity, which varies widely, partly due to the involvement of CYP3A5 in the metabolism of substrates of CYP3A4. Indeed, there appears to be great similarity in substrate specificity of the two enzymes [200-203].

The wild-type CYP3A5*1 variant of CYP3A is present in 5-30% of Caucasians and in 60-73% of African Americans [204-206]. Thus, among Caucasians, wild-type is
a minority allele. When CYP3A5 is expressed, it is mainly so in the gut, and contributes only to a minor extent to hepatic drug metabolism [207]. The low frequency of CYP3A5 expressors among Caucasians is predominantly due to the high frequency of the splice variant CYP3A5*3 [204, 208, 209]. Although CYP3A5 is thus principally expressed in the black population, influence of CYP3A5 genotype on the metabolism of drugs in Africans was not very well investigated.

Thus, available knowledge on the putative clinical impact of CYP3A5 expression was low, the field having been hampered by the difficulty of recruiting individuals homozygously expressing CYP3A5 in Caucasian populations. Indeed, the maximum number of CYP3A5*1 homozygotes in the same study, that I could find upon reviewing the field, was 16. Data on the putative impact of CYP3A5 genotype on otherwise well described CYP3A substrates was sparse. For instance, while it could be considered demonstrated that CYP3A5 expression did impact the oral clearance and dosing requirement of tacrolimus, data on cyclosporin and midazolam were conflicting [210].

7.2 The Tanzanian connection

Co-workers on this paper had recently conducted a study on the impact of CYP3A5 on quinine metabolism, in collaboration with workers at the Muhimbili University College of Health Sciences, Dar es Salaam [209]. As part of this effort, 144 healthy, black Tanzanians had been genotyped for the presence of the A6986G polymorphism (defining the globally most common, CYP3A5*3 allele, yielding a splice defect causing premature termination of translation), the G1690A SNP (CYP3A5*6 yielding a splicing defect) and the G27131-32insT (CYP3A5*7 insertion yielding a frame-shift), which collectively have made up the vast majority of alleles in African populations [204]. The allele frequency in this cohort was: *1: 0.51, *3: 0.19, *6: 0.18 and *7: 0.12 [209]. It was decided to pick individuals in the cohort that either lacked the wild-type CYP3A5*1 allele (that is, having any combination of *3, *6 or *7), or who were homozygous for CYP3*1 (lacking *3, *6 or *7 alleles), and to compare saquinavir exposure after a single dose of Fortovase®. The intra-individual variability of exposure for this drug is rather large, driving up the number of subjects necessary to show a gene effect. We opted to compare only the respective homozygotes, as the impact of heterozygotism on PK parameters of various drugs had often been difficult to demonstrate.
7.3 Fortovase®

Though saquinavir was not much used in Sweden, the aforementioned pilot data and the possibility of cooperating with the group that had generated them, prompted us to choose this drug as a model substrate for studying the effect of CYP3A5 genotype on the metabolism of a PI (a decision that may seem strange in retrospect, as this first-generation PI has been notorious for its problematic bioavailability). Since our in-house saquinavir assay was an HLPC-UV with low sensitivity, and we did not have an assay for the metabolites, we contacted professor Haefeli at the University of Heidelberg, whose group had been responsible for the initial report of a putative impact of CYP3A5 on saquinavir metabolism. It was agreed that the samples from our study was to be analysed at Heidelberg using their LC-MS/MS method.

Two different formulations of saquinavir have been produced by Roche, a hard gel capsule (Invirase®) and a soft gel capsule (Fortovase®). When administered without ritonavir, they have a bioavailability of 4 and 12% respectively [211]. As stated, at the time of the study, unboosted saquinavir was no longer used; furthermore the Invirase® hard gel capsule formulation had been selected for use together with ritonavir, since its bioavailability turned out to be more favorable than the soft gel formulation Fortovase® in this situation. Yet, Fortovase® was chosen for our study, as its pharmacokinetics when given without ritonavir display a lower variability than does Invirase®. Our study was to be conducted without ritonavir; therefore this would be advantageous.

7.4 Saquinavir is a substrate of CYP3A5

The recruitment and execution of the trial was performed in Dar es Salaam, Tanzania. Prior to analyzing date, the genotype assignments of the volunteers were reconfirmed. The subjects received a single dose of 1200 mg Fortovase®, and saquinavir concentrations were followed from 0-24 hours. As, according to protocol, we did not follow plasma drug concentrations beyond 24h, and in many cases there were not more than two measurement points for an estimation of the terminal k, we did not calculate AUC0-inf, but only 0-24h. Our results, though nominally statistically “non-significant” (P=0.0533) indicated that oral clearance of saquinavir was on average approximately 50% higher when 2 active alleles rather than 0 was present. As expected, however, the overlap in exposures between groups was striking. The primary analysis was supported
by a comparison of the metabolic quotas of saquinavir/M2+M3 metabolites in urine, showing a significantly lower quota in homozygotes for the wild-type allele. We considered our study a demonstration of the functional involvement of CYP3A5 in the metabolism of saquinavir. This may also be the case for other PIs (e.g., indinavir, as implied by the data of Anderson et al [197]), but would have to be demonstrated for each of these drugs.

We never considered that we had generated evidence of a clinically relevant effect, though given the average magnitude of the genotype effect, it might actually have impacted outcome if saquinavir had been used unboosted, where the drug exposures reached are no doubt on the steep part of the concentration-effect curve. This, however, would certainly not have been considered standard-of-care at the time of the study.

A subsequently published paper reported a test of the hypothesis that CYP3A5 genotype might impact lopinavir trough concentrations in a clinical cohort of approximately 100 patients treated with ritonavir-boosted lopinavir (11 of whom where homozygous for the wild-type allele). In this, no effect of CYP3A5 genotype could be seen on exposure to this ritonavir-boosted PI [212]. In vitro, ritonavir is not only a potent inhibitor of CYP3A4 but also of CYP3A5 [213, 214]. Thus, as it seems today, though CYP3A5 may be involved in the metabolism of presently used PIs, the clinical impact is likely to be minimal.
8 PAPER II: ATAZANAVIR, BILIRUBIN AND BOOSTED PI MAINTENANCE MONOTHERAPY

8.1 The design of boosted PI monotherapy trials

This paper reports a study of maintenance (as opposed to induction) monotherapy, as patients were required to have been successfully treated with undetectable viremia for a minimum of one year, in combination with being PI naïve. Of note, the population of our study is somewhat distinct from a population that have achieved undetectable viral load on a ritonavir-boosted PI based first line triple combination, and is then randomized either to continue with triple therapy, or to discontinue the 2 NRTI and only remain on PI/r monotherapy. That, arguably, is the paradigm for PI/r monotherapy studies that has since evolved (though a recently reported study on darunavir monotherapy breaks this pattern [215]). The advantage of a population already experienced to the PI/r, as opposed to the one defined in paper II, from the point of view of the design of trials meant to show non-inferiority, is that tolerability issues for the experimental agent are conveniently reduced through enrichment in the induction run-in period.

