Part 1
Transmissible diseases with long development times and vaccination strategies

Overview of Data Analysis: Diseases with Long Development Times

Sheila M. Gore

1 Introduction

Among diseases with long development times – from inception to diagnosis, from diagnosis to death, or both – feature breast and cervical cancer, end-stage organ failure and cardiovascular disease; all four have aspects in common with HIV disease, but its transmissibility by many routes sets it apart from the rest.

Breast and cervical cancers are detectable when asymptomatic (or precancerous) by screening; treatment of screen-detected lesions saves lives. Blood pressure lowering drugs reduce the incidence of stroke and coronary heart disease. An HIV antibody positive test leads to consideration of personal measures to prevent onward transmission of HIV disease and, from clinical management, to a better quality and length of HIV infected life, but as yet no cure. Cervical cancer, like HIV disease, is viral in origin and sexual transmission is implicated in its spread. Transplantation shares with HIV disease an immunological basis, recency, and unusual intensity of patient monitoring through laboratory markers.

Section 2 reviews the statistical problems posed by these other four applications, all of high public health or political profile, before consideration in Section 3 of chronic disease processes generally, and the transmissibility of HIV disease. Section 4 focusses on three data analytic themes in HIV disease – progression markers, incubation distribution and infectivity – and briefly reviews how they have been tackled. Future statistical directions in HIV disease are outlined in Section 5 with emphasis on transmission study design and overview, and on the non-proportionality over time of covariate influences. This includes unmeasured (frailty) as well as measured (and appropriately parametrized) covariates.

2 Review of applications: statistical problems posed

2.1 Breast cancer

Based on cases whose maximum follow-up from onset was six years, Greenwood (1926) wrote of breast cancer: 'At no observed epoch from onset is the

rate of mortality of the same order of magnitude as normal mortality' and noted also that the time-specific risk of death, or hazard, increased during the first three years after onset and was then constant or slowly declined. Boag (1949) observed that late relapse was not infrequent after five years in many forms of cancer and so 5-year survival rate, however defined, was an unsatisfactory estimate of the proportion permanently cured. Boag's view was that a proportion C of patients was cured by treatment and so liable only to causes of death other than the original cancer, but for the remaining patients time to death from breast cancer, distributed lognormally, was not modified by unsuccessful treatment. In 1975, Brinkley and Haybittle reported on the Addenbrooke's breast cancer series in which overall mortality was similar to that in the normal population after 21 years, but deaths from breast cancer were 16 times more likely in the study group (eight deaths reported). Because careful follow-up of breast cancer patients may bring benefit by the early detection and treatment of intercurrent illness, they warned: 'to define cure in terms of the general population mortality may have its limitations' (see also Langlands *et al.* 1979). And so, more than 50 years after Greenwood's astute observations on the 'National Duration of Cancer', no satisfactory description of the natural history of treated breast cancer had emerged. Curability was in doubt. Breast cancer is a chronic disease and age-related intercurrent deaths are not infrequent.

Standard regression models – Weibull, proportional hazards (Cox 1972), log-logistic – failed to represent the survival of patients with breast cancer in the Western General series (Gore *et al.* 1984) despite *agreement* on the relative influence of clinical covariates. The latter finding has been generalized by Solomon (1986), and by d'Agostino *et al.* (1990) to time-dependent covariates. The proportional hazards model accommodated neither the observed convergence of hazard functions nor diversity of times to peak hazard in breast cancer (see Figure 1). To investigate whether the influence of baseline covariates diminished in time, proportional hazards models were applied in three distinct epochs of follow-up, a different constant of proportionality being estimated for each. This led to a modified Cox model in which the influence of covariates was allowed to change smoothly with time according to a prescribed function. Differential frailty, through unmeasured covariates such as 'viable metastases' or otherwise, is a possible explanation for the observed hazard patterns, see Aalen (1988).

In the Western General series, covariate information was elicited only at the time of diagnosis so that patients' menopausal status, for example, was not updated during follow-up. Besides date and cause of death, only the date of first recurrence or metastasis (together with site(s)) was recorded – but not analysed because of retrospective data retrieval from case-notes. Recently, Zedeler *et al.* (1992) have considered the differential influence of

prognostic factors on the simultaneous occurrence of metastases at various anatomical sites in human breast cancer.

