ORIGINAL RESEARCH

Predicting smoking cessation and its relapse in HIV-infected patients: the Swiss HIV Cohort Study

J Schäfer,[1] J Young,[1] E Bernasconi,[2] B Ledergerber,[3] D Nicca,[4] A Calmy,[5] M Cavassini,[6] H Furrer,[7] M Battegay,[8] HC Bucher[1,8] and the Swiss HIV Cohort Study*

[1]*Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland,* [2]*Division of Infectious Diseases, Regional Hospital of Lugano, Lugano, Switzerland,* [3]*Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland,* [4]*Division of Infectious Diseases and Hospital Epidemiology, Cantonal Hospital St Gallen, St Gallen, Switzerland,* [5]*Division of Infectious Diseases, University Hospital Geneva, Geneva, Switzerland,* [6]*Division of Infectious Diseases, University Hospital Lausanne, Lausanne, Switzerland,* [7]*Department of Infectious Diseases, Bern University Hospital and University of Bern, Bern, Switzerland and* [8]*Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland*

Objectives

The aim of the study was to assess whether prospective follow-up data within the Swiss HIV Cohort Study can be used to predict patients who stop smoking; or among smokers who stop, those who start smoking again.

Methods

We built prediction models first using clinical reasoning ('clinical models') and then by selecting from numerous candidate predictors using advanced statistical methods ('statistical models'). Our clinical models were based on literature that suggests that motivation drives smoking cessation, while dependence drives relapse in those attempting to stop. Our statistical models were based on automatic variable selection using additive logistic regression with component-wise gradient boosting.

Results

Of 4833 smokers, 26% stopped smoking, at least temporarily; because among those who stopped, 48% started smoking again. The predictive performance of our clinical and statistical models was modest. A basic clinical model for cessation, with patients classified into three motivational groups, was nearly as discriminatory as a constrained statistical model with just the most important predictors (the ratio of nonsmoking visits to total visits, alcohol or drug dependence, psychiatric comorbidities, recent hospitalization and age). A basic clinical model for relapse, based on the maximum number of cigarettes per day prior to stopping, was not as discriminatory as a constrained statistical model with just the ratio of nonsmoking visits to total visits.

Conclusions

Predicting smoking cessation and relapse is difficult, so that simple models are nearly as discriminatory as complex ones. Patients with a history of attempting to stop and those known to have stopped recently are the best candidates for an intervention.

Keywords: HIV, relapse, smoking cessation

Accepted 1 April 2014

Correspondence: Dr Juliane Schäfer, Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, CH-4031 Basel, Switzerland.
Tel: +41 61 265 31 19; fax: +41 61 265 31 09; e-mail: juliane.schaefer@usb.ch

*See Appendix 1.

Introduction

Smoking is common among HIV-infected patients. It is also the most common modifiable risk factor for cardiovascular disease (CVD) within this population [1], with lower rates of CVD if patients stop [2]. Despite the strong association between smoking and both CVD and cancer [3–5], few cessation programmes have been established to help HIV-infected patients stop smoking [6,7]. Clinicians treating HIV-infected patients typically have other priorities, including maintaining viral suppression, managing antiretroviral therapy-related adverse effects and addressing either psychosocial problems or addiction to even more harmful substances. However, smoking cessation programmes have proved worthwhile in the management of HIV-infected patients, increasing cessation rates and decreasing relapse rates [8,9].

Many smoking cessation programmes are based on the transtheoretical model of behaviour change [10]. According to this model [11], individuals move from pre-contemplation of a change in behaviour to contemplation, preparation, action and then finally to maintenance after the change. In the pre-contemplation stage, smokers do not intend to stop (at least not within the next 6 months), while those in the contemplation stage think seriously about stopping in the next 6 months. There, smokers weigh up the pros and cons of cessation and are open to information [11,12]. Before stopping for good, smokers may cycle through the contemplation to action stages many times. While the value of this model is debated [13], literature suggests that motivation drives an attempt to stop, while dependence drives relapse when making an attempt [14].

The aim of this study was to assess whether information routinely collected at 6-monthly follow-up visits within the Swiss HIV Cohort Study (SHCS) could be used to predict (1) the smokers most likely to stop and (2) among smokers who stop, those most likely to relapse. These predictions could be included in a patient's cardiovascular risk profile (see fig. 1 in [15]) to encourage intervention by the clinician.

Methods

Patients

The SHCS is a multicentre, prospective, observational cohort study with continuing enrolment of HIV-infected adults and routine clinical follow-up scheduled every 6 months [16]. Routine collection of self-reported smoking status (yes or no) and of the self-reported number of cigarettes smoked per day began in April 2000. In this study, we analysed the data as at January 2012.

When predicting cessation, our population of interest was all patients reporting smoking at three follow-up visits (these reports did not need to be consecutive) and with at least two additional follow-up visits. With cessation defined as two consecutive reports of not smoking, patients were divided into two groups: patients who stopped smoking (Table 1, example patient 1) and those who did not (Table 1, example patients 2 and 3). If smoking status was missing at any follow-up visit following three reports of smoking, patients were assumed to still smoke – cessation had to be explicitly reported.

When predicting relapse, patients who stopped smoking and who had at least one additional follow-up visit were further subdivided into patients known to relapse (Table 1, example patient 4) and those not known to relapse (Table 1, example patients 5 and 6). If smoking status was missing at any follow-up visit after a patient stopped, patients were assumed to have not started smoking again – relapse had to be explicitly reported.

For both cessation and relapse, we used only the first of possibly multiple cessation and relapse episodes per patient.

Predictors of cessation and relapse

Literature suggests that motivation drives smoking cessation, while dependence drives relapse in those attempting to stop [14]. We assumed that motivation to stop would be related to recent events in patients' lives and dependence would be related to recent patient reports of the number of cigarettes smoked per day.