8.2 The evolving view of PI monotherapy

The introduction to paper II begins as follows: “Presently the recommended initial therapy for HIV-1 infection consists of 2 NRTIs combined with 1 NNRTI or 1 PI. Important long term adverse effects include lipoatrophy, fat accumulation and the metabolic syndrome, with component drugs often acting in synergy. Thus, it would be of great value if treatment could be simplified using as few drugs as possible, with these preferentially being those with the least long term toxicity.”

Here we have reference to the previously mentioned lipoatrophy syndrome, today thought primarily to be due to the mitochondrial toxicity of thymidine analogues [75, 216]. In critical retrospect, the talk of preferentially using the drugs with least toxicity may sound a little bit funny, as the only category of antiretrovirals that could feasibly be given as monotherapy is in fact ritonavir-boosted PIs. However, we refer particularly to the favorable metabolic profile of atazanavir [217]. Of interest, as
mentioned in the methods section, there was to be an accompanying study of body fat composition over 72 weeks of treatment, which failed for methodological reasons.

Reading further through the introduction, it is unmistakable how the spook of earlier “functional monotherapy” experiences was overshadowing this endeavor. Note that this was the time prior to the approval of darunavir, etravirine, maraviroc and raltegravir. The strict criteria of treatment failure and the careful stopping rules of this small pilot trial demonstrate the feeling that we could potentially seriously limit future treatment options and implicitly health outcomes, due to resistance following failure of boosted PI monotherapy. Knowledge on this has certainly progressed, and today we have a relatively good view of what can be accomplished with monotherapy over 48-96 weeks. An excellent systematic review and meta-analysis has recently been published on this topic [218].

In brief, indications are that induction monotherapy is less effective than standard triple therapy [219], and this is no longer contemplated as a widely used option for treatment initiation, if ever it was. When boosted PI monotherapy is used in patients that already have suppressed viral replication the picture is different. Depending on the definition of the population analyzed (ITT or some form of OT population) and the definition of failure (e.g. strict ITT where successful NRTI salvage is not considered failure, or the so-called TLOVR criteria [220] or only counting failure with consequences for future treatment options) efficacy is roughly similar or only slightly lower than that of traditional triple therapy [218]. On the whole, however, it seems undeniable that drug pressure with, at least lopinavir monotherapy, is lowered to an extent that generates detectable traces. For instance, an increased number of episodes of transient low grade viremia (blips) have been detected in several trials, and there is a higher risk of acquiring PI resistance when failing virologically with boosted PI monotherapy, than with boosted PI-based triple therapy [127], again highlighting the importance of the relation of drug pressure and barrier to resistance.

As I am writing, two studies on boosted darunavir maintenance monotherapy have recently been reported as abstracts at the International AIDS Society meeting (one with once-daily and one with twice daily dosing) [215, 221]. Both of these describe formal non-inferiority of darunavir monotherapy and triple therapy in the per protocol population, though again more blips were reported with monotherapy in one study [215] whereas in the other non-inferiority of monotherapy was not met in the ITT population [221]. Of note, it is difficult to draw solid conclusion from studies yet presented only as abstracts, particularly when one did not attend the meeting. Still, the
initial impression is that these reports confirm the general picture of boosted monotherapy given above, though darunavir, given earlier comparative data when co-treating with other agents, may have some edge over lopinavir.

It is interesting that lopinavir concentrations were not predictive of response to monotherapy, and no MEC could be identified [219, 222]. This may be an indication that MEC is not significantly increased in boosted PI monotherapy, perhaps despite a lower maximal efficacy (Emax), as suggested below.

Finally remains the important spectre that some agents might not fully penetrate to deep compartments such as CNS. This was fuelled by reported observations of detectable CNS viremia despite virologic control in plasma, in patients treated with boosted atazanavir monotherapy [223]. Again, data emerging as I am writing are of interest; it is notable that CNS viremia and CNS symptoms (in one case described as “HIV encephalitis”) was reported in 2/112 patients treated with boosted darunavir monotherapy [221]. The CNS viremia was reversible on addition of 2 NRTI. These data remain to be published. It will be interesting to see how they will affect our perception of boosted PI maintenance monotherapy.

8.3 The TDM approach in the atazanavir monotherapy study

In the discussion section of paper II it is stated that drug concentrations were to be measured at weeks 4 and 8 and at a presumed steady state after any dose adjustment. Then follows a statement which is in much need of further explanation: “Shortly afterwards sufficient data was available at the Department of Clinical Pharmacology, Karolinska University Hospital, to distinguish adequately the range of atazanavir plasma Ctrough measurements reached in treatment with the ritonavir-boosted atazanavir regimen (presumed not to be inferior to lopinavir/ritonavir) and the range seen in treatment with the unboosted 400 mg once daily regimen (which has failed to show a lack of inferiority with the same regimen). On this basis, the study protocol was changed allowing a dose increase of atazanavir in patients with a plasma concentration less than 500 ng/mL, if this was possible without inducing manifest jaundice. The ritonavir dose was specified always to remain at 100 mg once daily.”

What needs to be justified in retrospect is not the factual statements on the relative efficacy of atazanavir at the two different dose regimens (400 mg q.d unboosted and 300 mg q.d boosted by 100 mg ritonavir), and that underlying this there should be some sort of concentration-response relationship (in treatment-experienced
patients), but the tacit assumption that this fact could be utilised to optimise therapy in a trial of a different sort of population than that from which my general conclusion was drawn (that is, boosted PI monotherapy with fully susceptible virus, rather than treatment-experienced with varying degrees of resistance), and also that the subsequent prediction would be sufficiently accurate to be used to guide dosing and thereby avoid therapeutic failure. Underlying all this it is, perhaps surprisingly, clear that we considered this knowledge sufficiently well-founded, that disregarding it could be to put the patients at disadvantage. In fact, some words from the discussion section cited below, will further guide us as to how this came to be.

