Clinical data based optimal STI strategies for HIV: a reinforcement learning approach

Damien Ernst

Department of Electrical Engineering and Computer Science
University of Liège

Montefiore - March 9, 2006

Presentation based on the paper: "Clinical data based optimal STI strategies for HIV: a reinforcement leanring approach". D. Ernst, G.B. Stan, J. Gonçalves and L. Wehenkel

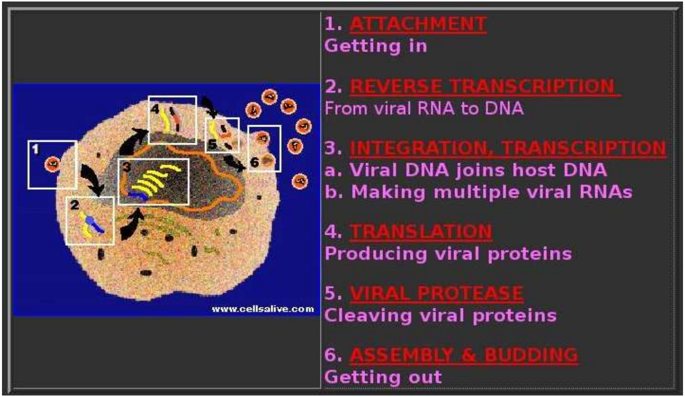
HIV

- *Human Immunodeficiency Virus (HIV) is a retrovirus at the source of the Acquired Immune Defficiency Syndrome (AIDS)*
- HIV particles target cells of the immune system (mostly CD4$^+$ lymphocytes and macrophages)
- Inclusion of HIV particles in immune cells lead to massive production of new viral particles, death of the infected cells and, ultimately, devastation of the immune system

Current anti-HIV drugs

Two main categories:

1. Reverse Transcriptaese Inhibitors (RTI)
2. Protease Inhibitor (PI)



1. ATTACHMENT
Getting in

2. REVERSE TRANSCRIPTION
From viral RNA to DNA

3. INTEGRATION, TRANSCRIPTION
a. Viral DNA joins host DNA
b. Making multiple viral RNAs

4. TRANSLATION
Producing viral proteins

5. VIRAL PROTEASE
Cleaving viral proteins

6. ASSEMBLY & BUDDING
Getting out

Figure: Taken from http://www.cellsalive.com/hiv0.htm

Treatments for infected patients

- Highly Active Anti-Retroviral Therapy (HAART): combination of two or more drugs. Usually one or more RTIs in combinations with a PI.
- Two main concerns about the long-term used of anti retroviral drugs: undesirable side effects (leading to poor compliance) and mutation of the virus (need to change drugs or even inability to find appropriate pharmaceutical treatments).
- Need for efficient drug scheduling strategies.
- Idealistically, a drug-scheduling strategy should bring the system to a state where the immune system has control over the virus (with low amount of drugs and low systemic effects).

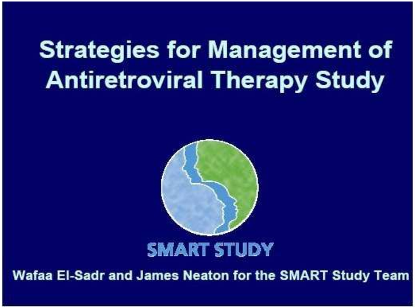
Structured Treatment Interruption (STI)

- STI: to cycle the patient on and off drug therapy
- STI strategies often well received by patients since they offer them period of relief from treatment
- In some remarkable cases, STI strategies have enabled the patients to maintain immune control over the virus in the absence of treatment

Goal of this research: to compute optimal STI strategies

STI: A glimpse at today's practice

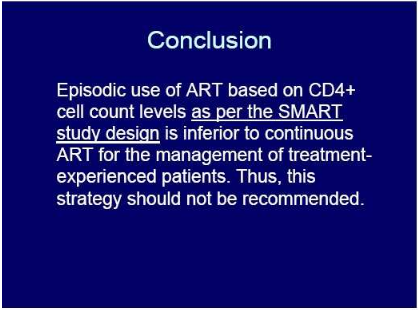
If CD4+ cell count falls below a certain threshold, put the patient on drugs. Otherwise put him off. This practice has met some problems:



Strategies for Management of Antiretroviral Therapy Study

SMART STUDY

Wafaa El-Sadr and James Neaton for the SMART Study Team



Conclusion

Episodic use of ART based on CD4+ cell count levels as per the SMART study design is inferior to continuous ART for the management of treatment-experienced patients. Thus, this strategy should not be recommended.