In the early 1980s breast cancer management was different between countries, between specialties, and by prognosis – with only one third of doctors claiming that treatment plans were distinguishable in terms of patient survival (Gore *et al.* 1988). Statistical overview of randomized trials in breast cancer had only just begun; it has now dispelled the need for radical surgery, shed light on increased long-term mortality associated with radiotherapy, and revealed a persistent survival advantage for women treated systemically by hormonal, cytotoxic or immune therapy (Early Breast Cancer Trialists' Collaborative Group 1992). Meanwhile, reduction in mortality from breast cancer after mass screening with mammography had also been established in randomized trials. Breast screening, at 3 year intervals, is now offered throughout the UK to women aged 50–65 years; research continues to optimize both screening interval and the management of screen-detected cancers, and to investigate screening at younger ages (Breast Cancer Screening 1986). Tabar *et al.* (1992) argue that approaches to therapy based on results obtained in clinically diagnosed, so-called 'early', breast cancers may be inappropriate for most screen-detected cancers – because the disease becomes systemic with viable metastases between the time at which it can be diagnosed by mammography and the time at which clinical diagnosis usually occurs.

The links between diet and breast cancer, randomized trial of systemic hormonal prophylaxis for women at high risk of breast cancer, and determining the genetics of familial breast cancer (Aalen 1992, Clayton 1991), are outstanding problems.

2.2 Cervical cancer

Cervical cancer is thought to originate with viral infection. It is rare in celibate women. Young age at first vaginal intercourse is a risk factor for disease progression, for reasons that may be associated with pubertal cervical physiology. Cervical screening aims to identify women with treatable cervical lesions which, if ignored, may develop into invasive cancer. Critical considerations in designing screening programmes include: smear accuracy in detecting underlying precursors, rates of progression, and the possibility of spontaneous regression. The management of women with mild or moderate dyskaryotic cervical smears remains controversial.

Kirby *et al.* (1992) concluded that mild or moderate dyskaryotic smears should not be an indication for immediate referral for colposcopy, since under Grampian's conservative management policy most women returned to normal without needing treatment, but the increased risk associated with abnormal smears justified rigorous surveillance. They had identified 500 women with

mild or moderate cervical dyskaryosis in 1978 or 1979 and 500, matched by age, who had had a normal smear at the same time. The standard statistical technique of censoring at the time of last smear could have led to serious bias in comparison of biopsy rates because last known smear result is to a certain extent informative about a woman's prognosis. After a negative smear, women were therefore assumed negative for three further years, the approximate mean inter-smear interval for control women, and censored thereafter. Analysis of the full screening history for these women included the use of graphical models and Gibbs sampling. Gilks *et al.* (1993) characterized woman i, who has had n_i smears, at time intervals t_{ij} between the $(j-1)$th and jth smear, as having smear results y_{ij} when the true underlying disease state was x_{ij}. The x_{ij} were unobserved realizations of a discrete state, continuous time, time-homogeneous Markov process, except for the last state $x_{i\,n_i+1}$ which was *observed* as a result of biopsy.

2.3 Transplantation for end-stage organ failure

The antecedents of organ failure vary: inborn errors of metabolism such as α_1-antitrypsin deficiency, progressive disease of adult onset such as diabetes, or acute events such as corneal trauma. Acceptance criteria for patients awaiting transplantation, and for donors, have changed dramatically over time and by centre; so too have results. Cadaveric kidney donors up to 70 years and heart donors up to 55 years are now accepted in the UK. Since 1985 patients with hepatic carcinoma have only been accepted for liver transplantation in Cambridge after laparotomy to exclude pre-existing metastases. However, before 1980, 45 out of 92 liver transplants were to patients with hepatic carcinoma and this diagnostic group had the highest risk of dying from three months after transplantation (Gore *et al.* 1987). In the second half of the 1980s, immunosuppression with Cyclosporin brought marked improvements in one year survival for all types of transplanted organ, as did beneficial matching in renal transplantation, and was therefore established as the immunological basis of kidney exchange in the UK.