From a theoretical perspective, we assumed that the MEC might be increased in monotherapy, much as I have argued that it was in the efavirenz dose ranging trials, when efavirenz did not receive full support from the NRTI. This, of course, was the rationale for suggesting that the therapeutic range we had (rightly or wrongly) identified for atazanavir in treatment-experienced patients might in fact be applicable in the setting of monotherapy, though we fully recognised that it was not applicable to the treatment naïve treated with triple therapy. At second thought it is, as mentioned above, clear that, rather than a shift in the MEC, monotherapy may in fact only be characterised by a certain decrease of the Emax of the treatment regimen. In that case, the optimisation of atazanavir exposure attempted in our study may have been as useless as increasing the dose of ritonavir-boosted atazanavir in a treatment naïve patient with reasonably ordinary drug exposure.

It is interesting to compare our TDM approach reported in paper II with the aforementioned paper by Ståhle et al, that for a long time informed our TDM practice [152], in that retrospectively assumptions on the MEC seem exaggerated in both cases. I hope that my general introduction to the development of antiretroviral therapy has served to demonstrate the intellectual context that prompted this tendency. I also hope that this makes it clear why the preceding discussion of the paper by Ståhle et al is important for the understanding of paper II. For a long time, ultimately based on unboosted PI experiences, our TDM service had a tendency to recommend dose increases beyond labelling for efavirenz and atazanavir. In retrospect, it is my personal view that these were mistakes that need to be justified – which I hope to do with this account.

Now, to skip right to the results section, in relation to the protocol stating that the ritonavir dose should always remain 100 mg/d, it is stated that “the first patient with virologic treatment failure was repeatedly shown to have a low plasma concentration
of atazanavir (250 ng/mL)”. Reading this with that just stated in mind, one wonders: how were we so sure that this was low?

Yes, it was relatively low for a patient with ritonavir-boosting, but too low? Adherence was good and there was no use of acid-reducing agents (which reduce the bioavailability of atazanavir). And then - the ritonavir dose was raised. This would be expected to have very minor, if any, effect on atazanavir exposure. I vaguely recall our discussion about this patient, where for some subjective reason of the patient’s (which I do not remember), it was considered that increasing the atazanavir dose was not practically feasible. I cannot reconstruct how we argued between ourselves as we raised the dose of ritonavir. And indeed, as subsequently stated: “this procedure did not increase the atazanavir concentration”.

And then in the discussion: “The first patient to fail in our study showed repeatedly low atazanavir concentrations, despite, apparently, being fully compliant with the study regimen and not using any interacting drugs. After this, the protocol was changed, based on local therapeutic drug monitoring (TDM) experiences (Josephson et al, manuscript submitted), to enable an increase in the atazanavir dose until a Ctrough of >500 ng/mL was reached, if this was possible without inducing manifest jaundice. Despite this, another 4 patients failed. In the whole study population, no correlation was seen between atazanavir drug concentrations and virologic outcome.”

What was the full basis of the TDM policy that apparently failed to influence the results of our monotherapy trial? I have previously given a general outline of this. Yet, for the reader to understand the perhaps somewhat confusing intervention of the clinical pharmacologist in this work, I must discuss the manuscript that was forthcoming according to the discussion section of paper II. This paper contains the full argument behind the 500 ng/mL cut-off (which we had lowered to approximately 430 ng/mL as the dataset at our lab grew larger).

8.4 The unpublished manuscript

The abstract of this manuscript, which was submitted to and rejected by, Therapeutic Drug Monitoring, reads as follows:

Objectives: Atazanavir is an HIV protease inhibitor (PI) that has been approved in two different dosage regimes: (a) 300 mg daily + ritonavir 100 mg, and (b) 400 mg daily. Non-inferiority to state-of-the-art PI lopinavir has been shown for the ritonavir-boosted
atazanavir regime in treatment-experienced HIV-infected patients, whereas unboosted atazanavir has been inferior to lopinavir in two trials. By comparing the ranges of drug exposures seen in the two regimes, we attempted to define a lower atazanavir target concentration in treatment-experienced patients. Furthermore, since hyperbilirubinemia is a concentration-dependent side effect of atazanavir, we evaluated s-bilirubin as a biomarker for adequate atazanavir exposure.

**Methods:** Routine TDM atazanavir plasma concentrations were analyzed by HPLC with UV detection. Samples taken >15 hours post dose were defined as “trough”.

**Results:** In patients treated with atazanavir 300 mg daily and ritonavir 100 mg daily, the median concentration was 1.13 µmol/L, and the interquartile range 0.60-2.02 µmol/L. In patients treated with unboosted atazanavir 400 mg daily, the median concentration was 0.43 µmol/L, with an interquartile range of 0.1-0.63 µmol/L. Thus 75% of boosted samples were above 0.6 µmol/L, while 75% of unboosted samples were below 0.6 µmol/L. A s-bilirubin >25 µmol/L had a positive predictive value of 85% for an atazanavir trough concentration >0.6 µmol/L.

**Conclusions:** We suggest 0.6 µmol/L (429 ng/mL) as a lower atazanavir target concentration >15 hours post dose in treatment-experienced patients. An elevated s-bilirubin (> 25 µmol/L) is predictive of an atazanavir exposure above this limit.

This argument is further illustrated by the graph on page 61, showing descriptive statistics of atazanavir trough concentrations (post > 15 hours) in the two dosage regimens that have been investigated in phase III clinical trials: atazanavir 300 mg + ritonavir 100 mg daily, and atazanavir (ATV(r)), and atazanavir (ATV) 400 mg daily.

We did not further pursue the publication of this paper. This was because our particular dataset is of little importance; it was the approach that mattered. Still, apart from the attempt to derive a MEC for atazanavir on the basis of differential dose effects, it heralds the interest in plasma bilirubin as a biomarker of atazanavir exposure, an approach that was to be utilized in the analysis of the monotherapy data (see below).

