Immunologic marker paths for seroconversion: single determinations of immunoglobulin A and β_2-microglobulin are not adequate to estimate time of HIV infection

Multicohort Analysis Project Workshop. Part II

Objective: To investigate the usefulness of single determinations of serum immunoglobulin (Ig) A and β_2-microglobulin (β_2M) levels in estimating the time from HIV seroconversion.

Subjects: Five cohorts were represented in the workshop. The Multicohort Analysis Project (MAP) workshop database comprised 1744 HIV-infected patients with documented HIV seroconversion times, 363 of whom had AIDS. Overall, 1430 patients had two or more pre-AIDS CD4+ cell counts (13056 counts); 896 patients had two or more pre-AIDS IgA measurements (6081 observations); but only 2964 β_2M measurements were available.

Main outcome measures: Marker paths for $\log_e$ IgA and $\log_e$ β_2M and for $\sqrt{}$CD4+ cell count. Dependence on cofactors (age, sex, mode of HIV transmission) and attention to inter-individual variation in seroconversion level and annual decline in $\sqrt{}$CD4+ cell count. Logistic discrimination was performed using cofactor-adjusted paired $\log_e$ IgA and $\log_e$ β_2M to date HIV infections as recent (within 3 years of seroconversion), intermediate, or distant (≥ 6 years after seroconversion). Transition intensities between immunologically defined states (using IgA or β_2M) and to clinical AIDS.

Results: Linear functional form described the decline in $\sqrt{}$CD4+ cell count, except for the Italian cohort where a steeper annual loss of $\sqrt{}$CD4+ cell count occurred in the first year than thereafter. $\sqrt{}$CD4+ cell count at seroconversion depended on age and mode of transmission. Annual loss of $\sqrt{}$CD4+ cell count was less severe in those infected by sexual transmission. Non-monotone functional form emerged for $\log_e$ IgA with an initial decrease in the first year after seroconversion followed by an increase thereafter. $\log_e$ IgA levels were higher in older subjects and in those infected by sexual transmission, but lower in women. A quadratic growth curve described the marker path for $\log_e$ β_2M, which increased for 5–6 years after seroconversion but declined thereafter. $\log_e$ β_2M values were significantly higher in older patients and injecting drug users, but lower in women. The considerable heterogeneity of marker paths between individuals affected all three markers, but marker values appeared to track within an individual. Discrimination based on paired IgA and β_2M measurements from single blood samples performed poorly in classifying HIV infections as recent, intermediate or distant. High intensities of backward as well as forward transitions between immunologically defined states explained the poor discrimination based on single sample per individual.

Conclusions: Further study of how serum IgA reflects HIV infection and careful clinical and statistical assessment of reduced β_2M from 5 or 6 years after HIV infection are needed. The dependence of marker paths on mode of HIV transmission suggests that sexual transmission may have implications for HIV disease progression. The feasibility of using serum IgA and β_2M, evaluable in single stored blood samples, to estimate time of HIV infection has been set back by our results.

Introduction

In Part I, we showed that immunoglobulin (Ig) A and β_2-microglobulin (β_2M) are jointly prognostic of time to AIDS. Here, we investigate their usefulness in estimating the time from HIV seroconversion by considering the marker paths of CD4+ cell count (for reference), IgA and β_2M levels from seroconversion.

The stimulation of lymphoid cells, especially T cells, with mitogens results in increased β_2M production [1]. Increased levels are thought to reflect activation of the immune system and raised values of β_2M are seen in a number of acute viral infections [2–4], liver disease [5], renal failure and lymphoproliferative and autoimmune diseases. Levels have been shown to be higher in injecting drug users (IDU) [6] and in patients with haemophilia [7], partly due to injecting behaviour and the high incidence of liver disease in these two groups. There is a suggestion that homosexual men may also have slightly increased values as a result of exposure to infectious agents. Furthermore, it has been suggested that African Americans may have lower levels than other races [8].

IgA levels reflect B-cell activation and elevated levels are seen in a variety of immune disorders such as autoimmune diseases, malignancies, liver and renal diseases and Epstein–Barr virus infection [9]. It has been suggested that levels of IgA are higher in men than in women [10], and that there may be racial differences with Africans having slightly higher, and more widely varying IgA levels [11], possibly because of the high prevalence of parasitic and other infections in this group.

Published studies have suggested that both β_2M and IgA rise after HIV seroconversion and remain raised throughout infection [9,12–15]. Levels of β_2M have been shown to decrease after the initiation of zidovudine therapy [16,17], although this effect is thought to be transient. Little work [15] has been published regarding the effect of zidovudine on IgA levels. Of particular relevance for public health surveillance of HIV infection is whether marker paths, or parametrizations of them, depend on current age group, sex, exposure category and geography. Because unlinked anonymous testing programmes [18,19] in hospitals, including antenatal and genitourinary clinics, are likely to exclude or at least differentiate patients with AIDS, we have only considered marker paths up to AIDS diagnosis in one of our evaluations but have included post-AIDS values in the other. Data on zidovudine use are not available from unlinked anonymous testing programmes and so we used marker values before and after the start of therapy. Elsewhere [20], we have assessed whether initiating zidovudine therapy affects marker values.

Methods

To stabilize variance and normalize residuals, a logarithmic transformation was used for IgA and β_2M and a square root transformation for CD4+ cell counts [20]. Box and whisker plots (Figs 1 and 2) were used to display cross-sectional marker changes in the years post-seroconversion, their relation to age and exposure category (Figs 1 and 2). Each person was represented by their first CD4+ cell count in any year (if measurement occurred in that year) and data sets for all five cohorts were pooled.

Two statistical approaches to the modelling of marker paths were used in parallel. First, multilevel modelling [21,22] was chosen because it allowed for inter-individual random variation in fitted linear or quadratic growth curves, and flexible analysis of intercept, slope or quadratic parameter dependence on covariates. The basic linear growth curve, or random effects model, is specified by an average (fixed) intercept and slope, the variance of individualized intercepts and slopes (random effects) about their respective average, and the residual variance of observations about these individual growth curves. All random effects are assumed to be normally distributed. Extension to accommodate the regression effect of cofactors such as age, sex and exposure category on the intercept and slope in a fixed or random manner, and quadratic and higher order growth curves can also be easily modelled.

Marker values following AIDS diagnosis are not considered and analyses were restricted to patients with two or more pre-AIDS marker values. Multilevel modelling for each marker was performed separately for each cohort and for the Italian subcohorts before performing a combined analysis.

A series of straight lines with appropriate slope changes (piecewise linear) can be used to approximate any functional form. Thus, we adopted an alternative approach of piecewise linear regression, using a covariance structure to describe the tracking of serial measurements for each individual [23,24]. Covariates can also be accommodated but have been assumed here only to affect intercept, i.e., marker level at seroconversion.