Epoch-specific proportional hazards regression identified both the transient immunodominance of HLA matching (see Gilks *et al.* 1990 with corroboration by Thorogood *et al.* 1990) and that the risk of early renal graft failure (in the first two weeks post-transplant) was not reduced by Cyclosporin but was alleviated by beneficial matching. Neither tissue matching nor Cyclosporin has substantially reduced the 5% per annum hazard of late renal graft failure (that is: after the first year). Accelerated failure of regrafts (Gore and Bradley 1988), ascertainment bias and measurement error in DR typing were other features of note in transplantation studies. Ascertainment bias occurred at the inception of DR-typing, which was first applied in investigation of patients who had a failed renal graft. HLA-DR types, which were unknown at the time

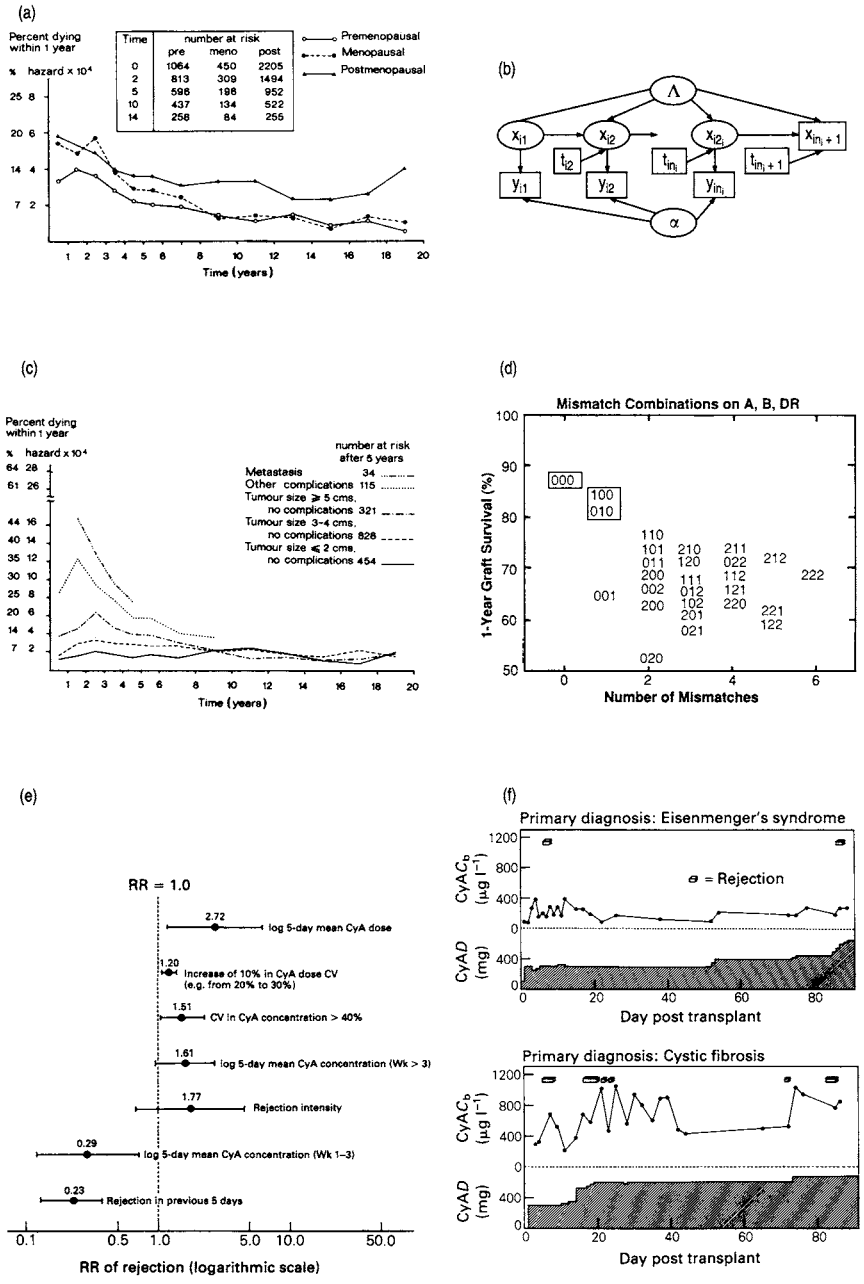

Figure 1: (a) Western General breast cancer series; (b) Cervical cancer: Gibbs sampling; (c) Western General breast cancer series; (d) HLA matching in kidney transplantation; (e) Risk of lung graft rejection; (f) Cyclosporin dose and blood levels.

of first graft, were subsequently infilled on the national database so that the strongest predictor of graft failure became knowledge of the recipient DR type! Such data-management errors are not uncommon in registry or clinical databases; moreover, it is often the case that new tests or drugs are used first in high risk patients. Unless due care is exercised, similar problems recur. Investigating the influence of HLA-phenotype on progression of HIV disease has been complicated by preferment for HLA typing of patients whose life expectancy is short. Moreover, there is serological failure of DR typing in 20% of patients whose HIV disease has progressed.