Criticism that may be extended toward our approach include that it may be rather insensitive in terms of clearly defining valid MECs due to, e.g., the overlapping exposures between populations and the variable levels of resistance, and perhaps the varying activity of other drugs in the regimens of treatment experienced patients. Also, the definition of a Ctrough (>15 hours post dose for a drug with an 8 hour half-life and a 24 hour dose interval) appears overly wide, as indeed was noted by
one of the peer reviewers. Some sort of imputation or a modelling approach would have been of value.

Remember, however, as already related, that any TDM strategy, regardless of its target concentration, could probably be confirmed to be effective in properly sized clinical trials, if the strategy served to increase median doses in a situation where a large portion of patients given the standard dose did not reach exposures over the MEC. In other words, our approach might retrospectively be considered justified if the MEC in these populations (monotherapy and treatment experienced) had been such that the general trend of guidance based on our target range had served to bring a substantial proportion of patients above it, regardless of whether our particular target was correct.

What was important, and driving our application of atazanavir TDM, was that all patients should have exposures typical of ritonavir-boosted atazanavir rather than of unboosted atazanavir therapy. Evidence clearly pointed to unboosted therapy being less potent in terms of viral effect, and that boosting prevents the emergence of resistance [95, 160]. Also, early maintenance studies of unboosted PI had failed. [224-226]. If the
exposures indeed were in this “normal range” of ritonavir-boosting, we felt that we were on solid ground in our boosted PI monotherapy study.

8.5 Failure due to the use of acid reducing agents?

Our monotherapy study was prematurely terminated according to protocol, since the number of failures had exceeded what, for statistical reasons, was the lowest number of failures not indicative of an increased failure rate in monotherapy. After all, you do pilot studies when you really do not know. In the context of understanding the causes of failure, it is interesting to revisit the discussion around acid-reducing agents. As mentioned above, such drugs lower the bioavailability of atazanavir; indeed co-treatment with proton pump inhibitors is contraindicated, and the use of other agent should be appropriately separated in time from atazanavir intake [227]. In our study, two patients out of five (though it says four in the discussion of the paper) that subsequently failed had been using acid-suppressing drugs in violation of the study protocol. Interestingly, our TDM service had documented drug concentrations in these patients that are quite ordinary when treating with ritonavir-boosted atazanavir (ranging between 700-1400 ng/mL). “Excluding such use may actually be difficult in routine clinical practice”, we wrote. “The clinical consequences of cotreatment with acid-suppressing drugs in triple ART based on ritonavir-boosted atazanavir are not fully clarified. Our results imply the quite feasible notion that the deleterious effects of this interaction may be particularly acute in the setting of atazanavir monotherapy.”

Again this paper demonstrates a failure of our TDM service. Concentration monitoring apparently failed to affect therapeutic outcome, did not in any way indicate any particular effect on atazanavir exposure due to the use of a proton pump inhibitor or a H2-blocker, and did not predict failure. In this context, it is worthwhile noting that we had unthinkingly stepped into the very trap of which van Heeswijk had warned us; we were both evaluating TDM and using it to guide therapy at the same time.

8.6 Plasma bilirubin as a marker of virologic outcome in atazanavir therapy

The main exposure-limiting side effect of atazanavir is unconjugated hyperbilirubinemia, caused by inhibition of the bilirubin-conjugating UGT1A1 enzyme. The increase in unconjugated bilirubin is variable and dependent on the genotype of the UGT1A1 promotor region [183-185, 228], which is also the cause of
the Gilbert syndrome, a benign state of mild, chronic unconjugated hyperbilirubinemia in the absence of liver disease or overt hemolysis. In this condition, which is due to a number of genetic polymorphisms, primarily in the promotor region of the gene coding for UGT1A1, hepatic glucuronidation activity, essential for efficient biliary excretion of bilirubin, is reduced to about 30% of normal [229]. Though over a hundred UGT1A1 variants have been described, with varying functional consequences, a TA insertion in the promotor region, defining an allele termed UGT1A1*28, is the most common cause of Gilbert’s syndrome in Caucasians [186]. The allele frequency is 0.4 with approximately 10% being homozygous and at risk for symptomatic disease.

In fact, due to its UGT1A1 blockade, atazanavir causes what one may call a “functional” Gilbert syndrome (which is more likely to cause symptomatic jaundice if the patient indeed has a genetic Gilbert syndrome). The proportion of patients using ritonavir-boosted atazanavir with levels of bilirubin >2.5 times the upper reference level are 50-60% [109, 160] though, as stated previously, the proportion discontinuing due to jaundice have been 0-3%.

A short report by Petersen et al in AIDS where they described the use of bilirubin for the monitoring of adherence to atazanavir therapy had caught our attention [230]. Once we were alerted to this possibility, we certainly did start noticing a prevalent pattern in s-bilirubin in patients initially responding but then rebounding on atazanavir-based therapy. A normal baseline bilirubin would increase to somewhere above the upper level of normal and remain so until viral rebound, when bilirubin would be back within normal limits. As data did not support any attenuation over time of the effect of atazanavir on UGT1A1 function [231], it was albeit impossible not to interpret this as a sign that adherence had lapsed or that the patient had discontinued atazanavir. The graph on page 64 represents the relation between on-treatment bilirubin (rather than increase from baseline, since we often did not have baseline bilirubin measurements) and trough (<15 hours post dose) atazanavir concentration in the cohort at the Karolinska University Hospital in Stockholm.
As documented in paper II, when analyzing the monotherapy study, the average Ctrough measures did not differ between subjects failing virologically and those who did not. Yet, the average serum bilirubin value was significantly lower in those who failed. Indeed, in figure 1 of the paper, it can be seen that 7/7 patients with mean on-treatment plasma bilirubin > 40 µmol/L remained virologically suppressed, as against 1/6 with bilirubin ≤ 40 µmol/L (this OT population analysis excludes one patient that discontinued due to icterus and one patient excluded from the study between screening and baseline).