For cessation
For cessation, patients were classified into highly, poorly and typically motivated patients (with the latter being those not in one of the other two groups). The highly motivated group comprised patients with: (1) a previous cessation attempt (defined as any nonsmoking visit since the patient first reported smoking); (2) a CVD event within the last 6 months; (3) an AIDS-defining disease within the last 6 months; or (4) a pregnancy within the last 9 months. The poorly motivated group comprised patients with: (1) alcohol or drug dependence; (2) psychiatric comorbidities; or (3) a stressful life event within the last 6 months (death of a spouse, divorce, marital separation or imprisonment – the top four stressful life events on the Holmes and Rahe stress scale [17]).

In a sensitivity analysis, the highly motivated group was extended to include patients receiving lipid- or blood pressure-lowering medication for the first time within the last 6 months, and patients hospitalized for any reason during the last 6 months.

Table 1 Definitions for study populations and index dates. A patient became a confirmed smoker (x_S) after reporting smoking at three follow-up visits (these reports did not need to be consecutive). Smoking cessation was defined as two consecutive reports of not smoking. All confirmed smokers with at least two additional follow-up visits were assessed for cessation (x_C). All smokers who stopped with at least one additional follow-up visit were assessed for relapse ($\checkmark_R$).

Example patient follow-up: x = report of smoking; ✓ = report of not smoking

Visit		1	2	3	4	5	6	7	8	9	10	11	12	Outcome	Rationale
Patients assessed for cessation	1	x	✓	x	x_S	✓	x_C	✓	✓					Stops – the index date is visit 6, the last time the patient was still smoking prior to stopping	At this point, efforts to encourage patients to stop would be likely to have most effect as clearly patients are ready to do so
	2	x	✓	x	x_S	x_C	x							Did not stop – the index date is visit 5, the latest smoking report when the patient still had not demonstrated sufficient self-motivation to stop	This should achieve a conservative comparison between patients who stopped and those who did not, because patients who did not stop were compared at a point when their age and probably their FRS were highest
	3	x	✓	x	$x_{S,C}$	x	✓							Did not stop – the index date is visit 4, the latest smoking report when the patient still had not demonstrated sufficient self-motivation to stop. Since the last visit is a first (but unsubstantiated) report of not smoking, the index date is the visit before the penultimate visit	This should achieve a conservative comparison between patients who stopped and those who did not, because patients who did not stop were compared at a point when their age and probably their FRS were highest
Patients assessed for relapse	4	x	✓	x	x_S	✓	x_C	✓	✓	$\checkmark_R$	x			Known to relapse – the index date is visit 9, the last time the patient had still stopped smoking prior to restarting	At this point, patients could be encouraged to not start smoking again
	5	x	✓	x	x_S	✓	x_C	✓	$\checkmark_R$	✓	✓	✓	✓	Not known to relapse – the index date is visit 8, the second of the two consecutive reports of not smoking following the last time the patient was still smoking prior to stopping	This should achieve a conservative comparison because patients not known to relapse were compared when they stopped – at that point their age and probably their FRS were lowest
	6	x	✓	x	x_S	✓	x_C	✓	$\checkmark_R$	✓				Not known to relapse – the index date is visit 8, the second of the two consecutive reports of not smoking following the last time the patient was still smoking prior to stopping	This should achieve a conservative comparison because patients not known to relapse were compared when they stopped – at that point their age and probably their FRS were lowest

6-monthly follow-up visits within the SHCS →

FRS, Framingham risk score; SHCS, Swiss HIV Cohort Study.

Additional candidate predictors of cessation were age and gender, the Framingham risk score (FRS) and the change in the reported number of cigarettes per day between the last two follow-up visits [18]. The FRS (see tables III.1-5 and III.1-6 in [19]) is a gender-specific measure of 10-year cardiovascular risk [20]. We also considered the ratio of nonsmoking visits to the total number of follow-up visits as a candidate predictor. This ratio may carry more information than simply recording whether a patient has had a previous cessation attempt.

Perceived health risks are higher among smokers who stop than those still smoking [21,22], and there is evidence that the propensity to stop is age-related [23,24], and therefore potentially related to increasing FRS, as age is a component of this score [8]. For patients who stopped smoking, we took the visit prior to stopping as the index date because, at this point, efforts to encourage patients to stop would be likely to have most effect, as clearly patients were ready to do so. For patients who did not stop, we took the penultimate visit as the index date as at this point patients still had not demonstrated sufficient motivation to stop. This should achieve a conservative comparison, because patients who did not stop were compared at a point when their age and probably their FRS were highest.

For relapse

For relapse, we used the maximum number of cigarettes per day reported at the last three follow-up visits prior to stopping as a measure of the degree of dependence. Additional candidate predictors of relapse were the three motivational groups, age and gender.

Assuming again that the propensity to stop is age-related [23,24], and therefore potentially related to increasing FRS [8], for patients known to relapse we took the visit prior to a first report of restarting smoking as the index date because, at this point, patients could be encouraged to not start smoking again. For patients not known to relapse, we took the second of the two consecutive reports of not smoking as the index date. This should again achieve a conservative comparison, because patients not known to relapse were compared when they stopped – at that point their age and probably their FRS were lowest.

We also considered the ratio of nonsmoking visits to the total number of follow-up visits and weight gain (defined as weight gain within the last 6 months) as candidate predictors of relapse. For these two variables, we changed the index date for patients not known to relapse, using the penultimate visit instead of the date patients were known to have stopped. This change in index date allows the ratio of nonsmoking visits to total visits to increase for patients not known to relapse. And because most cessation-related weight gain occurs within a few months of stopping [25],

the date patients were known to have stopped is not a suitable index date for patients not known to relapse.

Statistical analyses

Our aim was to build simple prediction models consistent with literature that suggests that motivation drives smoking cessation, while dependence drives relapse in those attempting to stop [14]. We built models first using clinical reasoning ('clinical models') and then by selecting from a large number of candidate predictors using advanced statistical methods ('statistical models').