Reciprocal creatinine has been used successfully, but retrospectively, in Kalman-filter monitoring of post-transplant renal function (Smith and West 1983): slope change signalled imminent kidney graft rejection. But in liver transplantation, so-called 'biochemical rejection' is no longer a basis for bolus steroids being used, unless rejection is confirmed by liver biopsy. Short-term risks of nephrotoxicity and acute graft rejection in Papworth heart-lung transplantees are being assessed by the monitoring of Cyclosporin dose and blood levels, together with reciprocal creatinine (retrospectively in the first instance, but with a view to development of an online clinical decision aid). High coefficient of variation of individual Cyclosporin dose (and blood level) over the previous 10 days significantly increased the risk of lung graft rejection (Best *et al.* 1992).

In heart transplantation, patients who survive the short-term competing hazards of acute graft rejection, infection and nephrotoxicity – and 80% do – are then at risk of developing coronary occlusive disease, and so are monitored at predetermined intervals by angiographic assessment (see Gilks *et al.* 1993). The narrowing of major coronary arteries is graded on a three-point scale: normal, mild stenosis (50% or less narrowing) or severe stenosis; and analysis has featured a discrete state, continuous time, time-homogeneous Markov process with irreversible disease states.

2.4 Cardiovascular disease prevention and treatment

Assessment of the real strength of the relation between 'usual' blood pressure and the primary incidence of coronary vascular disease has required correction for the biassing effects of purely random variations in baseline blood pressure measurements. MacMahon *et al.* (1990) gave a lucid account of how to correct indirectly for regression dilution bias in prospective observational studies, but stressed the importance for future studies of remeasuring risk factors at a later follow-up in at least a representative proportion of survivors so that internal correction for regression dilution could be made.

After correction for regression dilution, and by considering several studies in combination, MacMahon *et al.* (1990) showed that usual blood pressure was positively related to the risk of stroke and of coronary heart disease,

not only among individuals who might be considered 'hypertensive' but also among those who would usually be considered 'normotensive'; and that these relations were at least 60% stronger than previously thought. MacMahon *et al.* thus set in an epidemiological context their subsequent overview of randomized drug trials involving short-term reductions in blood pressure (see Collins *et al.* 1990). Whereas the incidence of strokes was significantly reduced in randomized trials of blood pressure lowering drugs to the extent suggested by epidemiological studies, a lesser short-term effect was reported in respect of coronary heart disease: 14% reduction (but with wide confidence interval from 4% to 22%) instead of the 20 to 25% reduction suggested by observational epidemiological data. Whether intervention or follow-up was too short to observe the full epidemiologically-expected benefit, or the drugs insufficiently effective remains to be resolved. Correction for regression dilution biasses (by using repeat measurement of risk factors to determine usual levels) is, of course, relevant to other risk factors that may be subject to substantial measurement error, such as dietary characteristics (salt reduction for example: see Frost *et al.* 1991 and Law *et al.* 1991a,b).

Cardiovascular disease prevention and treatment has demonstrated cohesive interplay of epidemiology, statistical overview and trial design leading to good accounts of the scientific rationale for chosing between clinical trials (see Sandercock *et al.* 1986), advocacy of factorial designs, and *a priori* determination of plausible treatment effect sizes, thus ensuring that appropriate numbers of patients are randomized , on occasion over 40,000 as by ISIS-3 Collaborative Group (1992). Increasingly, strategies for preventing ischaemic heart disease are directed towards whole populations, with emphasis on changes in national diets, reduction in cigarette smoking and control or prevention of hypertension (but see also Shaper *et al.* (1986) and the OXCHECK Study Group (1991)).

2.5 Summary

Statistical science has contributed importantly to understanding the natural history of breast and cervical cancer, to the design and implementation of interventions such as mass screening and organ exchange policy, and to quantification of existing evidence from therapeutic trials, as well as to directing attention to treatment options which merit sound evaluation. Statisticians have also shown the importance of dealing with ascertainment and measurement error in assessing risk relationships; and that risk relationships commonly change over time, because of heterogeneity of frailty (Aalen 1988) or for other reasons including unmodelled treatment by covariate interaction. Statistical monitoring of marker processes for individualized acute patient managements remains largely a research activity; prospective evaluations and implementation are rare.