Only the first observation per individual in each 4-month period was used and post-AIDS marker values were retained for this second analysis. Regression slopes varied yearly, β_1 being the initial slope. The estimated slope in a given year interval was the sum of the slopes from the preceding and current intervals. For example, the slope for the time 12–24 months would be $\beta_1 + \beta_2$ if there had been a slope change after 1 year. One slope change was needed, at most, to characterize the marker path for $\sqrt{CD4+}$ cell count [20], $\log_e$ IgA (both had slope changes after 1 year) and $\log_e \beta_2$M (slope changes after 5
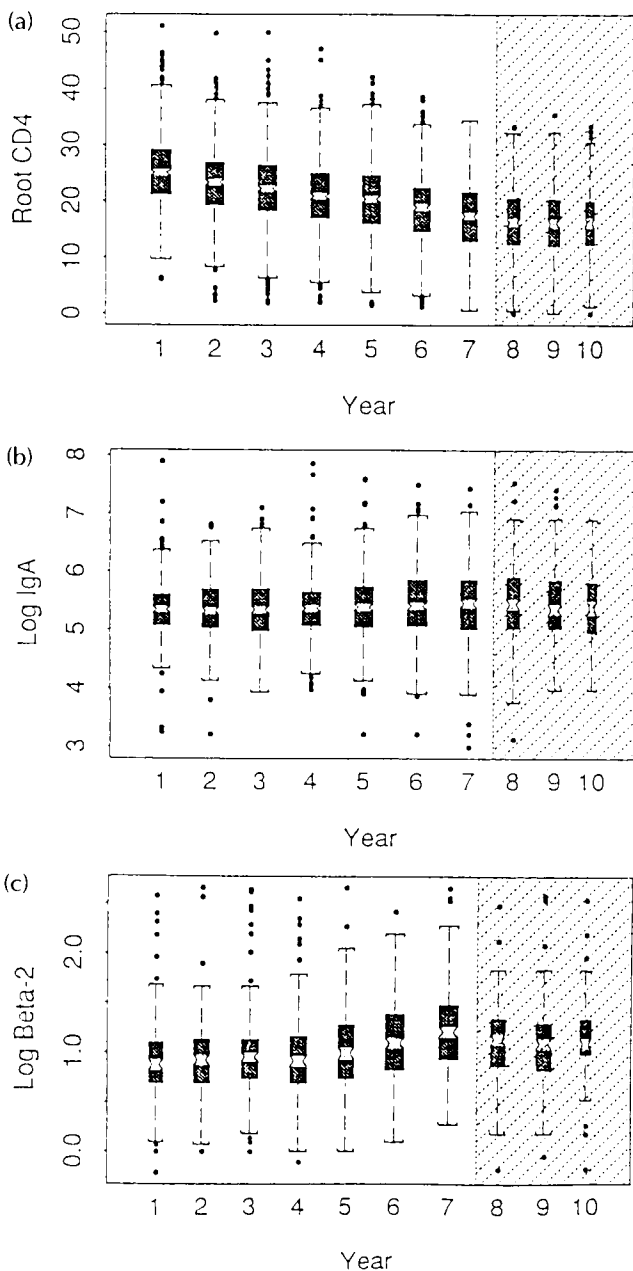
Fig. 1. Box and whisker plots for combined cohorts. Boxes show the interquartile range and the whiskers show 1.5 times the interquartile range. Misalignment of the notches on different boxes indicates significant differences between medians. (a) $\sqrt{CD4}$ cell count; (b) $\log_e$ immunoglobulin A; (c) $\log_e$ β_2-microglobulin for all cohorts.

years). The dependence of serial marker measurements, separated by s multiples of 4 months for a given patient, was formulated as $\gamma^{s\theta}$, where γ (Table 1) represents the correlation between marker values measured 4-months apart on the same individual, and θ allows various forms of serial dependency between repeated measures (at multiples of 4 months apart) on the same individual, ranging from an exponentially decaying structure ($\theta = 1$, most information from the previous observation) to constant correlation ($\theta = 0$, all past measurements are weighted

equally). Values of θ closer to 0 indicate higher degrees of tracking.

The above methods were applied to each marker in turn and separate analyses were performed for each cohort. A change in laboratory methods for determining IgA and β_2M in the Edinburgh cohort from September 1990 onwards was later discovered and adjusted for at analysis (but not in the box and whisker plots). Of particular interest was the form of the resultant piecewise linear regression, its reproducibility across cohorts, and the influence of surveillance covariates (age, sex, mode of transmission) on marker level at seroconversion. Covariate influences on slope and changes of slope were investigated by multi-level modelling.

The covariate influences estimated from the combined data were used to produce a bivariate plot of covariate-adjusted $\log_e$ IgA and $\log_e$ β_2M for a basic (so-called training) set of 1500 samples. This allowed us to make ordered logistic discriminations between samples infected recently (within the last 3 years) versus intermediate HIV infections versus distant infections (≥ 6 years). Predictive performance was assessed by assuming that only a single blood sample was available per individual, as in unlinked anonymous testing programmes, and was shown to be poor. The reasons for this relate both to inter-individual variation and to the intensity of backward as well as forward transitions between immunologic states defined by marker levels. We used IgA and β_2M values to estimate these transitions between immunologic states and to the final state of AIDS [25–28] in the same manner as CD4+ cell counts have been considered previously by other authors [25,26].

Results

Measurement of IgA in µg/l (not performed in the National Cancer Institute cohort) and β_2M in mg/l (only available at enrolment in the Toronto cohort) was cohort-specific and included evaluations on stored as well as prospective sera.

The cross-sectional marker changes in the years post-seroconversion, their relation to age and exposure category are shown by box and whisker plots. Median levels of $\sqrt{CD4+}$ cell count fell in each year after seroconversion, with significant differences in medians from year to year post-seroconversion (Fig. 1). Cross-sectional medians should not be interpreted as a typical individual trajectory because patients who progress rapidly to AIDS or death do not contribute to the plot in later years. The plot thus selects for more healthy individuals and we have used shading as warning of this.

Median $\log_e$ β_2M shows little obvious upward trend in the box and whisker display (Fig. 1), despite the fact that dramatic changes can occur within indi-
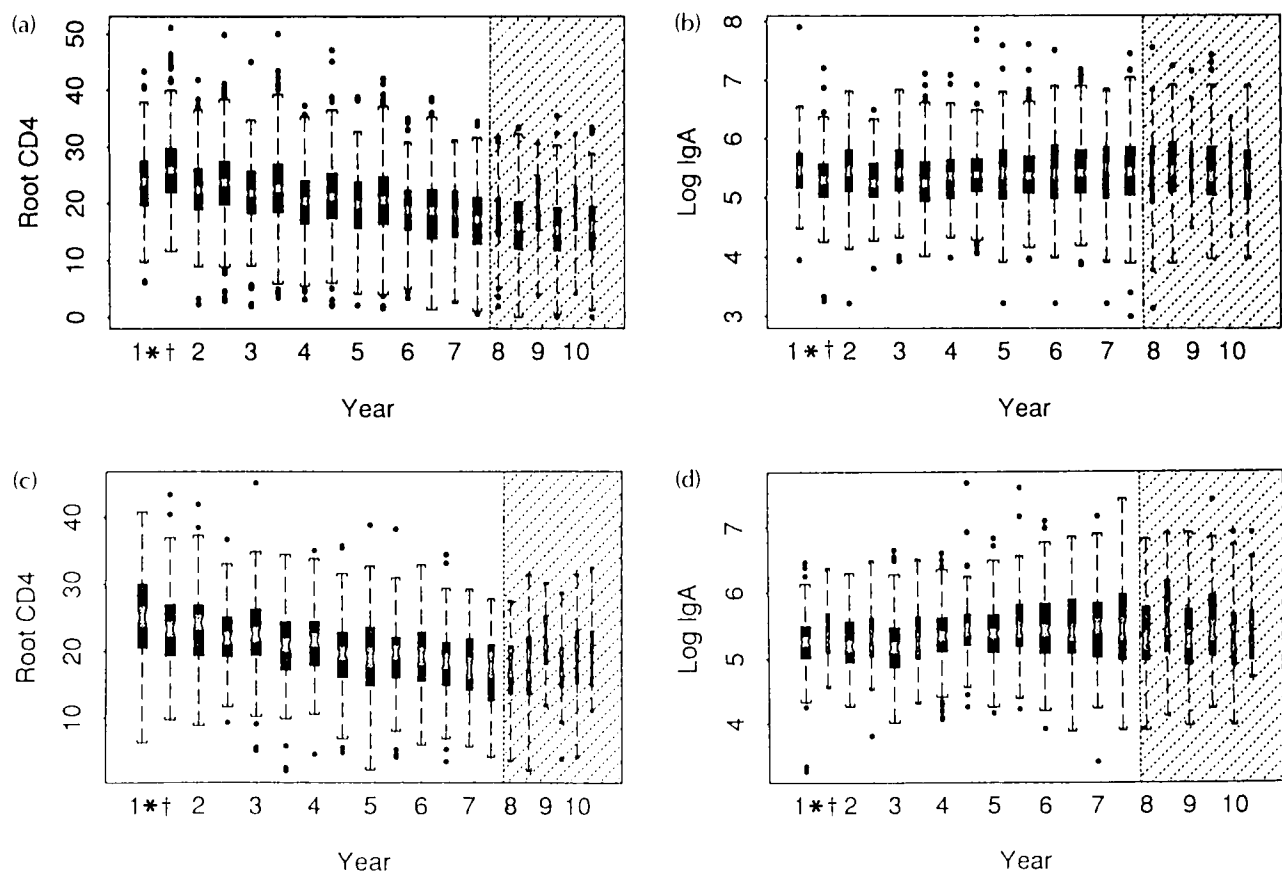
Fig. 2. Box and whisker plots for √CD4 and log$_e$ immunoglobulin (Ig) A by risk and age-group at seroconversion and currently. (a)√CD4 cell count (•, homosexuals; †, others). (b) Log$_e$ IgA (•, homosexuals; †, others).(c) √CD4 cell count for homosexuals by age (•, < 30 years; †, others). (d) Log$_e$ IgA for drug users by age (•, < 30 years; †, others).

viduals. β_2M determinations were not available in the London cohort until at least 5 years after seroconversion. In the log$_e$ β_2M plot for the combined cohorts (Fig. 1) a more obvious increase in marker distribution up to about year 7 is apparent, which decreases thereafter. The reversal in trend after year 7 may be genuine, an artefact of selection for slow progressors as described above, or a consequence of Edinburgh's changing method for β_2M determination (the newer diffusion method resulted in lower values; see below).

Exposure category differences in median were apparent in the pooled data, with different medians for homosexual men compared with other risk groups in the first years after seroconversion for √CD4+ cell count and log$_e$ IgA (Fig. 2). Homosexual men had more extreme values on average, except for log$_e$ β_2M levels. The effect of age on √CD4+ cell count, while controlling for exposure group, is also shown in Figure 2. In the case of homosexual men, both age at seroconversion and adjusted age (current age) in each year post-seroconversion had obvious effects on the median level; men aged > 30 years had lower medians. An age effect on log$_e$ IgA among drug users, who are younger in general and more homogenous in age at seroconversion, was easier

to discern using current age than age at seroconversion.

Marker plots thus suggested that functional forms for log$_e$ IgA and log$_e$ β_2M were more difficult to assess than for √CD4+ cell count.

√CD4+ cell marker path

A linear functional form was adopted for √CD4+ cell count in multilevel modelling based on previous work [29,30] but was re-examined via piecewise linear regression. The effects of sex (Edinburgh and Italian cohorts only) and age at seroconversion were first assessed in individual cohorts before combining cases across cohorts. In the combined analysis, route of HIV transmission was taken into account — sexual transmission (homosexual or heterosexual) versus otherwise (injecting drug use or blood products) — and thus exposure category in addition to geography (representing laboratory or cohort effects) was investigated as an explanatory factor for both intercept and slope.

For each cohort and the Italian subcohorts, the basic linear random effects model with no covariates was fitted first [20] followed by a second in which age at seroconversion had an estimated fixed effect on √CD4+ cell count at seroconversion (on

Table 1. Marker path summaries for Edinburgh and Italian cohorts allowing slope change, correlation and tracking within individuals, and covariate influences on seroconversion level.

	Coefficient (SE)				
	$\sqrt{CD4}$		$10^* \log_e$ IgA	$10^* \log_e \beta_2$ M	
Covariate influences	Edinburgh	Italy	Italy	Edinburgh[†]	Italy
Age and sex included					
Intercept : seroconversion level	27.55 (1.62)	29.53 (0.95)	49.94 (1.12)	9.73 (1.58)	8.60 (0.83)
Slope: initial slope	−1.50 (0.07)	−2.70 (0.32)	−0.81 (0.37)	0.52 (0.23)	0.24 (0.06)
Change in slope		1.25* (0.35)	1.37* (0.40)	−0.77‡ (0.28)	–
Age : effect on seroconversion level	−0.133 (0.056)	−0.107 (0.023)	0.189 (0.027)	0.030 (0.037)	0.045 (0.019)
Female sex : effect on seroconversion level	0.87 (0.59)	0.45 (0.40)	−0.98 (0.47)	−0.80 (0.39)	−0.86 (0.35)
γ: correlation within individuals	0.769 (0.014)	0.736 (0.011)	0.869 (0.009)	0.643 (0.047)	0.695 (0.023)
θ: tracking within individuals	0.315 (0.025)	0.380 (0.023)	0.511 (0.040)	0.251 (0.115)	0.298 (0.060)
SD	6.36	6.21	5.23	3.72	3.65
−2 $\log_e$ likelihood	18646.1	27978.3	1398.1	430.7	566.5
Age, sex and mode of transmission included					
Intercept : seroconversion level	–	29.35 (0.97)	51.04 (1.12)	–	8.35 (0.84)
Slope : initial slope	–	−2.69 (0.32)	−0.85 (0.37)	–	0.24 (0.06)
Change in slope	–	1.24* (0.35)	1.41* (0.40)	–	–
Age : effect on seroconversion level	–	−0.097 (0.025)	0.127 (0.029)	–	0.063 (0.021)
Female sex : effect on seroconversion level	–	0.49 (0.40)	−1.19 (0.46)	–	−0.83 (0.35)
Sexual transmission versus IDU	–	−0.38 (0.38)	2.18 (0.43)	–	−0.64 (0.33)
Effect on seroconversion level					
γ: correlation within individual	–	0.736 (0.011)	0.865 (0.009)	–	0.692 (0.23)
θ: tracking within individuals	–	0.380 (0.023)	0.513 (0.040)	–	0.297 (0.060)
SD	–	6.21	5.14	–	3.63
−2 $\log_e$ likelihood	–	27977.3	1373.7	–	562.7

*Change in slope from 1 year after seroconversion; †adjusted for change in method; ‡change in slope from 5 years after seroconversion. IgA, immunoglobulin A; β_2M, β_2-microglobulin; IDU, injecting drug use.

intercept). From the first models, we found that $\sqrt{CD4}$+ cell count at the time of seroconversion was lower in the Toronto (22.2), NCI (25.7) cohorts and Italian sexual transmission subcohorts (homosexual, 25.3; heterosexual, 25.1) than for Edinburgh's mainly IDU cohort (26.5), the Italian IDU subcohort (27.3) and the London haemophilic cohort (29.2). Rate of loss of $\sqrt{CD4}$+ cell count appeared to be less rapid in association with sexually transmitted HIV infection (slopes, −0.7, −1.1, −1.3, −1.3 for the Toronto and NCI cohorts and the Italian subcohorts, respectively) than for non-sexually transmitted HIV infection (−1.6, −1.6, −1.8 for Edinburgh, the Italian drug user subcohort and London haemophilic cohort, respectively). $\sqrt{CD4}$+ cell count at seroconversion was significantly lower in older patients in all cohorts except the NCI cohort [20]: by 1.0, 0.9 and 1.0 per 10 years of age for the Toronto and Italian sexual transmission subcohorts and by 1.6, 1.2 and 1.4 for Edinburgh, the Italian drug user subcohort and London haemophilic cohort, respectively. The effect of sex on $\sqrt{CD4}$+ cell count intercept could be estimated for the Edinburgh cohort and separately for the Italian IDU and heterosexual subcohorts. All three estimates, controlling for exposure category, were equivocal in their suggestion that females may have higher $\sqrt{CD4}$+ cell count at seroconversion [20].

Several interesting features emerged from corresponding analyses using piecewise linear regression with damped exponential covariance structure. First, the Italian cohort, which had higher $\sqrt{CD4}$+ cell counts soon after seroconversion than the others, revealed that the decline in $\sqrt{CD4}$+ cell count was more dramatic in the first year after seroconversion (−2.7; SE, 0.3) than subsequently (slope changes to −2.7 + 1.3 = −1.4; Table 1). Second, a high degree of tracking was evident in serial $\sqrt{CD4}$+ cell counts for individual patients with estimated θ: 0.31 (SE, 0.03), 0.38 (SE, 0.02), 0.48 (SE, 0.05), 0.60 (SE, 0.10) and 0.42 (SE, 0.04) for the Edinburgh, Italian, NCI, London and Toronto cohorts, respectively.

Multilevel modelling of the combined database of 1430 patients with two or more pre-AIDS $\sqrt{CD4}$+ cell counts summarizes the effects of age and sex, while allowing for sexual transmission and cohort effects on both $\sqrt{CD4}$+ cell intercept and slope (Table 2). $\sqrt{CD4}$+ cell count at seroconversion was 0.96 lower (SE, 0.18) with each additional 10 years of age; and lower by 1.82 (SE, 0.43) in patients who acquired their HIV infection sexually rather than via injecting drug use or blood products. However, sexually transmitted HIV infection was also associated with a significantly slower decline in $\sqrt{CD4}$+ cell count:

1.26 per annum compared with 1.62 for nonsexually transmitted HIV disease [31]. The Edinburgh rather than the Toronto cohort was used as the baseline reference because geographical heterogeneity in √CD4+ cell slope and intercept was largely accounted for by the slow decline from an already low √CD4+ cell count at seroconversion in the Toronto cohort. Being female was not demonstrably associated with a higher √CD4+ cell count at seroconversion (95% confidence interval, –0.3 to 1.0). The inter-patient SD in √CD4+ cell count at seroconversion and in the decline of √CD4+ cell count (slope), were 5.1 and 1.1, respectively, emphasizing the considerable individual variation despite adjustment for cofactors (Table 2).

Log_e IgA marker path

IgA measurements were not available for the NCI cohort. In the Toronto cohort, IgA was measured as frequently as CD4+ cell count but in other cohorts the availability of IgA measurements was roughly half that for CD4+ cell counts.

The results of fitting piecewise linear regression with damped exponential covariance structure to log_e IgA values, including those measured post-AIDS, for the Italian (Table 1), London and Toronto cohorts are detailed separately [20]. All three showed significant slope changes from negative to positive 1 year after seroconversion, and the damping parameter θ indicates a high degree of tracking. An apparent difference in log_e IgA marker path between the Edinburgh and other cohorts was discovered. This was a con-

sequence of Edinburgh's adoption of the diffusion assay to determine IgA in mid 1990; the diffusion assay had also been used to retrospectively determine IgA in earlier stored sera. To correct for this, an assignment to laboratory method for each sample was made that was expected to be correct for the majority of samples, based on the number of digits used in recording the measurements. The original non-diffusion method was associated with systematically higher log_e IgA determinations, by 0.165 (SE, 0.028) on average, or 3% [20].