Clinical models

For cessation, we first fitted a logistic regression model with only the three motivational groups as predictors ('basic clinical model'). We then assessed whether any additional model complexity achieved by adding either the FRS or age and gender, or the change in the number of cigarettes per day (reported at the last two follow-up visits) was justifiable in terms of the Akaike information criterion (AIC) ('extended clinical model'). The AIC is a measure of goodness of fit with a penalty for model complexity [26].

For relapse, the basic clinical model was a linear function of the maximum number of cigarettes per day reported at the last three follow-up visits prior to stopping. Again, we then assessed whether we could improve on this model (in terms of the AIC) through adding either the three motivational groups, or age and gender, or weight gain ('extended clinical model').

Statistical models

Our statistical models were based on automatic variable selection and model choice using additive logistic regression with component-wise gradient boosting [27]. The key aspects of this approach are: (1) selecting a subset of variables to include in the model, and (2) flexibility in representing selected variables so that both linear and nonlinear components are used to represent continuous predictors where appropriate [28]. Early stopping of the iterative boosting algorithm promotes a parsimonious model, with only a few variables or model components selected, and with their regression coefficients 'shrunk' towards zero to improve predictive performance [29,30]. Here, the model with the optimal number of iterations for predictive performance ('optimal statistical model') was then further constrained to reduce model complexity ('constrained statistical model'); see Appendix 2.

For cessation, we included as candidate predictors: the individual variables used to classify patients into motivational groups (except with a previous cessation attempt replaced by the ratio of nonsmoking visits to total visits),

Table 2 Characteristics of (1) patients who stopped and those who did not, and (2) patients known to relapse and those not known to relapse

Characteristic	Cessation		Relapse	
	Yes (n = 1261)	No (n = 3572)	Yes (n = 557)	No (n = 610)
Female gender (%)	26	28	26	28
Age (years) [median (IQR)]	43 (37, 49)	45 (40, 50)	44 (40, 50)	44 (38, 51)
Previous cessation attempt (%)	26	19	24	27
Ratio of nonsmoking visits to total follow-up visits [median (IQR)]	0 (0, 0.06)	0 (0, 0)	0.33 (0.23, 0.50)	0.51 (0.33, 0.70)
Cardiovascular event[1] (%)	1	0	1	0
CDC category C event[1] (%)	0	1	1	1
Pregnancy[2] (%)	1	0	0	2
Lipid- or blood pressure-lowering medication[3] (%)	2	2	1	2
Hospitalization[1] (%)	7	12	10	11
Alcohol or drug dependence[4] (%)	5	16	7	4
Psychiatric comorbidities[5] (%)	10	24	15	11
Stressful life event[6] (%)	9	12	9	7
Motivational group (%)				
Poorly motivated	14	33	20	14
Typically motivated	64	55	62	62
Highly motivated	22	13	18	24
FRS[7] [median (IQR)]	6 (2, 12)	6 (2, 12)	–	–
Change in the number of cigarettes per day between the last two smoking visits[8] [median (IQR)]	0 (0, 0)	0 (0, 0)	–	–
Maximum number of cigarettes per day at the last three smoking visits[9] [median (IQR)]	20 (10, 20)	20 (15, 25)	20 (10, 20)	15 (8, 20)
Weight gain[1,10] (kg) [median (IQR)]	–	–	0 (–2, 2)	0 (–1, 2)

IQR, interquartile range; CDC, Centers for Disease Control and Prevention; FRS, Framingham risk score.
[1]Within the last 6 months prior to the index date.
[2]Defined as evidence for the beginning of a pregnancy within the last 9 months prior to the index date.
[3]Defined as the uptake of lipid- or blood pressure-lowering medication for the first time within the last 6 months prior to the index date.
[4]Defined as reporting high daily alcohol consumption (≥20 g/day for women and ≥40 g/day for men) or injecting drug use.
[5]Defined as reporting treatment by a psychiatrist, treatment with antidepressants, or depression diagnosed by a psychiatrist or other physician.
[6]The top four stressful life events on the Holmes and Rahe stress scale [17]: death of a spouse, divorce, marital separation or imprisonment within the last 6 months prior to the index date.
[7]Available in 1217 (97%) and 3442 (96%) patients who stopped and those who did not, respectively.
[8]Available in 1242 (98%) and 3570 (100%) patients who stopped and those who did not, respectively.
[9]Available in 1259 (100%) and 3572 (100%) patients who stopped and those who did not, respectively; and in 555 (100%) and 610 (100%) patients known to relapse and those not known to relapse, respectively.
[10]Available in 545 (98%) and 597 (98%) patients known to relapse and those not known to relapse, respectively.

age, gender, the FRS, and the change in the number of cigarettes per day (reported at the last two follow-up visits). We also report results of a sensitivity analysis with an additional candidate predictor: the maximum number of cigarettes per day reported at the last three follow-up visits.

For relapse, we included as candidate predictors: the maximum number of cigarettes per day reported at the last three follow-up visits prior to stopping, the individual variables used to classify patients into motivational groups (reassessed at the new index dates and with a previous cessation attempt replaced by the ratio of nonsmoking visits to total visits), age, gender and weight gain.

To assess the predictive performance of our clinical and statistical models for cessation and relapse, we plotted a receiver operating characteristic (ROC) curve for each model and calculated the area under the curve (AUC) [31].

We used SAS version 9.2 (SAS Institute Inc., Cary, NC) for data preparation; for our analyses and graphics, we used R version 3.0.1 (R Foundation for Statistical Computing,

Vienna, Austria) and the R add-on package *mboost* version 2.2-2 [27].