This was an intriguing finding, and I go on to quote our discussion of this, which I find succinct, though in retrospect it fills me with some regret that we did not go on to investigate this further. In fact, one of the issues we considered at several times was to concomitantly study the diurnal evolution of plasma bilirubin and plasma atazanavir during one dosing interval. I am as yet not aware of any such data, which, as any one can see from the following, would have been of value for the interpretations of the bilirubin findings in the monotherapy study (indeed, Bristol-Myers Squibb may have data on file). This is what we wrote at the time:

"In this study, the average atazanavir Ctrough measurements did not differ between subjects failing virologically and those who did not. Yet, the average serum bilirubin value was significantly lower in those who failed. One interpretation of our finding might be that the drug concentrations measured in the patients with virologic
failure reflect improved drug adherence in the period immediately preceding a scheduled visit to the clinic and that serum bilirubin, to some extent, functions as a marker of long-term atazanavir exposure. Two sets of fact, however, are incongruent with this hypothesis. First, our adherence data do not support lower drug adherence in the failing patients. Second, published data on the kinetics of radioactively labeled unconjugated hyperbilirubinemia imply a typical terminal half-life of approximately 10 hours [232], which is of the same magnitude as the half-life reported for ritonavir-boosted atazanavir [228]. Thus, it is not evident that measurement of serum bilirubin would be a more powerful marker of cumulative atazanavir exposure than measurement of the actual drug concentration.”

A pharmacologist would note that if half-life is determined by a single dose of an injected radioactively labeled tracer, one might erroneously interpret the last detected elimination phase and its decay constant as representing the terminal elimination half-life. In the event of a truly deep compartment, such a study might not be able clearly to distinguish slow equilibrium with a putative deep compartment, confounding inter-compartmental and elimination clearance, and thus underestimating the physiological half-life of the substance (I owe it to my friend, doctor Jonatan Lindh, to recognize this possibility). This risk of this would depend on the sensitivity of the assay and the time-span over which the marker is measured after administration.

“Another possible interpretation of our finding, assuming that it is not attributable to pure chance, is that the degree of serum bilirubin increase at a given atazanavir exposure may not only be related to the UGT1A1 promoter genotype but to the intracellular availability of atazanavir. The site of drug action of PIs is intracellular, as is the site of action of atazanavir as a UGT1A1 inhibitor. In vivo intracellular drug concentration measurement is complicated and is seldom performed. Therefore, little is known about the relation between plasma and intracellular atazanavir concentrations. It is interesting to speculate that the low serum bilirubin values seen in patients experiencing virologic failure might correlate with low intracellular availability of atazanavir. This hypothesis should be further investigated.”

Here we are posing very relevant fundamental questions about the mechanism of a clinical finding. I admit, however, to having done nothing to further pursue the truth about these matters.

Finally, to end this discussion of the use of bilirubin as predictor of virologic outcome in atazanavir therapy, we were also able to tease out a hint of a relationship in the NORTHIV dataset, as demonstrated in paper IV of this thesis, to be further
discussed below. All these data are, of course, inconclusive. The best thing would be to retrospectively investigate the relation between bilirubin increase from baseline (or simply on-treatment bilirubin levels, as we did, due to the lack of baseline values in many cases) and outcome in the now numerous large clinical trials conducted with atazanavir, boosted and unboosted. In fact, I recently discussed this issue with doctor Zhu, a clinical pharmacologist working on atazanavir for Bristol-Myers Squibb, and responsible for the presentation of PK/PD data for the CASTLE study at the 10th Conference of Clinical pharmacology of HIV therapy in Amsterdam, April 2009. He denied being aware of any analysis of the relation between on-treatment serum bilirubin and outcome in the pivotal trials.
9 PAPER III: TESTING A BIOMARKER OF CYP3A ACTIVITY IN THE SETTING OF NORTHIV

9.1 The discovery of a putative endogenous CYP3A biomarker

The saga of 4β-hydroxycholesterol as a marker of CYP3A activity begins with an outlier observation in a study of plasma lipids. The investigators noted that one of the healthy volunteers showed very high levels of this cholesterol metabolite. At screening for the study, it had been overlooked that this volunteer was actually treated for epilepsy with carbamazepine, a known inducer of CYP450 enzymes, and, of particular importance, CYP3A. This led to the hypothesis that the elevated level might indicate that the formation of this particular cholesterol metabolite was due to one of the enzymes induced by carbamazepine. If so, its blood levels might correlate with the activity of this or that hepatic enzyme.

Finding out whether a substance is in fact an inducer of this or that hepatic drug enzyme is today considered an important part of the pre-marketing study of most putative therapeutic drugs. Though there are in vitro methods for investigating this (of which I am not very knowledgeable, and therefore will not further discuss), studies with probe drugs in healthy volunteers are often required to characterize this possibility. A maximally facile approach for determining CYP3A activity would be of great value in many a setting, in drug development, for academic studies, and perhaps even in the clinics.

The further line of inquiry is described in a paper by Bodin et al [233]. The observation that carbamazepine might in fact be causative of the increased plasma concentration of this particular cholesterol metabolite, was corroborated by the finding of high levels not only in patients treated with carbamazepine, but also with other anticonvulsants known to induce hepatic drug metabolizing enzymes, such as phenytoin and phenobarbital. However, as evidenced for instance by a recent publication with the Karolinska cocktail (probing the activity of CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A simultaneously), after enzyme induction by rifampicin [234] hepatic enzyme induction, (whose molecular mechanisms I shall not further discuss), is in many cases not specific to one single enzyme (indeed, the aforementioned paper reveals a varying dose-dependency of the induction of the different CYP450 enzymes). Therefore, Bodin et al also reported data from in vitro
incubation studies suggesting that cholesterol was formed through enzymatic conversion by CYP3A4 but not CYP1A2, CYP2C9 or CYP2B6 [233]. Further, it is considered that auto-oxidation is not a major source of 4β-hydroxycholesterol [235]. From this, the conclusion was drawn that 4β-hydroxycholesterol could be a reasonably specific marker of CYP3A activity (as opposed to, say, a more general indicator of hepatic enzyme induction).

In the program to further characterize this putative endogenous marker of CYP3A activity, the plasma half-life of 4β-hydroxycholesterol has been investigated. This was first estimated to approximately 60h after a small study of the fate of injected deuterium-labelled 4β-hydroxycholesterol [236]. A later study increased this estimate to approximately 17 days [237] (incidentally this discrepancy illustrates the previously stated concerns on the accuracy of the half-life estimate for bilirubin, based on the kinetics of a single-dose of radio-labeled substance). This latter study also demonstrated the stability over time of 4β-hydroxycholesterol levels in volunteers not treated with any drugs. There is no evidence of diurnal variation due to meal intake; indeed, this is congruent with the very long half-life evidenced by study. Also, at physiological cholesterol levels, the correlation between 4β-hydroxycholesterol and cholesterol levels is weak, as evidence by the data reported in paper III as well as elsewhere [238].