For the combined log_e IgA analysis (Table 2) the Edinburgh cohort was restricted to patients with two or more log_e IgA determinations by the old (non-diffusion) method. Table 2 summarizes the functional form for log_e IgA as a quadratic growth curve with random effects and shows the influence of surveillance covariates, including sexual versus other modes of HIV transmission, on log_e IgA at seroconversion. The results confirm significantly higher levels in older individuals, in those infected by sexual contact, and in men. Being female and older by 10 years had opposite and nearly equal effects on log_e IgA at seroconversion. Infection acquired sexually increased log_e IgA by the equivalent of an additional 20 years of age. Sexual transmission, which slowed the decline in √CD4+ cell count, did not significantly alter the linear aspect of the log_e IgA marker path. The level of log_e IgA around seroconversion was re-established within 2–3 years and then increased steadily (Table 1): IgA at seroconversion for a 30 year-old Italian male in-

Table 2. Linear and quadratic random effects describing √CD4 count and 10* log_e immunoglobulin (Ig) A and the influences of age, sex, mode of transmission and cohort for cases with two or more √CD4 counts or 10* log_e IgA.

Parameter for combined cohorts	√CD4 (1430 cases; 13056 counts)		10* log_eIgA (896 cases; 6081 observations)	
	Estimate (SE)	SD*	Estimate (SE)	SD*
Intercept: at seroconversion*	26.77 (0.89)	5.10*	54.03 (0.99)	4.16*
Age: effect on intercept	–0.096 (0.018)	–	0.10 (0.02)	–
Female sex: effect on intercept	0.31 (0.33)	–	–1.21 (0.38)	–
Sexual transmission: effect on intercept	–1.82 (0.43)	–	2.01 (0.41)	–
Cohort effects on intercept				
Edinburgh	1.72 (0.76)	–	BC	–
Italy	2.70 (0.59)	–	–2.76 (0.63)	–
London	4.94 (0.94)	–	–3.45 (0.18)	–
NCI	3.71 (0.78)	–	NA	–
Toronto	BC	–	–3.38 (0.78)	–
Slope: annual change*	–1.08 (0.17)	1.11*	–0.51 (0.18)	1.68*
Sexual transmission: effect on slope	0.36 (0.12)	–	NA	–
Cohort effects on slope				
Edinburgh	–0.54 (0.18)	–	BC	–
Italy	–0.55 (0.15)	–	0.31 (0.14)	–
London	–0.73 (0.22)	–	0.41 (0.17)	–
NCI	–0.38 (0.19)	–	NA	–
Toronto	BC	–	–0.23 (0.15)	–
Quadratic change*	NA	–	0.11 (0.02)	0.18*
Residual variance	–	11.63	–	6.53

*SD of random, i.e., individualised effects. NCI, National Cancer Insitute; BC, baseline cohort; NA, not applicable.

fected by sexual contact would be expected to be $\exp[51.04 + 30(0.127) + 2.18]/10 = 300$, falling after 1 year to $\exp[5.703-0.085] = 275$, recovering by the end of his second year to $\exp 5.703-0.085 + (0.141 -0.085) = 291$, and increasing thereafter. At the end of 3 years, his expected IgA would be $\exp 5.674 + 0.056 = 308$.

Table 2 indicates considerable individual variation in $\log_e$ IgA at seroconversion (random SD, 0.416), in slope (random SD, 0.168) and in quadratic terms (random SD, 0.018) even after adjusting for surveillance covariates.

$\log_e \beta_2 M$ marker path

$\beta_2 M$ measurements were not available for the Toronto cohort, except for the initial visit. In other cohorts, the availability of $\beta_2 M$ was about a quarter that of CD4+ cell counts but the range of $\beta_2 M$ measurement times varied between cohorts (Fig. 3). In particular, no measurements within 5 years of seroconversion were available from the London haemophilia cohort, and few beyond 5 years from the Italian cohort. Edinburgh changed to a diffusion technique for $\beta_2 M$ in 1990 at the same time as the IgA method changed and $\log_e \beta_2 M$ values were $1.5/10 = 0.15$ lower, on average, by the diffusion assay [20].
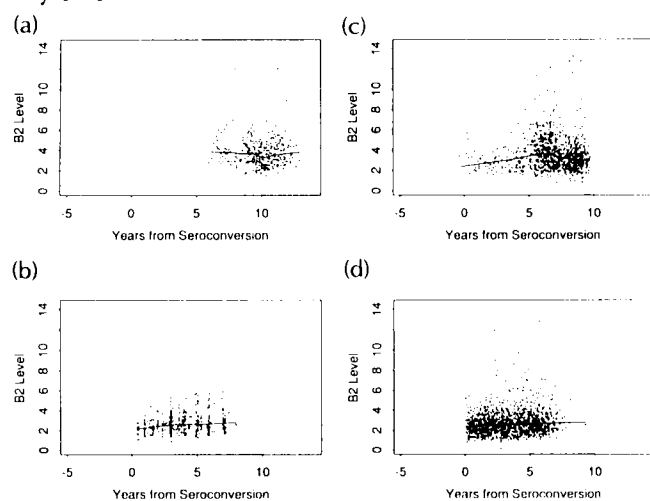
Fig. 3. Scatterplot and lowess smooth of β_2-microglobulin by cohort. (a) London Royal Free Hospital cohort; (b) National Cancer Institute (NCI) cohort; (c) Edinburgh cohort; (d) Italian cohort.

A linear functional form sufficed for $\log_e \beta_2 M$ in the Italian cohort but was inadequate for Edinburgh, the next largest cohort (Table 1) [20]. Both statistical approaches to the Edinburgh data suggested that $\log_e \beta_2 M$ rises until nearly the sixth year after seroconversion and then falls at a rate of 0.025 per annum according to piecewise linear regression. The Edinburgh data required a quadratic random effect, significantly negative on average (-0.016; random SD, 0.026; SE, 0.003, based on 217 cases and 743 observations) combined with a strongly positive slope

(0.192; random SD, 0.31; SE, 0.037) and intercept of 0.525 (random SD, 0.611; SE, 0.091).

A linear fixed effects regression with negative slope (-0.033; SE, 0.012) described the functional form of the London data 5.2 to 12.9 years after seroconversion. This was consistent with the Edinburgh slope of -0.025 per annum after 5 years. We were unable to fit the NCI data using either of the two statistical approaches.

For $\log_e \beta_2 M$, as for the other two markers studied, there was significant tracking within individuals (Table 1). Higher levels of $\log_e \beta_2 M$ were also found in older patients and in IDU, and lower levels in women.

Extending public health surveillance of HIV disease

For surveillance purposes, the early decrease in $\log_e$ IgA and late fall in $\log_e \beta_2 M$ suggest the possibility of bivariate discrimination between samples from HIV-infected individuals who seroconverted ≤3 years (recent), 3–6 (intermediate), or ≥6 years previously (distant), respectively. These time-since-infection intervals were defined roughly in accordance with the non-monotonicity of $\log_e$ IgA and $\log_e \beta_2 M$ marker paths, and because the greatest uncertainty relates to HIV infection in the last 3 years, for which back-calculation from AIDS diagnoses is largely uninformative [18]. This 3-way discrimination, if feasible, would be adequate for surveillance.

Figure 4 shows the covariate-adjusted bivariate plot of $\log_e \beta_2 M$ against $\log_e$ IgA using only the Edinburgh, Italian and London cohorts (which have serial measurements for both markers). Details of covariate adjustment, based on Tables 1 and 2, are given elsewhere [20] and take account of method changes in Edinburgh. The data shown in Figure 4 suggest that the required discrimination between samples within 3 years of seroconversion, 3–6 years after seroconversion, and ≥6 years after seroconversion, would be poor on an individual basis. To investigate whether it would be satisfactory for groups of HIV-antibody-positive samples, we estimated ordered logistic risk scores to give the probability of a sample being within 3 years of seroconversion or within 6 years of seroconversion, using the 1500 covariate-adjusted bivariate samples (409 recent, 469 intermediate, 622 distant) shown in Fig. 4 as a training dataset. A 50% sure classification performed better than raw logistic scores but still classified only 582 (39%) of samples in the training set, and correctly classified only 15 (13%) recent infections and 301 (48%) distant infections.

Intensity of transitions between immunologic stages defined by IgA and $\beta_2 M$ levels

We determined the intensities of backward as well as forward transitions between IgA and $\beta_2 M$ stages. These emphasize the difficulty of aging infections on
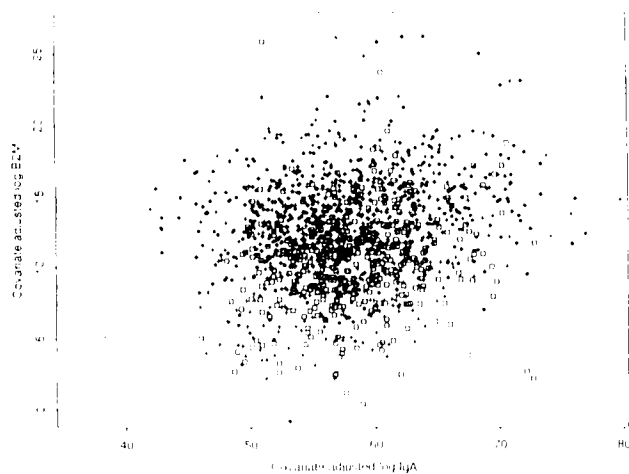
Fig. 4. Scatterplot of $\log_e \beta_2$-microglobulin (β_2M) versus $\log_e$ immunoglobulin (Ig) A (both covariate-adjusted) by time since HIV seroconversion. ( ), recent seroconversion (within 3 years); (+), intermediate seroconversion (3–6 years); (■), distant seroconversion (≥6 years). All values plotted effectively relate to an Edinburgh homosexual man aged 30 years at HIV seroconversion. Only the first pair of IgA and β_2M values per individual was included for any year after seroconversion; post-AIDS values are included.

the basis of a single determination. Four states were defined for each marker. For IgA, the intervals corresponding to the four states were < 200, 200–400, 400–600, and ≥ 600 μg/l. For β_2M they were 0–1.8, 2.8–4.5 and ≥ 4.5 mg/l. The transitions allowed between adjacent states in both directions and to the final state of clinical AIDS are shown in Fig. 5.
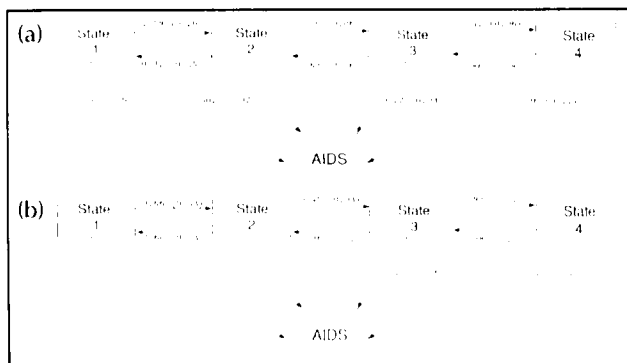

Fig. 5. Estimated transition intensities (SE) between immunologically defined states and to AIDS. (a) β_2-microglobulin state definitions and estimated transition intensities: 1 (0–1.8), 2 (1.8–2.8), 3 (2.8–4.5), 4 (≥ 4.5). (b) Immunoglobulin A state definitions and estimated intensities: 1 (0–200), 2 (200–400), 3 (400–600), 4 (≥ 600).

Edinburgh's β_2M measurements using the diffusion method have been adjusted by multiplication by exp (0.15). A comparably simple adjustment was not appropriate for IgA measurements and thus the non-diffusion IgA measurements from Edinburgh were used for the analyses reported here. No other geographic (between laboratory) or covariate adjustment was made. Initially, cohorts were analysed individually. The comparability of the results from the

individual cohorts [32] led to a pooling of data for subsequent analyses.

The estimated transition intensities and SE based on the pooled data are shown in Fig. 5. For β_2M, the forward transition rate dominated between states 1 and 2, forward and backward rates were comparable for states 2 and 3, and the backward rate dominated for states 3 and 4. These results were expected given the previous MAP analyses that suggested $\log_e \beta_2$M rises until the start of the sixth year after seroconversion and falls thereafter. For IgA, the forward and backward transition intensities between states 1 and 2 were comparable, while for states 2 and 3 and for states 3 and 4, the backward transition rates were somewhat higher than the forward rates. Given the initial decline and subsequent rise in IgA reported here, such estimates are surprising. The transition intensities in the AIDS state, however, show the expected gradient as a function of the marker state.

These results indicate that the use of β_2M and IgA to estimate time of HIV infection on the basis of a single determination is problematic because of the high probability of both increasing and decreasing values over time.

Discussion

Using statistical approaches for modelling marker paths has shown that $\sqrt{CD4+}$ cell count decays more rapidly in the first year after seroconversion than subsequently, as has been suggested by Weiss *et al.* [33] and Galai *et al.* [24]. However, there is considerable individual variation in the subsequent steady decline in $\sqrt{CD4+}$ cell count, and a high degree of tracking in serial $\sqrt{CD4+}$ cell count within patients. We have also shown that $\log_e$ IgA declines in the first year after seroconversion but rises linearly thereafter. Unexplained random variation between individuals results in heterogeneity in the time before $\log_e$ IgA regains its seroconversion level (typically 2–3 years after seroconversion) and for individual variation in that level. An initial, acute rise in $\log_e$ IgA in response to HIV infection may occur in the weeks prior to seroconversion (when measurement is unlikely), after which $\log_e$ IgA then declines before a chronic $\log_e$ IgA response to HIV infection is established. Data were frequently more fragmentary for $\log_e \beta_2$M measurements. Nevertheless, a coherent picture was constructed by piecing together information across cohorts: $\log_e \beta_2$M rises during the first 5 years or so after seroconversion but then declines, as demonstrated by the London and Edinburgh cohorts. Strong evidence of tracking in serial $\log_e$ IgA and $\log_e \beta_2$M measurements within individuals was found, as for $\sqrt{CD4+}$ cell count. It is, of course, possible that the ultimate fall in $\log_e \beta_2$M levels reported by the MAP workshop could be the result of zidovudine therapy, but a more basic expla-

nation would be that T-cell activation had peaked, resulting in lower β_2M levels. This phenomenon has been described previously for the expression of Class II major histocompatibility complex on T cells, which rises progressively during the natural history of HIV infection until the late stages when it declines [34].

Non-monotonicity in the functional forms for $\log_e$ IgA and $\log_e$ β_2M has implications for relative risks regression in cohort studies of progression to AIDS as well as for surveillance. Failure to allow for decreasing $\log_e$ β_2M levels may lead to an under-estimation of its prognostic significance; this caveat applies to the results from Part I of the workshop [35].

The MAP workshop focussed on covariate information available from anonymous HIV serosurveillance programmes, i.e., sex, age group, mode of HIV transmission and calendar year. Calendar year played an inadvertent role because of Edinburgh's change of methods in 1990 for determining IgA and β_2M. Age at seroconversion strongly influenced the seroconversion level for $\sqrt{\text{CD4+}}$ cell count, $\log_e$ IgA and $\log_e$ β_2M. In contrast, age appears to have little effect on CD4+ cell counts in HIV-antibody-negative individuals [36]. Annual change in marker level, even after the first year, was dramatic for $\sqrt{\text{CD4+}}$ cell count, being equivalent to 15 years' ageing, whereas for $\log_e$ IgA and $\log_e$ β_2M annual change equated to around only 4 years of additional age.

Higher CD4+ cell counts were not observed among women, but women had significantly lower IgA and β_2M levels, by 38.1 (13%) and 0.22 (8%), respectively, compared with a 30-year-old, heterosexually infected Italian man. Although the effect of sex on CD4+ cell count was not significant, it was comparable with that for HIV antibody-negative individuals [36]. Whether pregnancy in HIV-infected women further affects their marker values needs to be determined, since samples from antenatal women feature strongly in unlinked anonymous testing programmes.

Sexual versus other modes of HIV transmission (injecting drug use or the receipt of blood products) influenced all three markers but in diverse ways. $\sqrt{\text{CD4+}}$ cell count was significantly lower at seroconversion in patients who had been infected sexually (about 15% on the CD4+ cell scale); however, the subsequent decline in $\sqrt{\text{CD4+}}$ cell count was less marked (Table 2) than in patients who acquired their HIV infection by injecting drug use [31] or through contaminated blood products. Sexual transmission was associated with significantly higher $\log_e$ IgA levels at seroconversion. For $\log_e$ β_2M, IDU had higher levels at seroconversion as has been noted previously [5,6]. Presumptive mode of HIV transmission is clearly influential in characterizing marker paths.

Edinburgh's change to diffusion assays for determining IgA and β_2M levels had a number of implications for analyses. By the diffusion method, we could not demonstrate the initial fall in $\log_e$ IgA which was shown consistently across other cohorts and for Edinburgh's IgA determinations by the previous method. IgA determinations were also lower by diffusion assay, which is known to underestimate the quantification of dimeric IgA. The difference in functional form could simply be because of fewer IgA values within 2 years of seroconversion by the diffusion method, or because of a reduction in dimeric IgA early in HIV disease [37]. The MAP data may stimulate further immunologic investigation of the role of IgA in HIV disease. For β_2M, a straightforward additive adjustment between methods was adequate. The new method resulted in lower values, although there was no strong directional prior for this.

Covariate, geographic and method adjustments were made for all eligible ($\log_e$ IgA, $\log_e$ β_2M) pairs in the Edinburgh, Italian and London cohorts, and the adjusted marker pairs thereafter sorted into training and test data sets (1500 and 440 samples, respectively). Ordered logistic discriminations were used to categorize HIV infections as recent, intermediate or distant. For public-health surveillance purposes, the proportion of recent HIV infections, about which little can be inferred from AIDS backcalculation [18] could thus be estimated. Prediction was poor, however, even in the training dataset. This was due in part to the intensity of backward as well as forward transitions between marker states and the considerable heterogeneity between individuals in their marker trajectories. Such heterogeneity might be associated with the diversity of patients' human lymphocyte antigen phenotype or HIV genotype.

The strategic implications of these findings are twofold and focus attention back to CD4+ cell counts. In the interests of public-health surveillance it should be considered whether the design of unlinked anonymous testing programmes could be modified to accommodate rapid CD4+ cell count determination without breaching patient anonymity or whether technical developments can overcome the present time constraints on the handling of blood samples for CD4+ cell determination; second, whether the surveillance network can be extended, as in Scotland [38], to link first CD4+ cell count (or first pair of CD4+ cell counts [39]) from immunology laboratories with new HIV diagnoses [38] from virology laboratories. Further analytic work is needed on (changes in) test seeking behaviour before comparison of first CD4+ cell counts between groups of newly diagnosed HIV infections can be interpreted as an inference about the time of HIV infection prior to HIV diagnosis, or informing us about those who have not referred themselves for HIV testing.

Acknowledgements

References

1. Kin K, Kasahara T, Itoh Y, Sakurabayashi I, Kawai T, Morita M: β_2-microglobulin production by highly purified human T and B lymphocytes in cell culture stimulated with various mitogens. *Immunology* 1979, 36:47–54.

2. Evrin PE, Wibell L: Serum β_2-microglobulin in various disorders. *Clin Chim Acta* 1973, 43:183–187.

3. Karlsson FA, Wibeel L, Evrin PE: β_2 microglobulin in clinical medicine. *Scand J Clin Lab Invest* 1980, 40 (suppl 154):S27–S37.

4. Cooper EH, Forbes MA, Hambling MH: Serum β_2-microglobulin and C reactive protein concentrations in viral infections. *J Clin Pathol* 1984, 37:1140–1143.

5. Beorchia S, Vincent C, Revillard JP, Trepo C: Elevation of serum β_2-microglobulin in liver diseases. *Clin Chim Acta* 1981, 109:245–255.

6. Flegg PJ, Brettle RP, Robertson JR, Clarkson RC, Bird AG: β_2-microglobulin levels in drug users: the influence of risk behaviour. *AIDS* 1991, 5:1021–1024.

7. Howard MR, McVerry BA, Cooper EH: A longitudinal study of serum beta 2-microglobulin levels in haemophilia. *Clin Lab Haematol* 1988, 10:427–434.

8. Gorter RW, Vranizan KM, Osmond DH, Moss AR: Differences in laboratory values in HIV infection by sex, race, and risk group. *AIDS* 1992, 6:1341–1347.

9. Fling JA, Fischer JR, Boswell RN, Reid MJ: The relationship of serum IgA concentration to human immunodeficiency virus (HIV) infection: a cross-sectional study of HIV-seropositive individuals detected by screening in the United States Air Force. *J Allergy Clin Immunol* 1988, 82:965–970.

10. Heremans JF: Immunoglobulin A. In *The Antigens*, Vol II. Edited by Sela M. New York: New York Academic Press; 1974.

11. Rowe DS, McGregor IA, Smith SJ, Hall P, Williams K: Plasma immunoglobulin concentrations in a West African (Gambian) community and in a group of healthy British adults. *Clin Exp Immunol* 1968, 3:63–79.

12. Schechter MT, Craib KJP, Le TN, *et al.*: Susceptibility to AIDS progression appears early in HIV infection. *AIDS* 1990, 4:185–190.

13. Kramer A, Biggar RJ, Fuchs D, *et al.*: Levels of CD4+ lymphocytes, neopterin and β_2-microglobulin are early predictors of acquired immunodeficiency Syndrome. *Virology* 1990, 18:61–73.

14. Hofmann B, Wang Y, Cumberland WG, Detels R, Bozorgmehri M, Fahey JL: Serum β_2-microglobulin level increases in HIV infection: relation to seroconversion, CD4 T-cell fall and prognosis. *AIDS* 1990, 4:207–214.

15. Cuthbert RJG, Ludlam CA, Tucker J, *et al.*: Five-year prospective study of HIV infection in the Edinburgh haemophiliac cohort. *BMJ* 1990, 301:956–961.

16. Jacobson MA, Abrams DI, Volberding PA, *et al.*: Serum β_2-microglobulin decreases in patients with AIDS or ARC treated with azidothymidine. *J Infect Dis* 1989, 159:1029–1036.

17. Mastroianni CM, Paoletti F, Vullo V, Delia S, Sorice F: Serum β_2M in asymptomatic HIV-infected patients treated with zidovudine [letter]. *AIDS* 1990, 4:1297.

18. Public Health Laboratory Service Working Group: *Incidence of AIDS and Prevalence of Severe HIV-1 Disease, England and Wales 1992-1997: Predictions Using Data to End June 1992.* CDR, London: Public Health Laboratory Service; 1993.

19. PHLS AIDS Centre: Unlinked anonymous monitoring of HIV prevalence in England and Wales; 1990-92. *Communicable Dis Rep* 1993, 3:R1–R11.

20. MAP Workshop: Marker paths. *Stat Med* 1993 (in press).

21. Goldstein H: Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika* 1986, 73:43–56.

22. Goldstein H: Restricted unbiased iterative generalised least squares estimation. *Biometrika* 1988, 76:622–623.

23. Munoz A, Carey V, Schouten JP, Segal M, Rosner B: A parametric family of correlation structures for the analysis of 444 longitudinal data. *Biometrics* 1992, 48:733–742.

24. Galai N, Munoz A, Carey VJ, Chen K, Chmiel J, Zhou SYJ: Tracing of markers and onset of disease among HIV-1 seroconverters. *Stat Med* 1993, 12:2133–2145.

25. Longini I, Clark WS, *et al.*: Statistical analysis of the stages of HIV infection using a Markov model. *Stat Med* 1989, 8:831–843.

26. Lawless JF, Yan P: Some statistical methods for follow-up studies of disease with intermittent monitoring. In *Multiple Comparisons, Selection, and Applications in Biometry*. Edited by Hoppe FM. New York: Marcel Dekker; 1993:427–446.

27. Kalbfleisch JD, Lawless JF: The analysis of panel data under a Markov assumption. *JASA* 1985, 80:863–871.

28. Kalbfleisch JD, Lawless JF: Some statistical methods for panel life history data. In *Analysis of Data in Time*. Edited by A.C. Singh, P. Whitridge. Ottowa: Statistics Canada; 1991:185–192.

29. Lange N, Carlin BP, Gelfand AE: Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *J Am Stat Assoc* 1992, 87:615–626.

30. McNeil AJ: *Statistical Methods in AIDS Progression Studies with an Analysis of the Edinburgh City Hospital Cohort.* Cambridge: University of Cambridge; 1993.

31. Munoz A, Carey V, Saah AJ, *et al.*: Predictors of decline in CD4 lymphocytes in a cohort of homosexual men infected with Human Immunodeficiency Virus. *Acquir Immune Defic Syndr* 1988, 1:396–404.

32. MAP Workshop: Markov models. *Stat Med* 1993 (in press).

33. Weiss RA: How does HIV cause AIDS? *Science* 1993, 260:1273–1279.

34. Giorji JV, Detels R: T-cell subset alterations in HIV-infected homosexual men. NIAID Multicentre AIDS Cohort Study. *Clin Immunol Immunopathol* 1989, 52:10–18.

35. MAP Workshop: Markers as time dependent covariates in relative risk regression. *Stat Med* 1993 (in press).

36. Maini M, Gilson RJC, Adolphus N, Fakoya A, Ross E, Phillips AN: Normal range and sources of variability of CD4 counts in HIV-seronegative women and heterosexual men. *IX International Conference on AIDS*. Berlin, June 1993 [abstract PO-B43-2533].

37. Jackson S, Dawson LM, Kotlen DP: IgA1 is the major immunoglobulin component of immune complexes in the acquired immune deficiency syndrome. *J Clin Immunol* 1988, 8:64–68.

38. CD4 Collaborative Group: Use of monitored CD4 cell counts: predictions of the AIDS epidemic in Scotland *AIDS* 1992, 6:213–222.

39. Hoover DR, Graham NMH, Chen B, *et al.*: Effect of CD4+ cell count measurement variability on staging HIV-1 infection. *J Acquir Immune Defic Syndr* 1992, 5:794–802.

Appendix

The Multicohort Analysis Project workshop comprises: Graham Bird (Department of Immunology, Churchill Hospital, Oxford, UK); Richard Cook (Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada); Daniela De Angelis (PHLS AIDS Centre at Communicable Diseases Surveillance Centre and MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK); Vern Farewell (Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada); Katherine Fielding (MRC-BIAS Centre for HIV Research, Edinburgh, UK); Ami Fore (Division of Epidemiology, School of Public Health, Minneapolis, Minnesota,