Results

For cessation

Patient characteristics

As at January 2012, 4833 smokers in the SHCS were eligible for inclusion in our study, and of these 1261 (26%) stopped smoking, at least temporarily. More patients who stopped were in the highly motivated group relative to those who did not stop (22% *vs*. 13%, respectively) and fewer were in the poorly motivated group (14% *vs*. 33%, respectively) (Table 2). Patients who stopped were more likely to have a history of attempting to stop and less likely to have alcohol or drug dependence and psychiatric comorbidities than those who did not stop. However, recent hospitalization was more likely in patients who did not stop than in those who did.
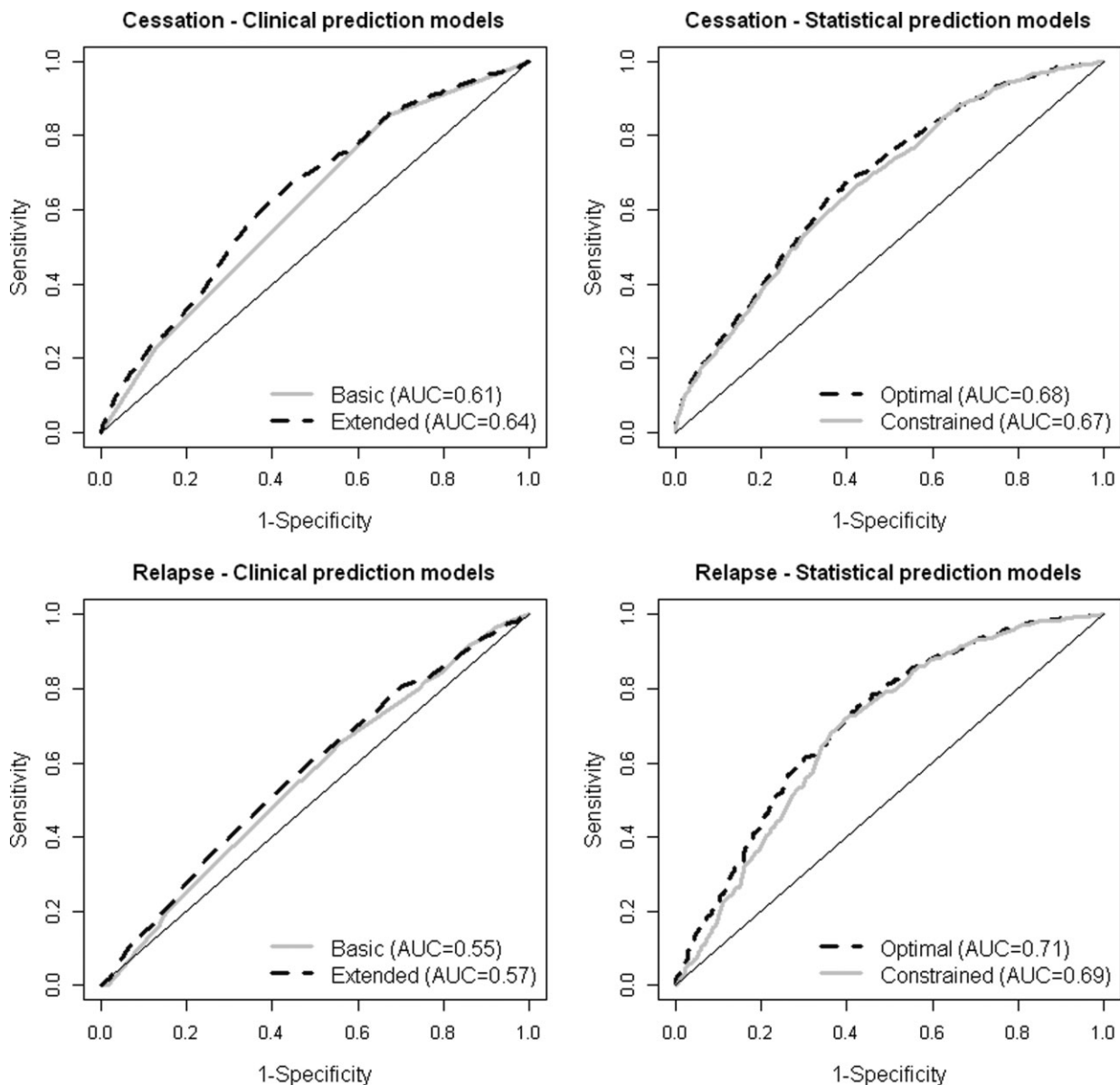
Fig. 1 Receiver operating characteristic (ROC) curves to assess the predictive performance of models for cessation and relapse. The area under the curve (AUC) can be interpreted as the probability that a patient who stops is given a higher predicted probability of cessation by the model than one who does not and that a patient known to relapse is given a higher predicted probability of relapse than one who does not [31]. The basic clinical model for cessation uses only three motivational groups as predictors; the extended clinical model adds age and gender. The basic clinical model for relapse is a linear function of a single predictor: the maximum number of cigarettes per day reported at the last three follow-up visits prior to stopping; the extended clinical model adds the three motivational groups. The optimal statistical model is an additive function of 12 and 10 predictors for cessation and relapse, respectively; this model is then constrained to just the most important predictor variables (five for cessation and one for relapse).