9.2 Long half-life a merit or a problem?

The long half-life and stability of 4β-hydroxycholesterol is an interesting feature of this bio-molecule. At first, it gives hope that this may be a reliable biomarker, not displaying the intra-individual variation due to, e.g., diurnal variation, that has been reported for another suggested endogenous metric of CYP3A activity, the urinary 6β-hydroxycortisol/cortisol ratio [239, 240]. Other workers in the field, however, recently noted that this long half-life might indeed be a problem for 4β-hydroxycholesterol as a biomarker to predict the effects of a CYP3A inhibitor on CYP3A activity, as measured by a pharmacological probe [241]. This so if the co-administered CYP3A inhibitor has a relatively short half-life, as for instance ritonavir, where full CYP3A inhibition may not last throughout the dosing interval. In fact, the results reported in paper III which did not fully agree with prediction (as will be discussed below), might be explained by this mechanism – that 4β-hydroxycholesterol is likely to reflect the average level of
CYP3A activity over time, rather than the maximum inhibition of CYP3A seen after administration of a CYP3A inhibitor with a moderately long half-life.

### 9.3 The NORTHIV trial as a vehicle to investigate 4β-hydroxycholesterol

The idea that would form paper III, was to use the NORTHIV clinical trial, where patients were randomized to antiretroviral therapy based on efavirenz (enzyme inducer), ritonavir boosted atazanavir (some induction, strong net CYP3A inhibition) or ritonavir-boosted lopinavir (strong underlying induction, strong net CYP3A inhibition), for the investigation of the operative characteristics of this biomarker. We were to test whether this putative marker of CYP3A activity would reflect well-known effects on CYP3A activity due to enzyme induction and inhibition in the respective treatment arms of NORTHIV. The expectation was that plasma 4β-hydroxycholesterol would increase with efavirenz therapy and decrease when treating with boosted PIs, though we knew that the results would somewhat be complicated by the fact that the boosted PI regimens entailed both induction and inhibition (as yet, the operative characteristics of 4β-hydroxycholesterol have not been well characterized in the presence of a strong CYP3A inhibitor that does not also have enzyme inducing properties).

### 9.4 CYP3A induction and inhibition

Revisiting the paper, there are some things on which I want to comment, including the interpretation of the point where the experiment failed to verify the hypothesis. Firstly there was the issue of 4β-hydroxycholesterol being a metabolite of cholesterol. It is well known that therapy with all the drug regimens of the NORTHIV trial have been associated with effects on blood lipids, including increases in total cholesterol [104, 129, 228]. Therefore we were aware that we would have to clearly demonstrate that any putative findings were not secondary to such general effects. I think that the data presentation in the paper, including figure 2, is successful in showing that this was not the case, as we also argued in the discussion when trying to make sense of the data obtained from the patients treated with ritonavir-boosted lopinavir, which were not in accordance with prediction.

Secondly, I want also to point out that no effort was made to correlate drug concentrations with changes in 4β-hydroxycholesterol (though I admit that we did fit
the data). The reason for this is no doubt already clear to the reader. On the one hand, drug exposure might certainly affect the levels of 4\(\beta\)-hydroxycholesterol, this being a marker of CYP3A activity (in fact, the paper by Kanebratt et al that has previously been mentioned clearly demonstrates the dose-dependence of hepatic enzyme induction with rifampicin [234]); on the other hand, as a marker of CYP3A activity, 4\(\beta\)-hydroxycholesterol levels were assumed to correlate with CYP3A activity. And CYP3A, of course, is known to be involved in the metabolism of all of the experimental drugs (though CYP2B6 is of principal importance for efavirenz metabolism). Thus, it is difficult to see how one could have made sense of concentration-effect data.

Thirdly, there were some outliers in this population, individuals going against the predicted and observed general trends after initiation of efavirenz and atazanavir. We never found any clear explanation for this in the available charts (e.g. concomitant medications). It is also notable that the pre-treatment levels of 4\(\beta\)-hydroxycholesterol seem to have been somewhat higher in these HIV-infected individuals than in healthy volunteers. We could not come up with any particular explanation (again no co-medications that were informative). Interestingly, in a subsequently published paper by Tomalic-Scharte et al, to which I shall return, the same phenomenon was observed [241].

The final point I want to discuss concerning the details of paper III, is the outcome in the arm treated with ritonavir-boosted lopinavir (which was recently corroborated in the study of Tomalic-Sharte et al [241]). Though the trend in the point estimate is in the right direction given the hypothesis, the effects were quite variable and the outcome was not statistically significant. Despite the fact that this regimen indubitably gives a significant background enzyme induction, I felt that data generated with the midazolam probe [89] (supported by later data with alfentanil [242]), as well as the clinical data on the very strong CYP3A inhibition characteristic of this regimen, did not lend any credibility to the notion that the relatively small (or perhaps lack of) change in 4\(\beta\)-hydroxycholesterol levels could be taken as evidence of the net CYP3A inhibition being relatively moderate with this regimen, as suggested during our discussions. After all, we were investigating 4\(\beta\)-hydroxycholesterol as a marker of CYP3A activity, rather than the extent of CYP3A inhibition due to ritonavir-boosted lopinavir; in, in this setting we failed to detect well-described effects on CYP3A activity, it would be methodologically unsound thereby to infer that these effects may
not be as profound as previously thought. The results from the lopinavir arm had to be explained in a different manner. In this context, I shall quote our argument from the discussion of paper III:

“Therefore we suggest that the absence of the expected decrease in 4β-hydroxycholesterol seen in the lopinavir+ritonavir group may be due to hepatic enzyme induction that cancels the effects of CYP3A inhibition on 4β-hydroxycholesterol formation. Using intravenous midazolam as a probe, Yeh et al demonstrated a 77% decrease in hepatic CYP3A activity following the administration of lopinavir 400 mg b.i.d. + ritonavir 100 mg b.i.d for 14 days [89]. Thus, in order to explain the plasma 4β-hydroxycholesterol data obtained in the lopinavir+ritonavir group using this line of argument, one must postulate either (1) that though CYP3A was the only one of the pathways tested through which 4β-hydroxycholesterol can be formed in vitro [233], this metabolite can be formed by other enzymatic pathways in the presence of hepatic enzyme induction, or (2) that the lopinavir + ritonavir drug regimen has differential effects on the CYP3A-mediated metabolism of midazolam and on the formation of 4β-hydroxycholesterol by the same enzyme”.