Figure: Taken from http://www.cpcra.org/docs/pubs/2006/croi2006-smart.pdf

More advanced techniques (not clinically tested)

- Some authors have proposed to design STI treatments by exploiting mathematical models of the HIV infection.
- Models are under the form of a set of Ordinary Differential Equations (ODEs)
- Deduction of STI strategies is done by using methods from the control theory.

But modelling of the HIV dynamics is a difficult task. Indeed, one has

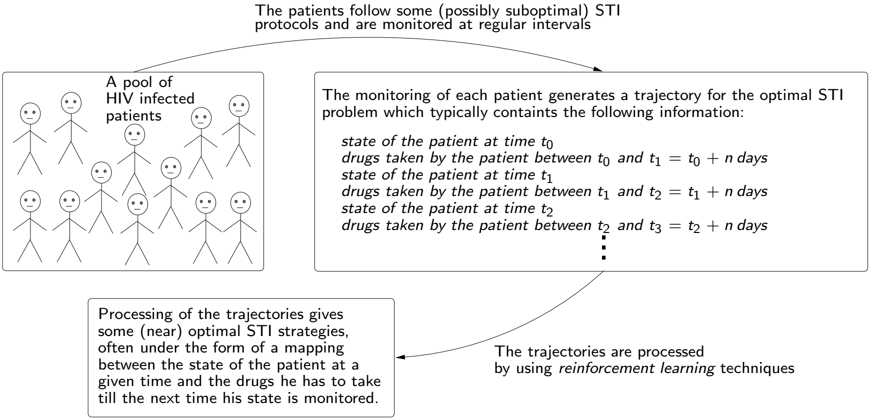
- to select the right parametric system of ODEs
- to fit the parameters to reflect quantitatively biological observations

An interesting alternative

- Infer directly from clinical data good STI strategies, without modelling the HIV infection dynamics.
- Clinical data: time evolution of patient's state (CD4$^+$ T cell count, systemic costs of the drugs, etc) recorded at discrete-time instant and sequence of drugs administered.
- Clinical data can be seen as trajectories of the immune system responding to treatment.

Inferring policies from trajectories

- Problem of inferring from trajectories appropriate control policy has been studied in control theory and computer science.
- One way to approach it: state an optimality criterion and search for strategies optimizing this criterion.
- Classical approach: infer a model and derive from it and the optimality criterion an optimal strategy.
- Reinforcement learning approach: compute optimal strategies directly from the trajectory, without identifying a model.



The patients follow some (possibly suboptimal) STI protocols and are monitored at regular intervals

A pool of HIV infected patients

The monitoring of each patient generates a trajectory for the optimal STI problem which typically containts the following information:

state of the patient at time t_0
drugs taken by the patient between t_0 *and* $t_1 = t_0 + n$ *days*
state of the patient at time t_1
drugs taken by the patient between t_1 *and* $t_2 = t_1 + n$ *days*
state of the patient at time t_2
drugs taken by the patient between t_2 *and* $t_3 = t_2 + n$ *days*
$\vdots$

The trajectories are processed by using *reinforcement learning* techniques

Processing of the trajectories gives some (near) optimal STI strategies, often under the form of a mapping between the state of the patient at a given time and the drugs he has to take till the next time his state is monitored.

Figure: Determination of optimal STI strategies from clinical data by using reinforcement learning algorithms: the overall principle.

Learning from a sample of trajectories: the RL approach

Problem formulation

Discrete-time dynamics:

$$x_{t+1} = f(x_t, u_t) \quad t = 0, 1, \ldots$$

where $x_t \in X$ and $u_t \in U$.

Cost function: $c(x, u) : X \times U \to \mathbf{R}$. $c(x, u)$ bounded by B_c.

Discounted infinite horizon cost associated to stationary policy $\mu : X \to U$: $J^{\mu}(x) = \lim_{N \to \infty} \sum_{t=0}^{N-1} \gamma^t c(x_t, \mu(x_t))$

Optimal stationary policy μ^* : Policy that minimizes J^{μ} for all x.

Objective: Find an optimal policy μ^*.

We do not know: The discrete-time dynamics.

We know instead: A set of trajectories $(x_0, u_0, x_1, \cdots, u_{T-1}, x_T)$.

Some dynamic programming results

Sequence of functions Q_N: $X \times U \rightarrow \mathbb{R}$

$$Q_N(x,u) = c(x,u) + \gamma \min_{u' \in U} Q_{N-1}(f(x,u),u'), \quad \forall N > 1$$

with $Q_1(x,u) \equiv c(x,u)$, converges to the Q-function, unique solution of the Bellman equation:

$$Q(x,u) = c(x,u) + \gamma \min_{u' \in U} Q(f(x,u),u').$$

Necessary and sufficient optimality condition:

$$\mu^*(x) \in \arg\min_{u \in U} Q(x,u)$$

Stationary policy μ_N^*:

$$\mu_N^*(x) \in \arg\min_{u \in U} Q_N(x,u).$$

Bound on the suboptimality of μ_N^*:

$$J^{\mu_N^*} - J^{\mu^*} \le \frac{2\gamma^N B_c}{(1-\gamma)^2}.$$

Fitted Q iteration

Trajectories $(x_0, u_0, x_1, \cdots, u_{T-1}, x_T)$ transformed into a set of one-step system transitions $\mathcal{F} = \{(x_t^l, u_t^l, x_{t+1}^l)\}_{l=1}^{\#\mathcal{F}}$.

Fitted Q iteration computes from $\mathcal{F}$ the functions $\hat{Q}_1$, $\hat{Q}_2$, ..., $\hat{Q}_N$, approximations of Q_1, Q_2, ..., Q_N.

Computation done iteratively by solving a sequence of standard supervised learning (SL) problems. Training sample for the k^{th} $(k \geq 2)$ problem is

$$\left\{\left((x_t^l, u_t^l),\ c(x_t^l, u_t^l) + \gamma \min_{u \in U} \hat{Q}_{k-1}(x_{t+1}^l, u)\right)\right\}_{l=1}^{\#\mathcal{F}} \text{ with}$$

$\hat{Q}_1(x, u) \equiv c(x, u)$. From the k^{th} training sample, the supervised learning algorithm outputs $\hat{Q}_k$.

$\hat{\mu}_N^*(x) \in \arg\min_{u \in U} \hat{Q}_N(x, u)$ is taken as approximation of $\mu^*(x)$.

In our simulations, SL method used is an ensemble of regression trees method named Extra-Trees.

Illustration

- We present results we have obtained by using the RL-based approach on artificially generated data.
- The example is directly inspired from
B.M. Adams, H.T. Banks, Hee-Dae Kwon and H.T. Tran. (2004). "Dynamic multidrug therapies for HIV: Optimal and STI Control Approaches". *Mathematical Biosciences and Engineering*, 1, 223-241.

Illustration: Kinds of STI strategies targeted

Bi-therapy treatments combining a fixed RTI and a fixed PI.

Revise drug administration every five days based on clinical measurements.

Four possible on-off combinations for the next five days: RTI and PI on, only RTI on, only STI on, RTI and PI off

We seek STI strategies that minimize J^{μ}.

Instantaneous cost at time t:

$$c(x_t, u_t) = 0.1V_t + 20000\epsilon_{1_t}^2 + 2000\epsilon_{2_t}^2 - 1000E_t$$

$\epsilon_{1_t} = 0.7$ (resp. $\epsilon_{1_t} = 0$) if the RTI is cycled on (resp. off) at t

$\epsilon_{2_t} = 0.3$ (resp. $\epsilon_{2_t} = 0$) if the PI is cycled on (resp. off) at time t

V: number of free HI viruses

E: number of cytotoxic T-lymphocytes

Decay factor γ: chosen equal to 0.98.

Illustration: A mathematical model as substitute for real-life patients

$$
\begin{aligned}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - (1-\epsilon_1)k_1 V T_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - (1-f\epsilon_1)k_2 V T_2 \\
\dot{T}_1^* &= (1-\epsilon_1)k_1 V T_1 - \delta T_1^* - m_1 E T_1^* \\
\dot{T}_2^* &= (1-f\epsilon_1)k_2 V T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V} &= (1-\epsilon_2)N_T \delta(T_1^* + T_2^*) - cV - [(1-\epsilon_1)\rho_1 k_1 T_1 + (1-f\epsilon_1)\rho_2 k_2 T_2]V \\
\dot{E} &= \lambda_E + \frac{b_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b}E - \frac{d_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d}E - \delta_E E
\end{aligned}
$$

T_1 (T_1^*) = number of non-infected (infected) CD4$^+$ lymphocytes
T_2 (T_2^*) = non-infected (infected) macrophages
V = number of free HI viruses
E = number of cytotoxic T-lymphocytes.

ϵ_1 and ϵ_2 = control actions corresponding to RTI and the PI.

Period during which the RTI (resp. the PI) is administrated to the patient: ϵ_1 (resp. ϵ_2) is set equal to 0.7 (resp. 0.3).

RTI (resp. the PI) not administrated: $\epsilon_1 = 0$ (resp. $\epsilon_2 = 0$).

Illustration: Some insight into this model

In absence of treatment, three physical equilibrium points:

1. uninfected state:

$$(T_1, T_2, T_1^*, T_2^*, V, E) = (10^6, 3198, 0, 0, 0, 10)$$

2. "healthy" locally stable equilibrium

$$(T_1, T_2, T_1^*, T_2^*, V, E) = (967839, 621, 76, 6, 415, 353108)$$

(small viral load, a high CD4$^+$ T-lymphocytes count, high HIV-specific cytotoxic T-cells count)

3. "non-healthy" locally stable equilibrium point

$$(T_1, T_2, T_1^*, T_2^*, V, E) = (163573, 5, 11945, 46, 63919, 24)$$

(T-cells depleted, viral load very high).

Illustration: Protocol for artificially generating the clinical data

Monitoring of patients: every five days during 1000 days.
Medication: can be revised every five days based on the information generated by the monitoring.
Iterative generation of the clinical data (ten iterations):

- First iteration. Thirty patients in "non-healthy" steady-state. Physiological data (T_1, T_2, T_1^*, T_2^*, V, E) recorded and a new type of medication randomly selected in U every five days. Monitoring of each patient generates a trajectory $(x_0, u_0, x_1, \cdots, x_{199}, u_{199}, x_{200})$.
- Second iteration. Only difference with first iteration: medication determined by the following STI strategy: in 85% of the cases, use strategy $\hat{\mu}^*_{400}$ computed by fitted Q iteration on previously generated trajectories; in the remaining 15% medication randomly selected in U.
- Third-tenth iteration: idem as second iteration.

Illustration: Simulation results

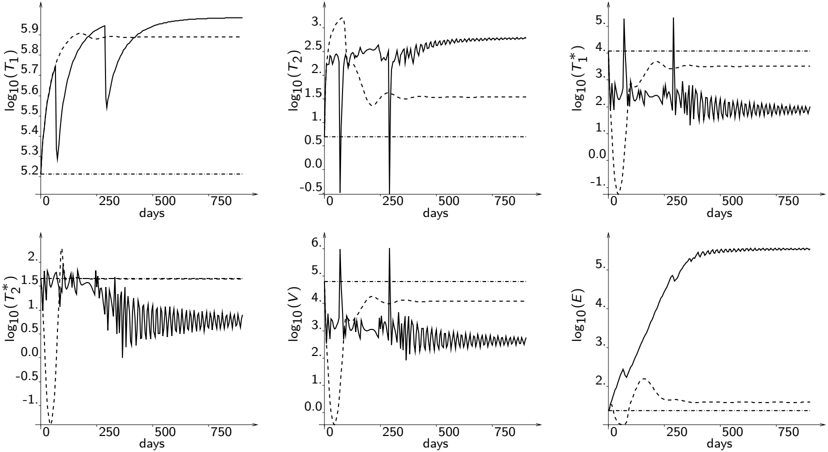


Figure: Solid curve ($-$) corresponds = patient which follows STI strategies; dashed curves ($- -$) = no interruption in the treatment; dotted curves ($- \cdot$) = no treatment

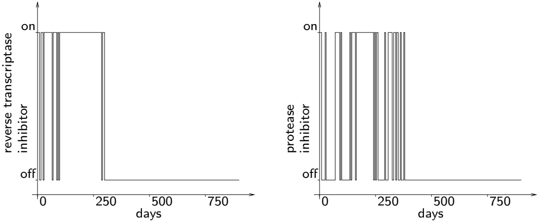


Figure: STI treatment for a patient treated from early stage of infection. Clinical data generated by 300 patients.

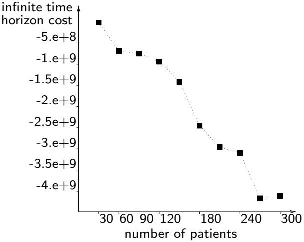


Figure: Influence of the number of patients on the infinite time horizon cost corresponding to the computed STI strategies.

From numerically simulated data to real-life patients

We expect to face four main difficulties:

- The HIV/immune system dynamics may be different from one patient to the other.
- Difficulty to state properly the optimal control problem
- Partial observability
- Corrupted measurements

Conclusions

- Reinforcement learning algorithms seem to be promising tools to extract from clinical data, good STI strategies.
- Lot of work is however still needed !!!
- But 40 millions of people are living with HIV/AIDS. Isn't it a good reason to keep working hard ?

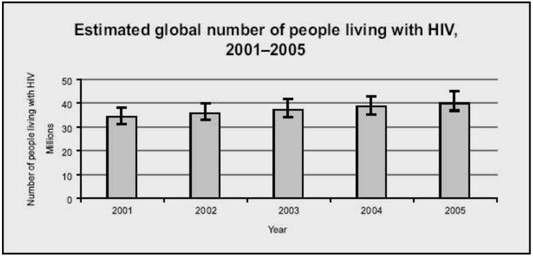


Figure: Taken from UNAIDS. AIDS epidemic update: December 2005. "UNAIDS/05.19E"