Predicting cessation

For predictive modelling, we included the 4639 patients (96%) without missing values in any of the candidate predictors; of these, 1199 (26%) attempted to stop. The predictive performance of our clinical and statistical models was modest (Fig. 1, top row), with AUC values of 61% to 68% (50% implies no ability to discriminate between patients who stopped and those who did not). The basic clinical model for cessation, with patients classified into the three motivational groups, was nearly as

Fig. 2 Estimated partial effects of selected continuous predictor variables of cessation and relapse, i.e. the combination of the linear and smooth curvilinear effect (solid red line), in the optimal statistical model as selected by the stability selection procedure. For comparison only, we also show the linear effect alone (dashed blue line). The predicted probability of cessation is highest for young adults, then decreases and may increase again for older individuals (a), and for light smokers with a low reported number of cigarettes per day (the maximum reported at the last three follow-up visits prior to stopping) (b); and the risk for relapse is highest for patients who have just stopped (c). Rugs at the abscissa indicate observed values of the relevant predictor variable, and the plot shows the partial contribution of this variable to the regression function.


discriminatory as the constrained statistical model with just the most important predictors: the ratio of nonsmoking visits to total visits, alcohol or drug dependence, psychiatric comorbidities, recent hospitalization and age (with both linear and nonlinear components). The extended clinical model for cessation consisted of the three motivational groups plus age and gender, while the optimal statistical model was an additive function of 12 predictor variables selected from a total of 13, with nonlinear components for age and for the ratio of nonsmoking visits to total visits. However, the increased model complexity of the extended clinical and optimal statistical models did not give appreciably better discriminatory ability compared with the basic clinical and constrained statistical models, respectively (Fig. 1, top row).

There was some evidence from both the optimal and constrained statistical models that the functional relationship for age may be nonlinear (Fig. 2a), with decreasing propensity to stop as age increases but with a turning point at around 55 years, and then a higher propensity to stop for individuals older than that. This particular functional form is consistent with the negative coefficient associated with age in the extended clinical model.

To illustrate the consequences of using these models, we assumed that clinicians treating HIV-infected patients might target an intervention at the 25% of their patients with the highest predicted probabilities of cessation. Such an intervention might involve offering additional encouragement and support to stop smoking through dedicated programmes or information about nicotine replacement products [32]. Following this strategy, our models imply that around 60% of the patients receiving the intervention could potentially benefit from it; however, 40% of these patients would have had sufficient motivation to stop on their own (Table 3).

Sensitivity analysis

In a sensitivity analysis, we considered the maximum number of cigarettes per day reported at the last three follow-up visits
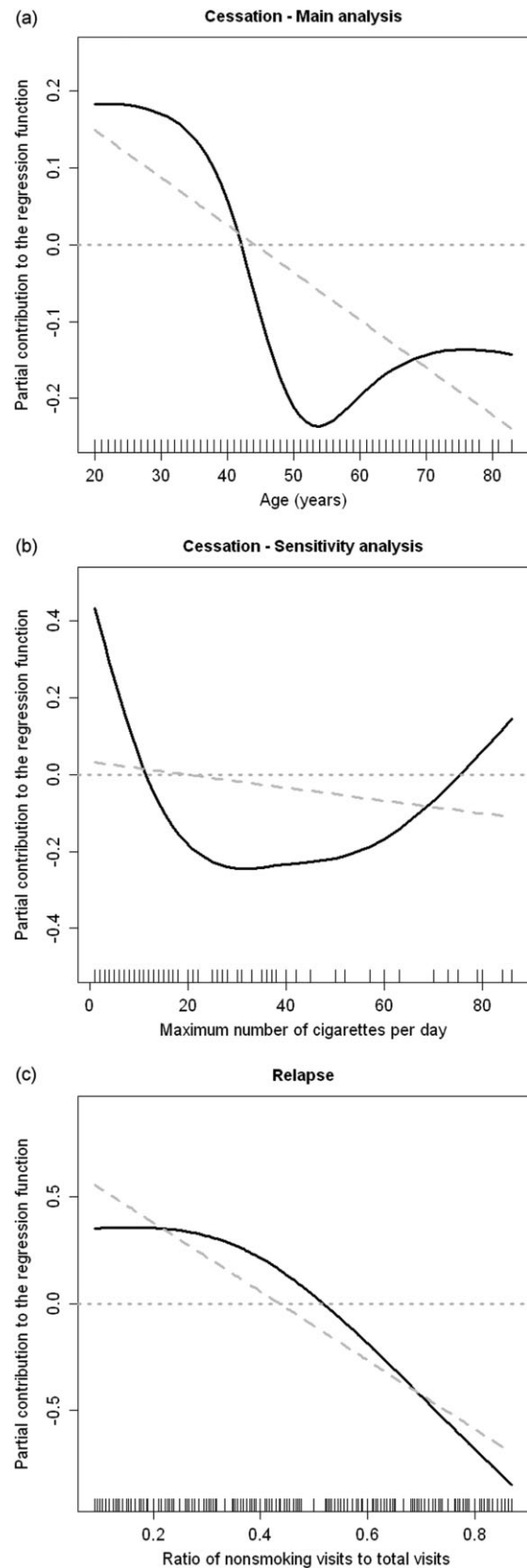
Table 3 Consequences of targeting a suitable intervention at the 25% and 50% of patients with the highest predicted probabilities of cessation and relapse, respectively, according to the clinical (basic and extended) and statistical (optimal and constrained) models

	Clinical models		Statistical models	
	Basic[1]	Extended[2]	Optimal[3]	Constrained[4]
Cessation: targeting an intervention at the 25% of patients with the highest predicted probabilities of cessation (n = 1160)				
Patients who did not stop on their own (%)	65	64	59	60
Patients who stopped on their own (%)	35	36	41	40
Relapse: targeting an intervention at the 50% of patients with the highest predicted probabilities of relapse (n = 570)				
Patients not known to relapse (%)	48	47	37	37
Patients known to relapse (%)	52	53	63	63

[1]For cessation, the three motivational groups; for relapse, the maximum number of cigarettes per day reported in the last three follow-up visits prior to stopping.
[2]For cessation, the three motivational groups plus age and gender; for relapse, the maximum number of cigarettes per day plus the three motivational groups.
[3]For cessation, 12 predictor variables selected from a total of 13; for relapse, 10 predictor variables selected from a total of 13.
[4]For cessation, constrained to five predictor variables (ratio of nonsmoking visits to total visits, alcohol or drug dependence, psychiatric comorbidities, hospitalization and age with linear and nonlinear model components); for relapse, constrained to one predictor variable (ratio of nonsmoking visits to total visits with linear and nonlinear model components).