Other diseases with long development times are likely to pose interesting statistical problems in the 1990s. Besides HIV disease, there are the demographically compelling problems of aging, particularly dementia; the conundrum of early 'programming' by nature or nurture; and the esoteric prion diseases (proteinaceous infectious particle: see Brown *et al.* 1993). In humans, prion diseases include Creutzfeldt–Jakob disease (Weber *et al.* 1993) and, in animals, scrapie (affecting sheep and goats and prevalent in most sheep rearing countries except Australasia or Japan) and bovine spongiform encephalopathy (currently a major epidemic confined almost entirely to the cattle of the British Isles: Hughes (1993)).

3 Chronic disease processes generally, and HIV disease

3.1 Chronic disease processes

The applications of the previous section were chosen to illustrate in context some major data analytic themes which have exercised statisticians. These are now drawn out schematically in Figure 2 with chronic disease processes generally in mind. Briefly, from *inception* of the disease process there is an interval until the disease, or its progression, is detectable by *screening* (such as by cervical smear, blood pressure, HIV test or CD4 count). Inevitably screening is an imperfect (and often indirect) measure of the *true underlying disease state*, but the intensity of subsequent screening, behaviour modification (such as condom use, diet or exercise) and drug prophylaxis (to reduce blood pressure) may depend upon what was observed. *Clinical signs* are another insight to the underlying disease state, and the signs which manifest (Kaposi's sarcoma, for example) may both determine the speed of (clinical) *diagnosis* and anticipate future transitions between underlying disease states. There may be a choice of patient managements, medical or surgical, some of which are *major interventions* in the sense that the progression marker which has been monitored hitherto is removed (tumour size by mastectomy) or its relevance alters (reciprocal creatinine in dialysed or transplanted renal recipient). Moreover, after such major interventions the focus of patient monitoring may shift to include the *monitoring of drug toxicity* as well as efficacy (as in the case of Cyclosporin) or of *new disease processes*, such as coronary occlusive disease in long-term survivors of heart transplantation. It may thus be anticipated that *covariates exert different influences* in distinct epochs of follow-up in which different outcomes, or markers thereof, are relevant. Finally *external data*, such as from clinical trials or epidemiological studies, may be adduced to infer the likely size of treatment effects, or the times of infection (in the case of HIV disease). Covariates and marker processes may be *measured with*

error and important but unknown influences may be *unmeasured*.

3.2 Progression of HIV disease

Figure 2 is now interpreted in the context of HIV disease: inception is the time of infection, which is seldom known exactly but often an interval can be defined in which the individual must have become infected – the start of this interval being determined epidemiologically (by when HIV infections began in the region) or individually (according to behavioural or serological history). Antibodies to HIV disease form within weeks or few months of infection – so-called seroconversion – and can be tested for in blood, urine or saliva. A flu-like seroconversion illness may occur, but is either rare or recognised rarely for what it is. Individuals differ in their susceptibility to HIV infection. Those with the haplotype A1 B8 DR3, for example, have been reported as more susceptible to infection and, once infected, indeed progress more rapidly to AIDS (Steel *et al.* 1988). There may be differences in the virulence of the virus strain by which individuals were infected, and this may have implications for their disease progression, or infectivity (see below). External epidemiological studies (such as contact tracing), or self referral to 'check out' HIV status, or clinical symptoms may lead to an individual's being HIV tested, the result of which (positive or negative) may modify risk behaviours. Once HIV disease has been diagnosed, the patient's CD4 count – and other markers such as IgA, β_2-microglobulin and haemoglobin – are monitored at irregular intervals. Irregularity is occasioned, for example, by 20% failure to attend appointments (Brettle *et al.* 1992), intensification of monitoring for clinical trials (Ellenberg *et al.* 1992), or alteration in three-monthly schedule because of good (or poor) prognosis (Munoz *et al.* 1992). Prophylaxis against pneumocystis carinii pneumonia is instituted according to CD4 count, the underlying true disease state being unobserved. Clinical symptoms and opportunistic infections, tumours and HIV encephalopathy are managed as they manifest – for example foscarnet is given to avert cytomegalovirus retinitis, a cause of blindness (Studies of Ocular Complications of AIDS 1992). No major intervention in the sense described above for breast cancer or end-stage-organ failure is yet applicable in HIV disease (Stablein 1990). Treatment with zidovudine, initially offered only to patients with ARC or AIDS, is now advertised for patients whose CD4 count is 'less than 500 and rapidly falling'. Optimal strategy for zidovudine use is unclear from clinical trials, and so is also being investigated in clinical cohorts (McNeil 1993). Intercurrent death rate is around 2.5% per annum among HIV infected drug users, but some suicides and deaths by overdose may be HIV-related.