At the time I strongly favored the first suggestion, somewhat in disagreement with my co-workers. Today, while I still think it may be correct, I recognize that the extremely long half-life of 4β-hydroxycholesterol (approximately 17 days as measured by Kanebratt et al [234]) compared to that of ritonavir (approximately 4-5 hours [189]) may make the relation between drug interaction PK data with ritonavir, and the effect of this inducer/inhibitor on plasma concentrations of 4β-hydroxycholesterol, very weak. As mentioned above, I owe my present understanding of this to Tomalik-Scharte et al [241]. Against this, though, one might argue that such effect would have had even more impact on the outcome in the atazanavir arm, where ritonavir is dosed once, rather than twice, daily. However, as stated in the paper, the background induction due to the ritonavir-lopinavir regimen is likely to be considerably greater than with ritonavir-atazanavir.

I shall finish the discussion of paper III by some brief remarks on this last mentioned paper [241], which describes the first important dataset on the operating characteristics of 4β-hydroxycholesterol as a biomarker of CYP3A activity produced by other investigators than the group at our department. In this investigation drug effects on midazolam clearance were compared with their effects on 4β-hydroxycholesterol. Though a statistically significant correlation between effects could
be detected in the whole material, the magnitude of this was not impressive, casting doubt on the ability of 4β-hydroxycholesterol to make a quantitatively exact prediction, e.g., of the extent of putative drug interactions. Without going into details about this, I think it is notable that the long half-life of 4β-hydroxycholesterol, which serves to reduce intra-individual and diurnal variability, may in fact compromise its application as a quantitative marker of drug metabolism in the presence of CYP3A inhibition. Perhaps ultimately this biomarker will be best fitted to qualitatively detect significant hepatic enzyme induction, rather than to more precisely determine, e.g., the direct clinical relevance of the metabolic drug interaction potential of an investigational compound. On this last issue, a further problem, apart from the concern about half-life, is the importance of pre-systemic CYP3A for drug interactions due to effects on the activity of this enzyme [243]. Plasma 4β-hydroxycholesterol seems unlikely to adequately reflect this parameter.
10 PAPER IV: TDM OF PI AND NNRTI

Paper IV reports the PK/PD sub-study of NORTHIV (this randomized controlled trial has been previously described). Since very much has already been said about TDM and the problems of defining the MEC, and furthermore, as the literature on concentration-effect relations for efavirenz, atazanavir and lopinavir in treatment-naïve subjects is rather extensively reviewed in paper IV itself, included in this volume, I shall be rather brief in my discussion of this paper, and instead focus on some methodological issues.

10.1 Notes on the PK sampling strategy and trough imputation

In retrospect, it is difficult to avoid the conclusion that the sampling schedule (trough sampling prescribed at week 4 and at week 48, as well as optionally at week 12) for the PK/PD substudy of NORTHIV is problematic. Week 48 samples were obviously only available for patients that were still in the study, whereas for those failing in the meantime, in most cases we had to rely on a single measurement of the drug concentration at week 4.

While a traditional approach was therefore difficult, my understanding is that the sampling strategy in NORTHIV, without any intense PK/PD subgroup, and with its single measure samples prescribed to be taken as troughs (though in many cases this was not done, as is often the case as the reality of the clinic imposes itself on the ideal world of TDM), was also inherently unsuitable for a population pharmacokinetic approach [244]. This impression I understood to be confirmed when I discussed the issue with the PK assessors at the Swedish Medical Products Agency (a present workplace of mine). Yet I am by no means expert in modeling, of which I have virtually no experience, and thus humbly open to the suggestion that I might have misjudged the situation. At any rate, when necessary in our study (as well as in our TDM practice), we imputed the trough concentration under the assumption of a uniform elimination rate constant for all patients, based on the average values for the elimination half-life given by the manufacturers of the respective drugs, a procedure that had been used previously in similar studies (e.g. [245, 246]).
10.2 TDM practice as an obstacle to its study

Furthermore and importantly, the fact that drug concentrations were reported to the treating physicians, and in some cases prompted dose alterations, may be a major obstacle to the interpretation of the study outcome. Indeed, when I look at this paper, there emerges the full image of a spectre invoked already by Van Heeswijk in the aforementioned position paper from 2002 [101]. Our \textit{de facto} implementation of, more or less, routine TDM had forced us to make assumptions that we then came to consider as part of the knowledge base, and which would distort our ability to make any progress on the issues we considered most closely at hand. Therefore I believe that some defects of both paper II and paper IV can be explained by the fact that there was a relatively widespread opinion (which I shared for a long time) that TDM data were integral to optimising HIV therapy.

10.3 The case of efavirenz related CNS side effects

I shall not spend many words discussing the outcome of this study that perhaps, with some goodwill, implied a relation between plasma bilirubin and atazanavir-based treatment outcome (though the cut-off value implied by our exploratory analysis is different from those previously suggested, as mentioned in the paper), but decidedly did not demonstrate what it, in fact, was incapable of showing: a concentration-effect relation for efavirenz, lopinavir and atazanavir in combination with 2 NRTI, in treatment-naïve patients. The reason for this is not only methodological (though these reasons may have been sufficient for it to fail), but, more importantly, that the doses used might be reasonably close to whatever would be optimal in this population (the argument is elaborated in paper IV).

Neither shall I dwell very long on my interpretation of likely MECs for these drugs in treatment naïve patients with fully active dual NRTI support, because it is evident that I, like everyone else, believe this to be, by all means, substantially lower than what is (but perhaps exceptionally) seen in patients that are adherent.

Instead I shall consider the subset of patients with CNS side effects of efavirenz. This has classically been supposed to be concentration dependent [147, 151], though as related in paper IV, several investigators have not been able to confirm this (e.g.,[247, 248]). However, as also previously discussed, there is evidence that the relation between efavirenz concentrations and the risk or severity of this sort of side effects might not be very easy to grasp. Indeed, Haas et al demonstrated what, no
doubt, most clinicians already knew - that this effect displays a great tendency for development of tolerance [147]. Since the time at which we wrote and referenced paper IV, two notable papers have appeared on the relation between efavirenz concentration and CNS side effects.