(our measure of dependence) as an additional predictor of cessation. A smooth curvilinear component for this predictor was then added to the constrained statistical model and the nonlinear age component was omitted. There seems to be a group of patients reporting a maximum of around 30 cigarettes per day who are least likely to stop smoking, with patients reporting either lower or higher maximums more likely to stop (Fig. 2b).

For relapse

Patient characteristics

Of the patients who stopped smoking, 1167 (93%) were eligible for our comparison between patients known to relapse and those not known to relapse. Of them, 557 patients (48%) reported smoking on a subsequent occasion. Patients known to relapse reported a greater maximum number of cigarettes per day at the last three follow-up visits prior to stopping than those not known to relapse (Table 2). More patients known to relapse were in the poorly motivated group relative to those not known to relapse (20% *vs.* 14%, respectively) and fewer were in the highly motivated group (18% *vs.* 24%, respectively). Patients not known to relapse had a higher ratio of non-smoking visits to total visits, indicating a history of attempting to stop prior to cessation or a longer period without smoking since then.

Predicting relapse

For predictive modelling, we included the 1140 patients (98%) without missing values in any of the candidate predictors; of these, 543 (48%) reported smoking on a subsequent occasion. The predictive performance of our clinical and statistical models was modest (Fig. 1, bottom row), with AUC values of 55% to 71%. The basic clinical model for relapse, based on the maximum number of cigarettes per day reported at the last three follow-up visits prior to stopping, was not as discriminatory as the constrained statistical model with just a single predictor variable – the ratio of nonsmoking visits to total visits – but with both linear and nonlinear components for this variable. The extended clinical model for relapse consisted of the maximum number of cigarettes per day plus the three motivational groups, while the optimal statistical model was an additive function of 10 predictor variables selected from a total of 13, with nonlinear components for both the ratio of nonsmoking visits to total visits and the maximum number of reported cigarettes per day. Again, the increased model complexity of the extended clinical and optimal statistical models did not give appreciably better discriminatory ability compared with the basic clinical and constrained statistical models, respectively (Fig. 1, bottom row). To further simplify the constrained statistical model – with just the ratio of nonsmoking visits to total visits selected as a predictor variable of relapse – a linear function could be used to approximate the weakly curvilinear relationship (Fig. 2c).

To illustrate the consequences of using these models, we assumed that clinicians treating HIV-infected patients might target an intervention to prevent relapse at the 50% of their patients with the highest predicted probabilities of relapse. Our models imply that around 50% to 60% of the patients receiving such an intervention could potentially benefit; however, the remaining patients would have had sufficient resolve to resist relapse without intervention (Table 3).

Discussion

Smoking cessation (and its maintenance) is the obvious way to reduce the risk of CVD in HIV-infected patients [2]. Targeting limited resources at the patients most likely to benefit is a logical step towards integrating smoking cessation programmes into the routine care of HIV-infected patients [33]. Our results illustrate that it is difficult to predict which patient will give up smoking or take up smoking again. As a result, simple prediction models are nearly as discriminatory as complex ones. As a rough rule of thumb for clinicians, patients in our highly motivated group, especially those with a history of attempting to stop, and those known to have stopped recently are the best candidates for an intervention.

In our data, 26% of smokers stopped, at least temporarily; because among those who stopped, 48% started smoking again. These percentages are consistent with cessation rates of 38% and 25% in two different cessation programmes for HIV-infected patients, respectively [8,34], and relapse rates in other populations of up to 50% even after such programmes [35]. We contrasted clinical reasoning with an advanced statistical method, but neither approach was able to clearly identify patients who would stop smoking or those who would start again. Note that there is a contradiction inherent in predicting smoking cessation: perfect prediction would identify patients who do not need an intervention. Hence, modest predictive performance may be sufficient if it identifies patients similar to those who stop anyway who would benefit from an intervention. For this reason, we illustrate the consequences of using our models (Table 3). This contradiction does not arise when predicting relapse: all those identified by a perfect model could benefit from an intervention. The value of targeting interventions in this way would need to be assessed in a trial in combination with a specific intervention, but the modest predictive performance of our models for relapse does not augur well for this approach.

Under the transtheoretical model of behaviour change, decreasing cigarette use ought to be a sign of preparation for change and hence predictive of cessation [18]. In our data, the change in the number of cigarettes smoked per day was not a strong predictor of smoking cessation. Also, despite many studies showing that smoking cessation is more likely with increasing age and in male individuals [23,24,36,37], neither age nor gender was a particularly good predictor of smoking cessation. While these predictors appeared in some models, the relationship between age and cessation may be nonlinear, consistent with earlier findings where cessation rates decreased in 30- to 50-year-olds but then increased for older patients [9,24]. The FRS measures the actual risk of CVD – and cessation is more

likely with an increase in perceived risk [21,22]. However, the FRS did not prove a good predictor of cessation, even though it was associated with smoking cessation in an earlier intervention study using SHCS patients [8] and has age and gender as components. The number of cigarettes smoked per day – assumed to be a proxy measure of dependence [18] – was not a strong predictor of relapse as expected; rather, it may be better as a predictor of cessation [38] because heavy smokers may lack confidence in their ability to stop [14]. We assumed that hospitalization might serve as a 'wake-up call', triggering smoking cessation. However, the limited SHCS data available suggest that CVD is seldom the reason for hospitalization; hospitalization may act more as a stressful life event, with injury or illness the sixth item on the Holmes and Rahe stress scale [17], so that smoking cessation then becomes less likely. Note that the frequency of cancer diagnoses within the last 6 months was 0.5% and 0.4% among smokers who stopped and those who did not, respectively. Such events were not considered as a candidate predictor because of their infrequent occurrence. Obviously a cancer diagnosis should prompt intervention by the clinician to encourage cessation and prevent relapse.

The strengths of this study are a comprehensive data set with relevant variables routinely collected since April 2000 and the use of advanced statistical methods capable of detecting nonlinear relationships where these would improve predictive performance. However, this study has a number of limitations. First, 6-monthly visits may be too infrequent to detect changes that would otherwise be predictive of smoking cessation and relapse. Secondly, other variables that are not routinely collected in the SHCS might be more predictive. Richmond [39] and Fagerström scores [40] would better measure motivation to stop and dependence, respectively, but these scores would require nine new questions. Rather than routinely ask these questions, an additional questionnaire might be used for a time at some cohort centres in a substudy. Thirdly, smoking status and alcohol consumption are self-reported in the SHCS and sometimes responses are missing. For patients who stopped, we assumed that a patient was still not smoking if smoking status was missing at any subsequent visit. The opposite may be true, as those who restart may be reluctant to admit it, but only 21 patients who reported stopping had subsequent missing responses. Routine collection of self-reported alcohol consumption began only in August 2005. Patients with missing responses were not assumed to have had high alcohol consumption; this needed to be explicitly reported. Fourthly, the SHCS lacks information about participation in cessation programmes or use of nicotine-replacement therapy. If heavy smokers were more likely to use these interventions, this could explain why the reported

maximum number of cigarettes per day was not a good predictor of relapse and why patients reporting a maximum of more than 30 cigarettes per day were more likely to stop. Because of these limitations, the modelling undertaken in this study could conceivably prove more successful in other settings.

Our study suggests that smoking cessation and its relapse cannot be accurately predicted from variables typically collected in observational cohorts of HIV-infected patients. Complex models proved no more predictive than simple ones, and it seems unlikely that formal modelling will be useful as a means of targeting interventions at the patients most receptive to them. In this study, the ratio of nonsmoking visits to total visits was not only the most important predictor in the constrained statistical model for smoking cessation but also the only predictor in the constrained statistical model for relapse. Targeting interventions at the patients most likely to benefit is logical, but our study suggests that clinicians cannot easily do better than what is obviously sensible – target those trying to stop or those who have just stopped. Clinicians should ask patients if they are trying to stop and help those who are. They should reassure patients trying to stop that – consistent with the transtheoretical model of behaviour change – it can take several attempts but the longer they stay off cigarettes the higher the chances of success.

Acknowledgements

Appendix 1

The members of the Swiss HIV Cohort Study are: J. Barth, M. Battegay, E. Bernasconi, J. Böni, H. C. Bucher, C. Burton-Jeangros, A. Calmy, M. Cavassini, C. Cellerai, M. Egger, L. Elzi, J. Fehr, J. Fellay, M. Flepp, P. Francioli (President of the SHCS), H. Furrer (Chairman of the Clinical and Laboratory Committee), C. A. Fux, M. Gorgievski, H. Günthard (Chairman of the Scientific Board), D. Haerry (deputy of 'Positive Council'), B. Hasse, H. H. Hirsch, B. Hirschel, I. Hösli, C. Kahlert, L. Kaiser, O. Keiser, C. Kind, T. Klimkait, H. Kovari, B. Ledergerber, G. Martinetti, B. Martinez de Tejada, K. Metzner, N. Müller, D. Nadal, G. Pantaleo, A. Rauch, S. Regenass, M. Rickenbach (Head of Data Center), C. Rudin (Chairman of the Mother & Child Substudy), P. Schmid, D. Schultze, F. Schöni-Affolter, J. Schüpbach, R. Speck, P. Taffé, P. Tarr, A. Telenti, A. Trkola, P. Vernazza, R. Weber and S. Yerly.

Appendix 2

Component-wise gradient boosting uses gradient descent techniques to select the most informative features of a set of candidate predictors, with model complexity determined by the number of boosting iterations. We used (25-fold bootstrap) cross-validation to find the appropriate number of iterations so that the resulting statistical model had optimal predictive performance ('optimal statistical model') [41]. To further constrain for model complexity ('constrained statistical model'), we then (1) selected the smallest iteration with a cross-validated negative binomial likelihood that was at most 1 standard error larger than that of the optimal statistical model [42]; (2) found the frequency with which a variable or model component is selected when using random subsamples of size $n/2$ of the original data [43]; and (3) re-fitted the boosting model with only those variables or model components that were selected with inclusion frequencies $\geq 90\%$ [44].

References

1 Glass TR, Ungsedhapand C, Wolbers M *et al.* Prevalence of risk factors for cardiovascular disease in HIV-infected patients over time: the Swiss HIV Cohort Study. *HIV Med* 2006; **7**: 404–410.

2 Petoumenos K, Worm S, Reiss P *et al.* Rates of cardiovascular disease following smoking cessation in patients with HIV infection: results from the D:A:D study(*). *HIV Med* 2011; **12**: 412–421.

3 Clifford GM, Polesel J, Rickenbach M *et al.* Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. *J Natl Cancer Inst* 2005; **97**: 425–432.

4 Friis-Moller N, Sabin CA, Weber R *et al.* Combination antiretroviral therapy and the risk of myocardial infarction. *N Engl J Med* 2003; **349**: 1993–2003.

5 Friis-Moller N, Reiss P, Sabin CA *et al.* Class of antiretroviral drugs and the risk of myocardial infarction. *N Engl J Med* 2007; **356**: 1723–1735.

6 Gritz ER, Vidrine DJ, Fingeret MC. Smoking cessation a critical component of medical management in chronic disease populations. *Am J Prev Med* 2007; **33** (6 Suppl): S414–S422.

7 Vidrine DJ. Cigarette smoking and HIV/AIDS: health implications, smoker characteristics and cessation strategies. *AIDS Educ Prev* 2009; **21** (3 Suppl): 3–13.

8 Elzi L, Spoerl D, Voggensperger J *et al.* A smoking cessation programme in HIV-infected individuals: a pilot study. *Antivir Ther* 2006; **11**: 787–795.

9 Huber M, Ledergerber B, Sauter R *et al.* Outcome of smoking cessation counselling of HIV-positive persons by HIV care physicians. *HIV Med* 2012; **13**: 387–397.

10 Riemsma RP, Pattenden J, Bridle C *et al.* Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *BMJ* 2003; **326**: 1175–1177.

11 Prochaska JO, Redding CA, Harlow LL, Rossi JS, Velicer WF. The transtheoretical model of change and HIV prevention: a review. *Health Educ Q* 1994; **21**: 471–486.

12 Prochaska JO, Velicer WF, Rossi JS *et al.* Stages of change and decisional balance for 12 problem behaviors. *Health Psychol* 1994; **13**: 39–46.

13 Sutton S. Back to the drawing board? A review of applications of the transtheoretical model to substance use. *Addiction* 2001; **96**: 175–186.

14 West R. Assessment of dependence and motivation to stop smoking. *BMJ* 2004; **328**: 338–339.

15 Bucher HC, Rickenbach M, Young J *et al.* Randomized trial of a computerized coronary heart disease risk assessment tool in HIV-infected patients receiving combination antiretroviral therapy. *Antivir Ther* 2010; **15**: 31–40.

16 Schoeni-Affolter F, Ledergerber B, Rickenbach M *et al.* Cohort profile: the Swiss HIV Cohort study. *Int J Epidemiol* 2010; **39**: 1179–1189.

17 Holmes TH, Rahe RH. The Social Readjustment Rating Scale. *J Psychosom Res* 1967; **11**: 213–218.

18 DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. *J Consult Clin Psychol* 1991; **59**: 295–304.

19 Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002; **106**: 3143–3421.

20 Law MG, Friis-Moller N, El-Sadr WM *et al.* The use of the Framingham equation to predict myocardial infarctions in HIV-infected patients: comparison with observed events in the D:A:D Study. *HIV Med* 2006; **7**: 218–230.

21 Borrelli B, Hayes RB, Dunsiger S, Fava JL. Risk perception and smoking behavior in medically ill smokers: a prospective study. *Addiction* 2010; **105**: 1100–1108.

22 McCoy SB, Gibbons FX, Reis TJ, Gerrard M, Luus CA, Sufka AV. Perceptions of smoking risk as a function of smoking status. *J Behav Med* 1992; **15**: 469–488.

23 van Loon AJ, Tijhuis M, Surtees PG, Ormel J. Determinants of smoking status: cross-sectional data on smoking initiation and cessation. *Eur J Public Health* 2005; **15**: 256–261.

24 Osler M, Prescott E. Psychosocial, behavioural, and health determinants of successful smoking cessation: a longitudinal study of Danish adults. *Tob Control* 1998; **7**: 262–267.

25 Aubin HJ, Farley A, Lycett D, Lahmek P, Aveyard P. Weight gain in smokers after quitting cigarettes: meta-analysis. *BMJ* 2012; **345**: e4439.

26 Akaike H. Information theory and an extension of the maximum likelihood principle. In: Kotz S, Johnson NL, eds. *Breakthroughs in Statistics.* New York, Springer, 1992: 610–624.

27 Bühlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting (with discussion). *Statistical Science* 2007; **22**: 477–505.

28 Kneib T, Hothorn T, Tutz G. Variable selection and model choice in geoadditive regression models. *Biometrics* 2009; **65**: 626–634.

29 Copas JB. Regression, prediction and shrinkage (with discussion). *J R Stat Soc Ser B (Methodol)* 1983; **45**: 311–354.

30 Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med* 1990; **9**: 1303–1325.

31 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.

32 Moscou-Jackson G, Commodore-Mensah Y, Farley J, Digiacomo M. Smoking-cessation interventions in people living with HIV infection: a systematic review. *J Assoc Nurses AIDS Care* 2014; **25**: 32–45.

33 Niaura R, Shadel WG, Morrow K, Tashima K, Flanigan T, Abrams DB. Human immunodeficiency virus infection, AIDS, and smoking cessation: the time is now. *Clin Infect Dis* 2000; **31**: 808–812.

34 Fuster M, Estrada V, Fernandez-Pinilla MC *et al.* Smoking cessation in HIV patients: rate of success and associated factors. *HIV Med* 2009; **10**: 614–619.

35 Quist-Paulsen P, Gallefoss F. Randomised controlled trial of smoking cessation intervention after admission for coronary heart disease. *BMJ* 2003; **327**: 1254–1257.

36 Osler M, Prescott E, Godtfredsen N, Hein HO, Schnohr P. Gender and determinants of smoking cessation: a longitudinal study. *Prev Med* 1999; **29**: 57–62.

37 Wetter DW, Kenford SL, Smith SS, Fiore MC, Jorenby DE, Baker TB. Gender differences in smoking cessation. *J Consult Clin Psychol* 1999; **67**: 555–562.

38 Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. *Addict Behav* 2009; **34**: 365–373.

39 Richmond RL, Kehoe LA, Webster IW. Multivariate models for predicting abstention following intervention to stop smoking by general practitioners. *Addiction* 1993; **88**: 1127–1135.

40 Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The fagerstrom test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119–1127.

41 Efron B. Bootstrap methods: another look at the Jackknife. *Ann Stat* 1979; **7**: 1–26.

42 Breiman L, Friedman JH, Stone CJ, Olshen RA. *Classification and Regression Trees*. Boca Raton, Chapman and Hall/CRC, 1984.

43 Meinshausen N, Bühlmann P. Stability selection (with discussion). *J R Stat Soc Ser B (Methodol)* 2010; **72**: 417–473.

44 Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc Ser C (Appl Stat)* 1999; **48**: 313–329.