I will finish this thesis with a few words about these, not least because of the somewhat curious incident that, despite the fact that these papers seem to carry contrary implications for the question of whether it might be worthwhile or not to use TDM in patients with CNS side effects, they are first-authored by the same person. The title of the first [249] is: “Absence of a relation between efavirenz plasma concentration and toxicity driven efavirenz discontinuations in the EuroSIDA cohort”.

This reports how patients in the EuroSIDA cohort with an efavirenz plasma concentration available were divided into those that discontinued efavirenz because of any toxicity and those that continued efavirenz for more than two years. Drug concentrations did not predict discontinuations due to side effects. The frequency of drug concentrations above the generally accepted upper limit of the therapeutic interval for efavirenz (4000 ng/mL, deriving originally from Marzolini et al [151]) was similar among those that discontinued and those that did not. This leads the authors to question the designation of efavirenz concentrations above this level as being “toxic”. It is also noted that positive hepatitis C status was an independent predictor of toxicity driven efavirenz discontinuations. This is indicative that patient-related factors, rather than efavirenz drug levels, may be the main determinant of clinically relevant efavirenz toxicity beyond the, say, first month of treatment.

The other paper on efavirenz toxicity by the same author, published the same year in Journal of Acquired Immune Deficiency Syndromes [150] has the following title: “Efavirenz Dose Reduction is Safe in Patients With High Plasma Concentrations and May Prevent Efavirenz Discontinuations.” In this observational study, patients with efavirenz concentrations above 4000 ng/mL were selected from the Dutch ATHENA cohort. These patients were classified into two groups - patients who had undergone dose reduction after the high plasma concentration measurement and those that did not. The dose reductions resulted in a 41% decrease in the median efavirenz concentration, and were associated with a significantly reduced discontinuation rate due to toxicity.

So what are we to believe? This discussion of my scientific work is already too long, so I shall not go on to analyse these observational studies in detail. Instead I leave the topic with this particular example of the difficulties in making practical sense of the PK/PD relations of antiretroviral agents.
11 CONCLUSIONS

My main conclusions of these studies (inferences greatly informed by the investigations of other workers on the same and similar topics that have been reported during the period in question) are the following:

- Genetic variability in CYP3A5 is unlikely to have any measurable clinical impact on present day boosted PI regimens.
- Though boosted PI maintenance monotherapy may be an option in selected patients, there are important concerns about, e.g., CNS penetration. Increasing PI doses beyond those presently recommended is unlikely to alter the efficacy of boosted PI monotherapy.
- Though no general target level can be suggested (e.g., due to the inter-individual variability of the atazanavir effect on bilirubin conjugation), continuous monitoring of plasma bilirubin is useful in patients treated with atazanavir, as a return to baseline in a patient with a previous increase after atazanavir initiation, may indicate lapsing adherence.
- Routine TDM is not useful in antiretroviral therapy, mainly due to the high efficacy and wide therapeutic intervals of present day antiretroviral drugs at recommended doses.
12 ACKNOWLEDGEMENTS

There are many persons that ought to be acknowledged for their role in the making of this thesis. I shall name some of these, and hope that those not named (that perhaps should have been) will not feel slighted.

First I want to express my deepest gratitude to professor Anders Sönnerborg. Anders has been my sensei when trying to make sense of HIV and the somewhat particular characteristics of antiretroviral therapy. It goes without saying that his influence is great on the research here reported, at times directly, but also indirectly insofar as my own input has informed the conduct and reporting of these studies.

Associate professor Ylva Böttiger has not only been my main supervisor and guide through this project, which may very well not have been completed without her support. She has also been of great importance for the development of my thoughts around the advantages and limitations of TDM, as well as on how to communicate and make clinically useful drug concentration data.

Professor Leif Bertilsson is a scientist truly working within what Karl Popper called “the context of discovery”. When struggling to confirm one’s hypotheses it is all too easy to somehow disregard outlier observations (or even to consider them annoying obstacles to statistical analysis). Leif taught me that such observations are in fact gateways to scientific discovery. Though I cannot claim to have made much use of this knowledge, I am very thankful to Leif for showing me a great mechanism of scientific progress.

Professor Magnus Gisslén is the principal investigator of NORTHIV. This thesis would have been very different without him. At all times he has generously shared his extensive knowledge and understanding of antiretroviral therapy.

Associate professor Lars Ståhle introduced me to the area of this research, and has taught me much of what I know about the fundamental principles of pharmacology. He is the father of Swedish HIV TDM, and another person without whom this work might not have been.

My understanding of pharmacokinetics, statistics and the formal aspects of science has to a large extent been developed in discussion with doctor Jonatan Lindh. His input has been of the greatest importance for my ideas on how science ought to be conducted and reported.
Doctor Olle Karlström was not only the principal investigator of the study reported here as paper II; he is also a person with whom I have had a great many fruitful discussions, and whose influence has been considerable on the views I have reported in this thesis.

I also acknowledge the work and influence of the co-authors not mentioned here, of the HIV clinicians at the Stockholm HIV policlinics at Karolinska University Hospital, Huddinge and Solna, and Venhälsan at Södersjukhuset, of my colleagues at the Department of Clinical Pharmacology, Karolinska University Hospital, and the efforts of all those responsible for the patient care, sampling and laboratory analysis necessary for the realization of these studies.
13 REFERENCES


156. Läkemedelsindustriföreningen, FASS. Available at: http://www.fass.se/LIF/home/index.jsp?UserTypeID=0. 2009.


158. Mayers, D., et al., *A double-blind placebo-controlled study to assess the safety, tolerability and antiretroviral activity of efavirenz (DMP-266) in combination with open-label zidovudine and lamivudine in HIV-1 infected patients*. The


http://www.hiv-pharmacogenomics.org/.


212. Arribas, J., et al., The MONET trial: darunavir/ritonavir monotherapy shows non-inferior efficacy to standard HAART, for patients with HIV RNA <50 copies/mL at baseline, in Abstract TuAB 106. 5th IAS conference on HIV Pathogenesis, Treatment and Prevention. 2009: Cape Town, South Africa.


217. CDER Guidance for industry: antiretroviral drugs using plasma HIV-RNA measurements – clinical considerations for accelerated and traditional approval. Available at:


