Statistical Science
1989, Vol. 4, No. 3, 264–281

An Annotated Bibliography of Quantitative Methodology Relating to the AIDS Epidemic

R. E. Fusaro, N. P. Jewell, W. W. Hauck, D. C. Heilbron, J. D. Kalbfleisch, J. M. Neuhaus and M. A. Ashby

Abstract. This paper provides an annotated bibliography of over 100 articles containing quantitative methodology relating to the AIDS epidemic. The majority of the work describes mathematical and statistical models of the growth and extent of the epidemic, and statistical procedures to estimate key components of the disease process. Among these components, attention has focused primarily on estimating the incubation distribution. It is hoped that the bibliography will not only interest those currently active in the field but also encourage other statisticians to become involved in AIDS research efforts. The general area of AIDS research appears to be a rich source of statistical problems of considerable interest and importance.

Key words and phrases: Acquired immunodeficiency syndrome (AIDS), epidemic models, human immunodeficiency virus (HIV), incubation distribution, infectivity, screening.

1. INTRODUCTION

The epidemic of AIDS cases and infections associated with the Human Immunodeficiency Virus (HIV) has now been in existence for a sufficiently long time that a substantial amount of data has been collected relating to many of its aspects. In order to interpret this information and to accurately predict the future course of the epidemic, an important body of quantitative research methodology has begun to evolve through the work of statisticians, epidemiologists and applied mathematicians. Indeed, there have been recent issues of statistical journals primarily dedicated to the review and promotion of this work. The collection of data related to AIDS is seldom routine and often suffers from incomplete ascertainment of key information, so that studies are rarely amenable to straightforward analyses. Thus, there is a continuing need for statisticians to aid in the analysis of data and in the planning of future studies. In addition, the

R. E. Fusaro is a Ph.D. candidate in Biostatistics at the University of California, Berkeley. His mailing address is: Group in Biostatistics, 140 Earl Warren Hall, University of California, Berkeley, California 94720.

N. P. Jewell is Professor of Biostatistics and Statistics at the University of California, Berkeley. His mailing address is: Group in Biostatistics, 140 Earl Warren Hall, University of California, Berkeley, California 94720.

W. W. Hauck is Professor in Residence at the University of California, San Francisco. His mailing address is: Department of Epidemiology and International Health, University of California, San Francisco, California 94143.

D. C. Heilbron is Adjunct Associate Professor of Biostatistics at the University of California, San Francisco. His mailing address is: Department of Epidemiology and International Health, HSW 1630, University of California, San Francisco, California 94143.

J. D. Kalbfleisch is Professor and Chair in the Department of Statistics and Actuarial Science, University of Waterloo. His mailing address is: Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, Canada, N1L 3G1.

J. M. Neuhaus is Adjunct Assistant Professor of Biostatistics at the University of California, San Francisco. His mailing address is: Department of Epidemiology and International Health, HSW 1630, University of California, San Francisco, California 94143.

M. A. Ashby is a Ph.D. candidate in Biostatistics at the University of California, Berkeley. His mailing address is: Group in Biostatistics, 140 Earl Warren Hall, University of California, Berkeley, California 94720.

general area appears to be a rich source of statistical problems of considerable interest and importance. It seems appropriate, therefore, to prepare an annotated bibliography of some of the methodological research that has been carried out to date. It is hoped that this will not only interest those currently active in the field but also encourage other statisticians to become involved in AIDS research efforts.

2. SCIENTIFIC BACKGROUND

Before briefly describing the topics included in the bibliography, it may be helpful to define some basic terminology. Figure 1 provides a somewhat simplistic picture of the natural history of the disease.

HIV (Human Immunodeficiency Virus) is now the standard name for the virus associated with AIDS. HIV antibodies can usually be detected in the blood shortly after infection, whereupon an individual has "seroconverted" and is thereafter referred to as "seropositive." The level of "infectivity" (i.e., the potential for transmission of the virus) of individuals with HIV disease is currently thought to vary and peak at seroconversion and the appearance of clinical symptoms of HIV disease. The "incubation" or "induction period" is the time from the occurrence of infection to the diagnosis of AIDS. This period is also referred to as the "latency period" by some authors, although the expression more accurately refers to the time between infection and seroconversion. AIDS itself is defined by a number of clinical life-threatening indications. It is, in fact, the advanced stage of HIV disease and many patients experience various symptoms of the disease prior to diagnosis of AIDS. These symptoms include generalized lymphadenopathy, various opportunistic infections and night sweats. Both "incubation" and "latency" suggest a silent disease up to diagnosis of AIDS and so are misleading in this respect. "Induction" is perhaps the best term, but "incubation" is most common. We have indicated the end of the disease process as death, since death from the disease is unfortunately the current prognosis for the overwhelming majority of AIDS cases.

For background reading on the immunology, epidemiology and other aspects of AIDS written at a level that is accessible to the scientific lay reader, we refer to a series of articles by Selwyn (1986), a recent report from the Centers for Disease Control (1987) and the October 1988 issue of *Scientific American* (volume 259), which contains ten articles outlining the major scientific issues relating to HIV infection and AIDS. Also of interest are review articles by Laurence (1985), Gallo (1987) and Levy (1988). Johnson (1988) surveys social and behavioral AIDS research in the gay, IV drug user, prostitute and general heterosexual populations. For bibliographies relating to all aspects of AIDS research, see *AIDS Bibliography* (University Publishing Group, Frederick, Maryland), Abrams (1987) and *AIDS Research Today* (BIOSIS Perspective Series, Biological Abstracts, Inc., Philadelphia, Pennsylvania).

3. CLASSIFICATION SCHEME

The articles in the bibliography are classified into five groups for convenience. These categories are necessarily broad, and the assigned codes should be considered as primary classifications. Occasionally, when the contents of an article substantially overlaps the described groups, multiple codes are assigned. The five groups are titled as follows:

I. General Models of Transmission Dynamics
II. Statistical Modeling of the Magnitude of the Epidemic
III. Natural History of AIDS
  A. Incubation Period
  B. Infectivity and Infection Distribution
  C. Survival Analysis and Disease Progression
  D. Association Between Cases
IV. Screening Tests
V. Clinical Trials

Group I contains the least statistical work since most of the AIDS epidemic models developed thus far are deterministic. We have attempted only to give an

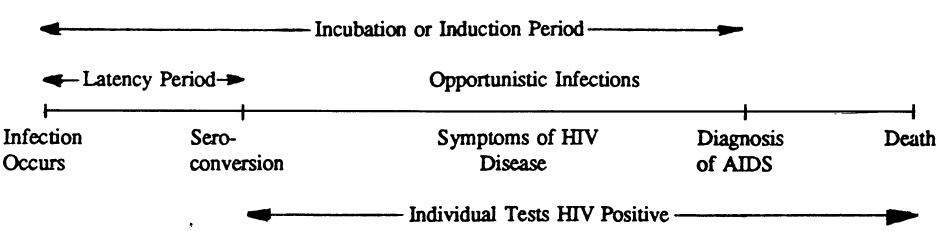

FIG. 1. *A simplified schematic of the natural history of HIV disease. Individuals with HIV disease are thought to be infective from the time of infection through to death, although the level of infectivity is thought to vary and be highest at seroconversion and at the appearance of symptoms. Diagnosis of AIDS corresponds to the development of certain specified life threatening conditions. Other infections and symptoms of HIV disease can and often do occur well before AIDS is diagnosed.*

overview of activity in this field, and also note here that Isham (1988) is an excellent review article. Future advances will center around the development of more empirically based epidemic models and stochastic models and will compare these with their deterministic counterparts.

Group II includes papers whose primary goal is estimation of the magnitude of the epidemic and short-term forecasting. These articles describe more empirically based methods than those of Group I and do not usually assume a dynamic model. Much of this work refers to the method of "back calculation," introduced by Brookmeyer and Gail (1986, 1988). This methodology obtains short-term forecasts by projecting the future number of AIDS cases from estimates of those already infected with HIV. It calculates the number of individuals previously infected from the known number of AIDS cases up to a point in time using a parametric model for the unknown infection times. The method assumes that the number of diagnosed cases has a distribution in chronological time that can be expressed as a convolution of the infection-time density and the incubation distribution and assumes full knowledge of the incubation-period distribution. These ideas are closely related to the subsequent work of Bacchetti and Moss (1989) mentioned below, although the two approaches assume different knowns and unknowns in the convolution equation.

Papers in Group III present statistical models of the components of the HIV disease process, which are key to the models of Groups I and II. These components include infectivity, the incubation period and the expression of the disease. Attention has largely focused on understanding the incubation period, which is covered in Group IIIA.

The distribution of the incubation period is particularly hard to estimate, in part due to its length (although the passage of time will help to solve this), but also due to the nature of the infected cohorts being followed or retrospectively identified. These cohorts include (a) prevalent infected homosexual men, (b) individuals infected by blood transfusion, (c) hemophiliacs and (d) pediatric cases. As yet there is little data on the incubation period for cases associated with intravenous drug abuse. Statistical efforts have involved primarily the first two groups and have concentrated on problems associated with the nature of ascertainment. For example, among prevalent cohorts of homosexual men, the time of infection is usually unknown and short-term incubation periods tend to be undersampled. Various imputation methods have been suggested to deal with the first of these issues; see, for example, Taylor, Schwartz and Detels (1986), Muñoz, Wang, Bass, Detels, Kingsley, Chmiel, and Polk (1989) and Jewell and Kalbfleisch (1989). A recent paper of Bacchetti and Moss (1989) estimates the incubation-period distribution by deconvoluting the observed distribution of chronological times of AIDS diagnoses using information on the distribution of HIV infections. Since transfusion-associated AIDS cases are retrospectively ascertained, longer incubation times are substantially undersampled (see Medley, Anderson, Cox and Billard, 1987; Medley, Billard, Cox and Anderson, 1988; Kalbfleisch and Lawless, 1988, 1989a, b). Several papers in this section allow estimation procedures to depend on covariates.

To date, only a few statistical papers examine the other components of the disease process such as infectivity and infection rates (Group IIIB) and survival of AIDS cases (Group IIIC). Group IIID describes papers on the analysis of contact networks or associations between cases.

Finally, Group IV contains several articles concerning the screening of individuals and the reliability of HIV testing procedures. Group V includes a paper that discusses the analysis of clinical trials of AIDS treatments.

The bibliography is not intended to be exhaustive but is aimed at highlighting central papers in these areas. It contains very few substantive papers where the statistical ideas are standard or straightforward. Generally, we have not included the many reports of governmental agencies. We have not included papers that deal with estimation and projection of the economic costs of AIDS and HIV disease. Interested readers may consult the paper by Hay, Osmond and Jacobson (1988) and the references contained therein.

4. Data Sets

A number of large data sets and studies have played an important role in understanding the natural history of AIDS in the United States. A selected list of such studies that figure prominently in the bibliography follows:

(1) *The San Francisco City Clinic Cohort or Hepatitis B Cohort*: A sample of homosexual men recruited from patients at the City Clinic to participate in Hepatitis B vaccine trials, some subsets of which have been followed over time. See Jaffe et al. (1985);

(2) *The San Francisco General Hospital Cohort*: A mixture of samples of homosexual men, some from STD (sexually transmitted diseases) clinics, some randomly selected, and some partners of AIDS cases, all followed over time. See Moss, Osmond, Baccheti, Chermann, Barre-Sinoussi and Carlson (1987);

(3) *The Multicenter AIDS Cohort Studies (MACS)*: a set of volunteer samples of homosexual men from selected urban areas throughout the

United States, followed over time. See Chmiel et al. (1987);

(4) *The San Francisco Men's Health Study*: a probability sample of homosexual men from San Francisco, followed over time. See Winkelstein et al. (1987);

(5) *CDC Transfusion Associated AIDS Cases*: a retrospective follow-up of AIDS cases associated with blood transfusions. See Lui, Lawrence, Morgan, Peterman, Haverkos and Bregman (1986) and Medley, Anderson, Cox and Billard (1987);

(6) *The National Cancer Institute Multicenter Hemophiliac Cohort*: a follow-up study of hemophiliacs from several hemophilia treatment centers in the United States. See Brookmeyer and Goedert (1989).

Of course, there are many other studies that involve these risk groups and the other major risk groups identified thus far, which include intravenous drug abusers, infants born to infected mothers, and heterosexuals. In addition, several national United States surveys are in the planning stages and it is clear that considerable statistical effort will be necessary throughout the planning and analysis of these investigations. Of related interest is a recent article by Layne, Marr, Colgate, Hyman and Stanley (1988) which advocates the creation of a national HIV database to provide epidemiologists, sociologists and mathematical modelers with raw data summarizing biological and behavioral parameters pertaining to the AIDS epidemic.

Several references in the bibliography are not annotated. These are articles cited elsewhere that we have been unable to obtain, but have included for completeness. The lack of an annotation in no way reflects on the importance of these papers. We apologize to authors for omissions from the bibliography or any misrepresentation of their work. Naturally we welcome appropriate additions to the list that we have missed or appeared after our deadline for inclusion which was essentially January 1, 1989.

ACKNOWLEDGMENT

This research was sponsored by SIMS (Societal Institute of the Mathematical Sciences) under support from the National Institute on Drug Abuse (NIDA Grant DA04722).

REFERENCES

To avoid unnecessary repetitions, references which appear in the annotated bibliography are not listed here.

ABRAMS, E. J. (1987). *AIDS Bibliography 1986–87*. U. S. Department of Health and Human Services.

CENTERS FOR DISEASE CONTROL (1987). Human immunodeficiency virus infection in the United States: A review of current knowledge. *MMWR* **36** (suppl. S-6) 1–48.

CHMIEL, J. S., DETELS, R., KASLOW, R. A., VAN RADEN, M., KINGSLEY, L. A., BROOKMEYER, R. and THE MULTICENTER AIDS COHORT STUDY GROUP (1987). Factors associated with prevalent Human Immunodeficiency Virus (HIV) infection in the Multicenter AIDS Cohort Study. *Amer. J. Epidemiol.* **126** 568–577.

GALLO, R. C. (1987). The AIDS virus. *Scientific American* **256** (1) 46–56.

HAY, J. H., OSMOND, D. H. and JACOBSON, M. A. (1988). Projecting the medical costs of AIDS and ARC in the United States. *J. Acquired Immune Deficiency Syndromes* **1** 466–485.

JAFFE, H. W., DARROW, W. W., ECHENBERG, D. F., O'MALLEY, P. M., GETCHELL, J. P., KALYANARAMAN, V. S., BYERS, R. H., DRENNAN, D. P., BRAFF, E. H., CURRAN, J. W. and FRANCIS, D. P. (1985). The acquired immune deficiency syndrome in a cohort of homosexual men. *Ann. Intern. Med.* **103** 210–214.

JOHNSON, A. M. (1988). Social and behavioral aspects of the HIV epidemic—a review (with discussion). *J. Roy. Statist. Soc. Ser. A* **151** 99–119, 125.

LAURENCE, J. (1985). The immune system in AIDS. *Scientific American* **253** (6) 84–93.

LAYNE, S. P., MARR, T. G., COLGATE, S. A., HYMAN, J. M. and STANLEY, E. A. (1988). The need for national HIV databases. *Nature* **333** 511–512.

LEVY, J. A. (1988). Mysteries of HIV: Challenges for therapy and prevention. *Nature* **333** 519–522.

MOSS, A. R., OSMOND, D., BACCHETTI, P., CHERMANN, J.-C., BARRE-SINOUSSI, F. and CARLSON, J. (1987). Risk factors for AIDS and HIV seropositivity in homosexual men. *Amer. J. Epidemiol.* **125** 1035–1047.

SELWYN, P. A. (1986). AIDS : What is now known. I. History and immunovirology; II. Epidemiology; III. Clinical aspects; IV. Psychosocial aspects, treatment prospects. *Hospital Practice* May 15, 67–82; June 15, 127–164; September 15, 119–153; October 15, 125–164.

WINKELSTEIN, W., JR., LYMAN, D. M., PADIAN, N., GRANT, R., SAMUEL, M., WILEY, J. A., ANDERSON, R. E., LANG, W., RIGGS, J. and LEVY, J. A. (1987). Sexual practices and risk of infection by the Human Immunodeficiency Virus. The San Francisco Men's Health Study. *J. Amer. Med. Assoc.* **257** 321–325.

ANNOTATED BIBLIOGRAPHY OF QUANTITATIVE METHODOLOGY RELATING TO THE AIDS EPIDEMIC

ABRAMSON, P. R. (1988). Sexual assessment and the epidemiology of AIDS. *J. Sex Research* **25** 323–346.

Reviews several HIV transmission models and finds that three broad categories of sexual parameters are crucial to each: sexual orientation, partner selection and manner of intercourse. Explores the feasibility of assessing these parameters and examines the reliability of existing data. (I)

ABRAMSON, P. R. and ROTHSCHILD, B. (1988). Sex, drugs, and matrices: Mathematical prediction of HIV infection. *J. Sex Research* **25** 106–122.

Rather than deriving models intended to accurately predict the development of the epidemic, this paper demonstrates how the spread of AIDS can be affected by parameters whose values are imprecise or unknown. Uses a simple epidemic model that

partitions the population according to (unspecified) physiological, behavioral or genetic cofactors. Assumes a constant rate of infection between these groups, proportional to the number of seropositives within each class. Suggests that each intergroup constant of proportionality may be the product of the expected number of contacts between a member of one group and members of the other multiplied by the probability that a contact results in transmission. Presents numerical examples illustrating that predictions based on this model are very sensitive to parameters that are currently difficult to estimate, such as the susceptibility of a group upon exposure and the rates of high risk behavior. (I)

AHLGREN, D. J., STEIN, A. C. and LYONS, P. A. (1987). Computer model of the AIDS epidemic. *Arch. AIDS Research* **1** 69–79.

Describes a compartmental model of HIV transmission via sexual contact, needle sharing and treatment of hemophilia with infected blood products. Reports that the model is implemented in the Apple Macintosh microcomputer language STELLA and is designed to allow planners, hospital administrators and educators to explore the impact of modified sexual behavior, sterilized needle usage, HIV testing programs and imperfect treatments and vaccines. (I)

ALLEN, A. D. (1988). Modeling AIDS and its treatment with immunostimulation. *Medical Hypotheses* **26** 55–58.

Models the pathophysiological effects of AIDS in terms of changing T4 to T8 cell ratios. Based on this model, correctly predicted the response of an AIDS patient to hyperimmunization with inactivated polio vaccine. (IIIC)

ALVORD, W. G., DRUMMOND, J. E., ARTHUR, L. O., BIGGAR, R. J., GOEDERT, J. J., LEVINE, P. H., MURPHY, E. L., JR., WEISS, S. H. and BLATTNER, W. A. (1988). A method of predicting individual HIV infection status in the absence of clinical information. *AIDS Research and Human Retroviruses* **4** 295–304.

Predicts infection status through latent structure analysis of data from four conventional HIV antibody assays. Reports nearly complete agreement with full clinical and laboratory diagnoses. Claims, therefore, that the simultaneous use of these four rapid, inexpensive assays, analyzed by these statistical methods, reduces the need for confirmatory Western blot assays. Also, employs latent structure analysis to determine the specificity and sensitivity rates of these assays absent previous definitive clinical or laboratory results. (IV)

ANDERSON, R. M. (1988). The epidemiology of HIV infection: Variable incubation plus infectious periods and heterogeneity in sexual activity (with discussion). *J. Roy. Statist. Soc. Ser. A* **151** 66–98; 124–125.

Suggests there may be two phases of peak infectiousness, each ranging in length from a few months to a year: the first occurring following the onset of primary HIV infection; the second immediately prior to AIDS emergence. Incorporates this two-episode infectious period in a simple deterministic transmission model. Argues that the doubling time in the initial epidemic depends mainly on the first episode, but that both phases determine the reproductive rate and the epidemic's magnitude and shape.

Discusses recent surveys of homosexual and heterosexual rates of partner change. Explores simple models describing observed frequency distributions of rates of partner change in these communities. Examines the implications of these models for future epidemiological research. (I)

ANDERSON, R. M. and MAY, R. M. (1988). Epidemiological parameters of HIV transmission. *Nature* **333** 514–519.

Reviews current epidemiological data for the parameters that drive HIV transmission within and between different risk groups. Examines sources of uncertainty regarding these parameters and the impact on predictions. (I)

ANDERSON, R. M., MAY, R. M. and MCLEAN, A. R. (1988). Possible demographic consequences of AIDS in developing countries. *Nature* **332** 228–234.

Investigates the impact of AIDS on demographic patterns in developing countries through deterministic models of transmission dynamics that incorporate demographic and epidemiological processes. Proposes a time-delayed recruitment model since HIV spreads predominantly among sexually active adults and births occur only to sexually mature females. Assumes homogeneous mixing and equal transmissibility between genders. Using a range of parameter values within bounds suggested by empirical studies, predicts that AIDS will result eventually in negative population growth rates in developing countries, although these effects may not become fully manifest for decades.

Also extends delayed recruitment model to reflect age structures. For currently plausible parameter values, predicts that AIDS will have little effect on the population dependency ratio, i.e., the ratio of the number of children and elderly to the number of adults. (I)

ANDERSON, R. M., MEDLEY, G. F., BLYTHE, S. P. and JOHNSON, A. M. (1987). Is it possible to predict the minimum size of the immunodeficiency syndrome (AIDS) epidemic in the d Kingdom? *Lancet* **1** 1073–1075.

Contrary to Brookmeyer and Gail (1986), argues that knowing the number of seropositive individuals and the heterogeneity of their sexual activity is essential in estimating the minimum size of the AIDS epidemic. Estimates total UK deaths using the model and methods described in Anderson et al. (1986) *infra*. Finds observed changes may be reproduced by various combinations of parameters, with predictions particularly sensitive to assumptions regarding the incubation period distribution, differences in frequency and pattern of sexual activity, and the proportion of seroconversions who later develop AIDS. Also contends that Brookmeyer and Gail employed an approximation of the underlying continuous process they examined rather than an exact equation. (II)

ANDERSON, R. M., MEDLEY, G. F., MAY, R. M. and JO SON, A. M. (1986). A preliminary study of the transmission dynamics of the Human Immunodeficiency Virus (HIV), the causative agent of AIDS. *IMA J. Math. Appl. Med. Biol.* **3** 229–263.

Describes preliminary deterministic models of the transmission dynamics of HIV infection in homosexual communities. Discusses previous modeling of other sexually transmitted diseases. Surveys existing epidemiological data to derive estimates for parameters such as duration of incubation period, proportion of seroconvertors who will develop AIDS, and infectious period.

Develops a series of models of increasing complexity, examining how various processes alter the course of the initial epidemic. Models the spread of AIDS in populations with homogeneous sexual practices, for both populations that are closed and ones with recruitment. Also accommodates variable incubation periods. Formulates more complex models reflecting heterogeneous sexual behavior and finds that high levels of heterogeneity decrease projections of the magnitude of the epidemic. Stresses that these models illustrate the effect of various parameters upon the infection's dynamics and thus should guide epidemiological research, but cautions that these models cannot yet generate reliable predictions regarding future trends. (I)

AUGER, I., THOMAS, P., DE GRUTTOLA, V., MORSE, D., MOORE, D., WILLIAMS, R., TRUMAN, B. and LAWRENCE, C. E. (1989). Incubation periods for pediatric AIDS patients. *Nature* **336** 575–577.

Analyzes the incubation period of maternally infected pediatric AIDS cases from New York State. Calculates a nonparametric maximum likelihood estimator of the incubation distribution, conditional upon developing AIDS within the observation interval, by a modified Kaplan-Meier product limit procedure that allows for delayed entry into the risk set. Reports an abrupt knee in this distribution, thus suggesting two risk populations. Calculates a parametric maximum likelihood estimate of the incubation distribution using a mixture of two Weibull distributions to capture this pattern. Also produces parametric and nonparametric estimates of the epidemic density curve, finding an increase in incidence consistent with exponential growth. (IIIA)

BACCHETTI, P. and MOSS, A. R. (1989). Estimating the incubation period of AIDS by comparing population seroconversion and incidence data in San Francisco. *Nature* **338** 251–253.

Estimates seroconversion rates for gay men in San Francisco from 1978–88, and then estimates the incubation distribution by comparing this to AIDS incidence over time in the same population. Uses the EM algorithm to perform the deconvolution, with a modification in the maximization step that produces smooth estimates. (IIIA)

BAILEY, N. T. J. (1988). Simplified modeling of the population dynamics of HIV/AIDS. *J. Roy. Statist. Soc. Ser. A* **151** 31–34.

Suggests that simple epidemic models may provide decision makers with prompt and yet sufficiently accurate information since many variables, such as sexual behavior, may have an overall average effect in a large population. Provides a simple HIV seroprevalence model that only includes parameters for effective transmission rate, initial HIV prevalence and fraction of the population engaged in safe behavior and thus not at risk. Estimates these parameters and appraises goodness of fit for the San Francisco City Clinic cohort.

Proposes a compartmental model for the incubation period, wherein an infected passes through several stages of immunologic compromise and then suffers an opportunistic infection. Presents a dynamic model for AIDS morbidity, assuming that all seroconverters develop AIDS. Estimates the effective transmission rate, proportion of population at risk, initial seroprevalence, mean of the compromise distribution, rate of AIDS onset following immunologic breakdown and the number of compartments. (I)

BARRETT, J. C. (1988). Monte Carlo simulation of the heterosexual selective spread of the human immunodeficiency virus. *J. Med. Virol.* **26** 99–109.

Describes a preliminary Monte Carlo simulation of short-term HIV transmission in a small heterosexual population. Generates each susceptible's constant infection risk per month of contact with an infected from a beta distribution. Also models partner change rates by a beta distribution. Examines the association between initial infection risk, the number of partners and final infection status. (I)

BLYTHE, S. P. and ANDERSON, R. M. (1988). Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). *IMA J. Math. Appl. Med. Biol.* **5** 1–20.

Incorporates variable incubation and infectious periods in a compartmental model of HIV transmission within a gay population assuming homogeneous mixing. Examines exponential, Weibull, gamma and rectangular distributions, and assumes that the infectious and incubation periods are equal. For each distribution, compares the model's properties by steady-state and local-stability analyses and by numerical methods. (I, IIIA)

BLYTHE, S. P. and ANDERSON, R. M. (1988). Variable infectiousness in HIV transmission models. *IMA J. Math. App. Med. Biol.* **5** 181–200.

Describes two methods for accommodating temporal fluctuations in infectivity during variable incubation periods in simple homogeneous mixing models of HIV transmission through a gay population. The first approach assumes that infecteds pass through a series of subclasses with different constant levels of infectivity, where the lengths of stay are exponentially distributed with different constant removal rates. The second method provides a mechanistic description of the relationship between the incubation period and infectivity based upon changes in viral abundance. Compares AIDS incidence projections over time generated by these models with forecasts that assume constant infectivity. Generally, finds that for models with two peaks in infectivity, the initial phase drives the early doubling time, while both episodes determine the overall magnitude of the epidemic and the level of the endemic equilibrium state. Discusses parameter estimation and future epidemiological research requirements suggested by formulation of these models. (I, IIIB)

BLYTHE, S. P. and ANDERSON, R. M. (1988). Heterogeneous sexual activity models of HIV transmission in male homosexual populations. *IMA J. Math. App. Med. Biol.* **5** 237–260.

Presents a proportionate mixing model of HIV transmission that describes the heterogeneity in homosexual partner change rates by treating sexual activity as a continuous variable in a set of integro-partial-differential equations. Develops a discrete approximation, which consists only of ordinary differential equations, by partitioning the range of sexual activity into discrete classes and then matching the equilibrium state and rate variables of the continuous model. Evaluates the accuracy of the approximation through numerical and analytic studies. (I)

BOLDSEN, J. L., JENSEN, J. L., SOGAARD, J. and SORENSEN, M. (1988). On the incubation time distribution and the Danish AIDS data. *J. Roy. Statist. Soc. Ser. A* **151** 42–43.

Fits a Weibull distribution to incubation-time data from U.S. transfusion-associated AIDS cases and plots the resulting contour curve. Obtains another estimate of the incubation distribution, based on San Francisco cohort data for seroprevalence and AIDS incidence, by comparing the expected number of cases under various Weibull distributions to reported cases via chi-squared statistics. Superimposes the contour curve for these chi-squared statistics on the transfusion AIDS curve and obtains an overall estimate of the Weibull parameters for the incubation period from the intersection.

For Danish, American, and Australian AIDS data, models number of infecteds by a logistic growth curve, and calculates expected AIDS incidence, assuming the incubation period follows this Weibull distribution. Estimates number of infected by comparing expected to observed AIDS incidence. (IIIA)

BONGAARTS, J. (1989). A model of the spread of HIV infection and the demographic impact of AIDS. *Statistics in Medicine* **8** 103–120.

Presents a compartmental model for long-term projections of the incidence and prevalence of HIV infection and AIDS and for assessing the epidemic's demographic impact. Employs a standard birth cohort component method of population projection for incorporating demographic characteristics, which also allows the epidemiologic and behavioral parameters in the model to vary with age. Further stratifies the population by gender, sexual behavior, marital status and infection status. Models the incubation process through four infection substates with exponentially distributed waiting times. This accommodates time-varying infectivity, and allows simulation of the entire model as a Markov chain. Points out that assuming equal risks of progressing through each substate results in a gamma distribution for the overall incubation period. Provides equations for estimating infection rates that assume random mixing among groups of individuals with multiple partners and frequent partner changes and that incorporate the effects of varying infectivity, transmission cofactors and condom usage. Reports the results of a computer simulation projecting the epidemic for 25 years in a hypothetical population with a Central African pattern of HIV transmission, i.e., no gay or IV drug user transmission. (I)

BROOKMEYER, R. and DAMIANO, A (1989). Statistical methods for short term projections of AIDS incidence. *Statistics in Medicine* **8** 23–34.

Discusses method of back-calculating number of individuals previously infected from AIDS incidence data through use of incubation distribution. Develops method for correcting incidence data for reporting delays and extends back-calculation method to account for new infectives in short-term projections of AIDS incidence, assuming that the infection rate remains constant in the immediate future. Applies this methodology to obtain short-term projections of cumulative U.S. AIDS incidence, assuming Weibull incubation distribution, adjusting for reporting delays and considering both step functions and log-logistic epidemic density models. (II)

BROOKMEYER, R. and GAIL, M. (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* **2** 1320–1322.

Proposes back-calculation method for estimating the magnitude of the AIDS epidemic without relying on knowledge of the number of seropositives or proportion of infectives that will develop AIDS. Instead, assumes the incubation period follows a Weibull distribution, citing Lui et al. (1986), and uses data regarding the increase in incidence over time. Employs the EM algorithm to solve for the distribution of infection times and ultimately for the total number of people infected prior to 1986 that will develop AIDS. Estimates the minimum size of the epidemic because the method projects cases only for those infected prior to the last incidence data. Uses Weibull and log-logistic distributions with various medians to investigate sensitivity of the methodology to the choice of the incubation period distribution. (II)

BROOKMEYER, R. and GAIL, M. (1987a). Biases in prevalent cohorts. *Biometrics* **43** 739–749.

Examines various biases resulting from investigating the natural history of AIDS through prevalent cohorts of seropositives whose time of infection is unknown. Considers the impact of using follow-up time in various time-to-event analyses instead of the unavailable time since infection, which is usually of primary interest.

Briefly discusses the one sample problem, but concentrates on the effects on fitting more general regression models, specifically the Cox proportional hazards model. For fixed covariates, shows that bias occurs when the time of infection distributions vary across the levels of the covariate ("onset confounding"). Points out that bias may persist, even absent onset confounding, due to nonconstant underlying hazard functions ("length biased sampling"). For time-dependent covariates, a bias may also occur when the infection distribution does not depend on the covariate ("frailty selection"). Covariates that change as a consequence of infection ("markers") are primarily subject to onset confounding since these markers are rarely associated with infection distributions that are constant across the level of the covariates. Provides qualitative descriptions of the directions of all three types of bias, based upon algebraic derivations. Calculations further reveal that the size and nature of the bias depends on the underlying hazard and infection distribution, and the relationship between these distributions and the covariates. (IIIA)

BROOKMEYER, R. and GAIL, M. (1987b). Methods for projecting the AIDS epidemic. *Lancet* **2** 99.

Responds to the criticisms of Anderson et al. (1987) claiming that their integral is in fact exact and that the critics' own results confirm that the incubation distribution is the preponderant influence in AIDS projections. Also suggests that their method may be useful in short term projections. (II)

BROOKMEYER, R. and GAIL, M. H. (1988). A method for obtaining short term predictions and lower bounds on the size of the AIDS epidemic. *J. Amer. Statist. Assoc.* **83** 301–308.

Provides theoretical foundations for the method of back calculation, described by Brookmeyer and Gail (1986), which uses AIDS incidence data to estimate the number of HIV infected. Considers the number of cases diagnosed in each calendar period to have a multinomial distribution. Assuming that the date of infection and incubation time are independent, expresses the cell probabilities as the convolution of the infection density (epidemic density) and the incubation distribution. Develops an EM algorithm for calculating the maximum likelihood estimate of the size of this multinomial distribution when the epidemic density is parameterized as a step function. Assumes that the incubation period distribution is known, but does not require knowledge of the proportion of infecteds that will develop AIDS. Applies this methodology to CDC AIDS incidence data through the end of 1986, adjusting for reporting delays and assuming a Weibull incubation distribution. Obtains short-term projections and estimates the minimum size of the epidemic, assuming no new infections occur subsequent to 1986.

Discusses several potential sources of error in back calculation. Reports that short-term projections are relatively robust to the shape and location of the Weibull incubation distribution. However, cautions that estimates of current HIV seroprevalence and long-term projections are highly uncertain: they are sensitive to assumptions regarding the incubation distribution and epidemic density; also, AIDS incidence data is less informative regarding the recent than the distant past.

Also discusses the method for estimating the incubation distribution from transfusion-associated AIDS data proposed by Lui et al. (1986). Indicates, for example, that Lui et al. present a likelihood equation that is actually conditional upon the time window of diagnosis and the transfusion date. Finds that for narrow windows this conditional maximum likelihood estimate is very inefficient relative to the full likelihood estimate. (II, IIIA)

BROOKMEYER, R., GAIL, M. and POLK, B. F. (1987). The prevalent cohort study and the acquired immunodeficiency syndrome. *Amer. J. Epidemiol.* **126** 14–24.

Companion article to Brookmeyer and Gail (1987a). Provides a qualitative discussion of technical results demonstrated in the Biometrics piece. (IIIA)

BROOKMEYER, R. and GOEDERT, J. J. (1989). Censoring in an epidemic with application to hemophilia-associated AIDS. *Biometrics* **45** 325–335.

Proposes two-stage parametric regression model for analyzing cohort study of an epidemic when exact time of infection is interval censored. Allows joint estimation of covariates' effect on risk of infection and on risk of progression to clinical disease.

For analyzing hemophiliac cohort data, employs a piecewise exponential regression model for calendar date at infection. This provides sufficient flexibility in modeling shape of hazard curve, obtained by nonparametrically estimating the cumulative distribution function of seroconversion times based on interval censored data. Considers Weibull regression model for time to AIDS following infection. Obtains maximum likelihood estimates and compares nested models using likelihood ratio test. Examines model adequacy by defining a sample of interval censored residuals, which are then treated as an interval censored sample from the unit exponential distribution. (IIIA)

BYERS, R. H., JR., MORGAN, W. M., DARROW, W. W., DOLL, L., JAFFE, H. W., RUTHERFORD, G., HESSOL, N. and O'MALLEY, P. M. (1988). Estimating AIDS infection rates in the San Francisco Cohort. *AIDS* **2** 207–210.

Estimates HIV infection rate in the San Francisco City Clinic hepatitis B cohort by fitting survival curves to the interval censored serologic data by maximum likelihood techniques. Compares the curves using the Akaike Information Criteria (AIC), finding that the log-logistic model fits the data somewhat better than the Gompertz, Weibull or logistic models. For comparison, also produced a life-table simulated survival curve, imputing the censored seroconversion times by first assuming the infection distribution to be uniform in the interval, choosing the best-fitting survival curve by life-table methods, and then imputing the infection time under the best survival distribution. (IIIB)

CASTILLO-CHAVEZ, C., COOKE, K., HUANG, W. and LEVIN, S. A. (1988). The role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS). In *Proc. International Symposium in Mathematical Approaches to Ecological and Environmental Problem Solving* (C. Castillo-Chavez, S. A. Levin and C. Shoemaker, eds.). *Lecture Notes in Biomath.* Springer, New York.

Cited in May, Anderson and McLean (1988). (I)

COLGATE, S., STANLEY, E. A., HYMAN, J. M., LAYNE, S. P. and QUALLS, C. (1989). A risk behavior based model of the cubic growth of AIDS in the United States. *Proc. Natl. Acad. Sci. U.S.A* **86**.

Argues that the cumulative number of U.S. AIDS cases has grown as the cube of time, after an initial transient, rather than exponentially. Presents a risk-based model, with biased mixing, for HIV transmission. This model reproduces the cubic growth of AIDS if the behavioral risk is distributed as an inverse cubic power of the number of individuals, such as when new partner frequency or sexual outlet frequency dominates the risk behavior. Also, assumes that the probability of progressing from seropositivity to AIDS is approximately constant over time.

Indicates that under this model, the apparent mean incubation period is an increasing function of time and, therefore, concludes that the decreasing rate of AIDS incidence is not yet due to behavioral changes. Also, indicates that the model implies that the mean risk behavior at infection is a decreasing function of time and thus that earlier AIDS cases engaged in riskier behavior. Calculates the average probability of infection per sexual contact under the model. Suggests that increased infectivity during the progression to AIDS can raise the growth rate to a higher power than cubic or at least mask the effects of lower overall risk behavior resulting from learning or the progression of the saturation wave to lower-risk groups.

Examines seeding models that describe the infection's passage to high-risk groups. Provides a numerical simulation of how an initial seeding in the average-risk group could spread the infection to the high-risk groups. (I)

DEGRUTTOLA, V. and LAGAKOS, S. W. (1989a). Analysis of doubly censored survival data, with application to AIDS. *Biometrics* **45** 1–11.

Presents a nonparametric, discrete time method for analyzing survival data that is doubly censored, i.e., when both the time origin and failure event may be right or interval censored. Obtains maximum likelihood estimates of the incubation period and infection time distributions by generalizing the self consistency algorithm for singly censored data proposed by Turnbull, B. W. (1976), *J. Roy. Statist. Soc. Ser. B* **38** 290–295. Demonstrates that for a broad class of censoring distributions, the expected log-likelihood equation has a unique maximum and the maximum likelihood estimator is consistent. Discusses a weakly structured continuous time parametric model that may be fit by this technique as well. Applies this method to survival data from a cohort of French hemophiliacs that received HIV infections through transfusions. (IIIA)

DEGRUTTOLA, V. and LAGAKOS, S. W. (1989b). The value of incidence data in assessing the spread of HIV infection. *Statistics in Medicine* **8** 35–43.

Examines the implications of the observed increase in doubling time for the AIDS epidemic in light of epidemic dynamics. Points out that saturation among high-risk groups and population heterogeneity (variation both among and within risk groups) tend to increase doubling times. Also, indicates that the induction time distribution appears to correspond to an increasing hazard rate, which also lengthens the doubling time. Thus, increasing doubling time alone does not indicate the extent to which behavioral changes have reduced the incidence rate, nor that the epidemic is beyond its peak. Argues that the extent of HIV infection cannot be accurately determined without knowing the shape of the infection distribution as well as the incubation distribution. (II)

DEGRUTTOLA, V. and MAYER, K. H. (1988). Assessing and modeling heterosexual spread of the Human Immunodeficiency Virus in the United States. *Rev. Infect. Dis.* **10** 138–150.

Presents a single-population epidemic model assuming a constant transmission rate per sexual act between infected and noninfected individuals. Compares this to models assuming constant risk of transmission over duration of relationship.

Generalizes model to describe two interacting homogeneous populations, one of which becomes infected much more rapidly, e.g., heterosexual IV drug users. Includes differential rates of intersexual transmission to account for the observation that most heterosexual AIDS patient had partners that used IV drugs or were bisexual. Also provides for nonrandom mixing depending upon drug usage. Incorporates a fixed lag in the start of the infectious period. Fits the two-population model using a range of parameter values obtained from epidemiologic studies. Reports that different assumptions about the lag in infectivity yield widely different projections of the magnitude of the heterosexual AIDS epidemic. Also discusses the role of epidemic models in analyzing case-control studies. (I)

DeGruttola, V., Seage, G. R., Mayer, K. H. and Horsburgh, C. R., Jr. (1989). Infectiousness of HIV between male homosexual partners. *J. Clinical Epidemiol.* To appear.

Estimates risk of HIV transmission per receptive anal sexual contact from a cross-sectional study of seropositive gay and bisexual men and their partners. Assuming constant risk for each exposure, fits a binomial model by maximum likelihood estimation. Since it is unclear how many contacts occurred following the unknown time of infection, performs several analyses truncating the total exposures at various levels. Reports that this model underestimates the number of infected that had few contacts and overestimates those with many, thus suggesting that infectivity may vary considerably between individuals. Accordingly, fits a model assuming that the risk per exposure was constant within each partnership, but that these risks were random variables sampled from a beta distribution. Parameter estimates, though relatively imprecise, suggest that infectivity is highly variable between individuals. Indicates that precise estimates of this variability, as well as quantification of the variability over time, require analysis of longitudinal data. (IIIB)

Denning, P. J. (1987). The science of computing: computer models of AIDS epidemiology. *Amer. Statist.* **75** 347–351.

Provides a brief nontechnical review of several HIV transmission models, focusing on those proposed by May and Anderson, Colgate et al. and Dietz. Discusses the history of epidemiologic modeling and the use of computers to evaluate complex mathematical models. (I)

Department of Health and Social Security, England (1987). *Future Trends in AIDS.* Her Majesty's Stationery Office, London (114 pages).

Contains the proceedings of a seminar sponsored by the English Department of Health and Social Security on March 23, 1987 in London. The lectures and discussions examine mathematical models and empirical extrapolation techniques for predicting the future of the AIDS epidemic and assess the data needs of such methods. (I, II)

Department of Health, The Welsh Office (1988). *Short-term Prediction of HIV Infection and AIDS in England and Wales. Report of a Working Group.* Her Majesty's Stationery Office, London (88 pages).

Employs empirical extrapolation, transmission dynamic models and back calculation methods to estimate the prevalence of HIV infection and to generate short-term projections of new AIDS cases in England and Wales. (I, II)

Dietz, K. (1988). On the transmission dynamics of HIV. *Math. Biosci.* **90** 397–414.

Presents a heterosexual transmission model, with homogeneous pair formation, that accounts for partnership duration and the number of contacts per partnership. Points out that classical models, on the other hand, assume that all same partner contacts occur instantaneously. Compared to such models, the pairing model predicts that the HIV infection will spread more slowly in the heterosexual population, under similar assumptions of partner change, per contact infectivity, and same partner contact frequency. Calculates the reductions in parameters, such as infectious period duration, infectivity, and contact, separation and pairing rates, that are necessary to reduce the reproductive rate to one. Outlines refinements to this model including the incorporation of heterogeneous contact rates. (I)

Dietz, K. and Hadeler, K. P. (1988). Epidemiologic models for sexually transmitted diseases. *J. Math. Biol.* **26** 1–25.

Points out that by pairing, two susceptibles gain a temporary immunity provided they have no outside sexual contacts. Extends a simple transmission model to account for pair formation by introducing explicit nonlinear pairing rates and separation rates. Assumes a constant rate of sexual contact and an exponentially distributed infectious period with constant infectivity throughout. Examines the relationship between various parameters and the existence of epidemic equilibria, concluding that equilibrium occurs only if the separation rate is large enough to ensure the necessary number of partners. (I)

Downs, A. M., Ancelle, R. A., Jager, H. J. C. and Brunet, J-B. (1987). AIDS in Europe: Current trends and short-term predictions estimated from surveillance data, January 1981–June 1986. *AIDS* **1** 53–57.

Analyzes European AIDS surveillance data reported to the World Health Organization by June 30, 1986, adjusted for reporting delays. Fits simple exponential models by regression methods over the entire period of the epidemic and over successive and overlapping 3-year time windows. Estimates doubling times based on this regression for the entire European Community and for ten countries individually. Also, predicts and constructs confidence intervals for the number of cases over one-half year periods by extrapolating from the curves for most the recent time windows (assuming constant doubling times). (II)

Eisenberg, B. (1989). The number of partners and the probability of HIV infection. *Statistics in Medicine* **8** 83–92.

Presents a simple model relating the probability of contracting HIV infection to the number of sexual contacts and partners. Assumes that partner choices are independent trials with a constant probability of selecting an infected partner. Similarly, assumes that successive contacts with an infective each present a constant risk, regardless of the type of contact. Examines numerical results for several values of these parameters. Employs various approximations to qualitatively compare the risks posed by different patterns of sexual contact. Finds that, for a constant total number of contacts, a monogamous relationship with a randomly selected partner poses a smaller risk than multiple contacts with randomly chosen partners. Also discusses a simple continuous time transmission model that assumes the number of contacts over time follows a Poisson distribution. (IIIB)

Gail, M. H. and Brookmeyer, R. (1988). Methods for projecting the course of the AIDS epidemic. *J. Natl. Cancer Inst.* **80** 900–911.

Reviews three methods for generating short term projections of the AIDS epidemic: empirical extrapolation, back calculation, and compartmental modeling. Compares the reliability of current data required by each method and the stability of their projections. (I, II)

Gail, M. H., Preston, D. and Piantadosi, S. (1989). Disease prevention models of voluntary confidential screening for human immunodeficiency virus (HIV) in isolated low risk and high risk populations and in mixed gay/heterosexual populations. *Statistics in Medicine* **8** 59–81.

Evaluates the potential benefits of voluntary confidential HIV testing through epidemic models separating susceptibles and infecteds into three compartments: those previously testing positive; those testing negative within previous year; and those not tested within previous year. The models also allow for further subdivision of the compartments into homogeneous subgroups based, for example, on sexual behavior.

Presents a transmission model for a single homogeneous population incorporating the following elements: initial population prevalence; transmission probability per new relationship between infected and susceptibles; average number of new relationships annually; effect of knowledge regarding HIV infectivity in preventing transmission. Generalizes simpler proportionate mixing models by accounting for differential attractiveness and transmissibility across subpopulations.

In general, finds that even assuming that knowledge of infectivity has low efficacy, voluntary testing of high-risk populations may provide a cost-effective means of retarding the spread of the epidemic, even within low-risk groups. (I, IV)

GANI, J. (1987). The challenge of AIDS modelling. *Appl. Probab. Newsletter.* Oper. Res. Soc. Amer., Baltimore, Md. **11** 1–3.

Presents a preliminary deterministic model of the AIDS epidemic assuming that some agent triggers AIDS symptoms in HIV carriers, e.g., a secondary viral infection such as hepatitis B. Provides equilibrium solutions and a numerical illustration. Indicates that this is one of many reasonable deterministic models for the epidemic and is intended to focus attention on the issues involved in such efforts. (I)

GASTWIRTH, J. L. (1987). The statistical precision of medical screening procedures: applications to polygraph and AIDS antibodies test data (with discussion). *Statist. Sci.* **2** 213–238.

Provides a large sample approximation to the standard error of the traditional estimator of the predictive value of a positive test (PVP), i.e., the probability that the subject is diseased given a positive result. Examines the ELISA test for AIDS and finds that the standard error of the estimated PVP increases as the prevalence in the population decreases. Discusses optimal sample size allocation for determining screening test accuracy. Concludes that moderately large samples are necessary for evaluating tests intended for groups with low prevalence rates. (IV)

GASTWIRTH, J. L. and HAMMICK, P. A. (1989). Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: application to estimating the prevalence of AIDS antibodies in blood donors. *J. Statist. Plann. Inference* **22** 15–27.

Describes a method for estimating seroprevalence that preserves donor anonymity by grouping samples and then testing these batches by a screening exam and confirmatory test where necessary. Derives a maximum likelihood estimator and a simple moment type of estimator of seroprevalence, assuming that the screening and confirmatory tests are independent and that the confirmatory test is perfect. Finds that for low prevalences, little information is lost by applying the moment rather than the maximum likelihood estimator.

Applies the moment estimator to Canadian AIDS data from individual testing by determining the probability distribution of the number of batches of a given size that contain a positive sample and then estimating the prevalence for each such outcome. Reports that the expected value of the moment prevalence estimate over this distribution essentially equals the results obtained by individual testing. Thus, since confirmatory tests are much more expensive than screening tests, and HIV prevalence is presumed low, in addition to preserving anonymity, this batch test method yields better seroprevalence estimates for the same total cost. (IV)

GIESECKE, J., SCALIA-TOMBA, G., BERGLUND, O., BERNTORP, E., SHULMAN, S. and STIGENDAL, L. (1988). Incidence of symptoms and AIDS in 146 Swedish haemophiliacs and blood transfusion recipients infected with human immunodeficiency virus. *Brit. Med. J.* **297** 99–102.

Estimates Kaplan-Meier survival curves for times to first symptom of HIV infection and to AIDS diagnosis for Swedish cohorts of hemophiliacs and recipients of HIV-infected transfusions. Imputes infection time for hemophiliacs, whose times are known only to within an interval, assuming constant and linearly increasing risks of infection. For comparison, fits Weibull distribution to the time to AIDS data. (IIIA)

GONZALEZ, J. J. and KOCH, M. G. (1987). On the role of "transients" (biasing transitional effects) for the prognostic analysis of the AIDS epidemic. *Amer. J. Epidemiol.* **126** 985–1005.

Argues that transients are phenomena associated with changes in the rate of spread of any infectious agent. For diseases with long and highly variable incubation periods such as AIDS, transients may mislead curve-fitting analyses of disease growth rates. Suggests that the apparent initial decline of the growth rate of AIDS (3–6 years) may be spurious, and the increase in doubling time may be actually caused by a positive, onset transient. Assumes that in initial stages the growth curve is well approximated by an exponential distribution, that most of the onset transient passes 5 years after the epidemic emerges and that a constant, unknown, proportion of infectives develop AIDS. Reports that this model fits early stages of the epidemic well. Also finds that, while the transient passes, changes occur in the observed incubation distribution, the age distribution of manifest cases, and the apparency of the epidemic, i.e., the ratio of cases to seroconversions. (I, II)

GRANT, R. M., WILEY, J. A. and WINKELSTEIN, W. (1987). Infectivity of the Human Immunodeficiency Virus: Estimates from a prospective study of homosexual men. *J. Infect. Dis.* **156** 189–194.

Estimates infectivity of HIV resulting from unprotected receptive anal intercourse, i.e., the per partner probability that such sexual unions with an infected partner will result in transmission of the agent. Assumes sexual partners are selected at random from a pool with known prevalence and thus treats the risk from multiple partners as a binomial experiment. Provides a maximum likelihood estimate of infectivity and suggests a moment estimator to supply an initial value for the iterative solution of the maximum likelihood equation. Also discusses a method for reweighting the reported number of partners to adjust for reporting delays. (IIIB)

GRIMSON, R. C. and GROSHEN, S. (1986). A statistical test for contact among individuals in an epidemic and a pattern among cases of acquired immune deficiency syndrome. *Statistics in Medicine* **5** 271–279.

Points out that to determine whether patient contact is a factor in disease transmission, it is important to ascertain the probability that an interviewed patient would randomly name one or more contacts that are also patients. Offers a model for ascertaining this probability based on the cumulative hypergeometric distribution. References earlier work in this area which appears in *Infectious Complications of Neoplastic Disease: Controversies and Management* (1985), Yorke Medical Books. (IIID)

HARRIS, J. E. (1987). Delay in reporting acquired immune deficiency syndrome (AIDS). Working Paper 452, Dept. Economics, Mass. Inst. Tech.

Provides maximum likelihood estimates, using an EM algorithm, of the empirical distribution of the delay in reporting AIDS cases. Examines the stationarity of the distribution.

Projects cumulative United States AIDS incidence by 1991 by extrapolating CDC data adjusted for reporting delays. (II)

HARRIS, J. E. (1988). The incubation period for human immunodeficiency virus (HIV-1). In *AIDS 1988: AAAS Symposia Papers* (R. Kulstad, ed.). Publication No. 88-19S, Amer. Assoc. Adv. Sci., Washington, D.C. (IIIA)

HEALY, M. J. R. and TILLET, H. E. (1988). Short-term extrapolation of the AIDS epidemic (with discussion). *J. Roy. Statist. Soc. Ser. A* **151** 50–65, 124.

To produce short-term predictions of number of AIDS cases, fits a variety of curves to the number of United Kingdom cases through 1986, adjusted for reporting lags. Models employed include an unweighted linear model using a log(no. cases + 1) scale and a log-linear model with Poisson errors, which yield similar forecasts. Also, in light of increasing doubling time in United States, fits a quadratic term in above models, which provide similar results over data period, but quite different forecasts.

Investigates change in slope by considering each month in 1985–86 as the last, and for each fitting a linear model for log(no. cases + 1) with weights decreasing geometrically into the past and a Poisson log-linear model on a 24-month time span. December 1986 versions of these models generate similar forecasts.

The evolution of empirical extrapolation leading to the ideas in this paper progressed through the following:

MCEVOY, M. and TILLET, H. E. (1985). Some problems in the prediction of future numbers of cases of the acquired immunodeficiency syndrome in the UK. *Lancet* **2** 541–542. Predicts United Kingdom AIDS cases for 1985–1988, assuming that epidemiological patterns remain unchanged, by extrapolating the results of a log-linear regression analysis of 1979–1984 annual incidence data.

ARTALEJO, F. R., ALBERO, M. J. M., ALVAREZ, F. V., LAGUARTA, A. B. and CABALLERO, J. G. (1986). Predicting AIDS cases (with reply). *Lancet* **1** 378–379. Contends that McEvoy and Tillet (1985) should have compared their log-linear predictions with those from polynomial models, since a quadratic model provided a better fit to Spanish AIDS data.

TILLET, H. E. and MCEVOY, M. (1986). Reassessment of predicted numbers of AID cases in the UK. *Lancet* **2** 1104. Revises prior projections to reflect updated AIDS incidence data, McEvoy and Tillet (1985). Recognizes that projections should account, however, for biasing transitional effects (see Gonzalez and Koch (1987)). (II)

HETHCOTE, H. W. (1988). A model for HIV infection and AIDS. In *Proc. International Symposium in Mathematical Approaches to Ecological and Environmental Problem Solving* (C. Castillo-Chavez, S. A. Levin and C. Shoemaker, eds.). *Lecture Notes in Biomath.* Springer, New York.

Briefly describes a preliminary series of HIV disease models, comprised of nonlinear ordinary differential equations, and outlines their intended use. Begins with a model of HIV transmission through a population divided into 14 compartments based upon sexual and needle-sharing behavior. Assumes that sexual contact between groups occurs according to mass action laws and that needle-sharing follows a proportionate mixing rule. Modifies the transmission model to reflect the progression of a fraction of infecteds to ARC, AIDS and death through a series of stages with exponential waiting times. Further adapts the model to incorporate perinatally and transfusion-acquired HIV infections. (I)

HYMAN, J. M. and STANLEY, E. A. (1988a). The effect of social mixing patterns on the spread of AIDS. In *Proc. International Symposium in Mathematical Approaches to Ecological and Environmental Problem Solving* (C. Castillo-Chavez, S. A. Levin and C. Shoemaker, eds.). *Lecture Notes in Biomath.* Springer, New York.

Presents a risk-based, deterministic model to examine the impact of social mixing on HIV transmission. Quantifies an individual's risk by his annual number of new sexual partners, and accounts for partnership duration by allowing a variable number of contacts per partnership. Stratifies seroconvertors and AIDS cases by time since infection or diagnosis. Models mixing patterns through a partnership acceptance function that allows varying degrees of like-partner biases. Numerical simulations indicate that for acceptance functions that allow some mixing between high- and low-risk groups, such as the inverse quadratic, the infection spreads to the low-risk groups earlier resulting in a faster epidemic given the larger sizes of these groups. Also reports that faster epidemics are predicted by models that allow infectivity and the rate of progression to AIDS to vary with time. (I)

HYMAN, J. M. and STANLEY, E. A. (1988b). Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* **90** 415–473.

Approximates cumulative diagnosed U.S. AIDS cases by a cubic equation in time. Expresses the cumulative diagnosed cases as the product of the probability an infected is eventually reported as an AIDS case times the convolution of the seroconversion rate and a Weibull incubation distribution. Solves for the seroconversion rate by the method of splines. Cautions, however, that since curve-fitting does not account for changing transmission dynamics, these methods provide unreliable long-term predictions.

Describes a transmission model for a homogeneous population, where partners are chosen randomly at a fixed rate from the entire population. Modifies this system to model purely heterosexual spread by splitting the population according to gender and accounting for partnership balance relationships. Expands the model to include time since infection and emergence of AIDS, since, for example, infectivity may vary widely with disease progression.

Further adapts these models to encompass risk behavior according to the number of new partners per year and assuming an average number of contacts per partnership. Assumes no migration between risk groups. Presents models assuming random partner choice from the entire population or biased partner selection. Specifically considers a strong selection preference towards partners from similar risk groups which gives rise to a diffusion model.

Presents numerical simulations illustrating the sensitivity of the nonrisk-based model to varying infectivity and compares the random and biased mixing models. Stresses that these models are too simplistic to accurately predict the course of the epidemic, but that they provide qualitative insights into the epidemic's dynamics, thus clarifying priorities for epidemiological research.

ISHAM, V. (1988). Mathematical modeling of the transmission dynamics of HIV infection and AIDS: A review (with discussion). *J. Roy. Statist. Soc. Ser. A* **151** 5–30, 44–49, 120–123.

Presents a comprehensive review of mathematical models of the sexual transmission dynamics of the AIDS epidemic. Discusses homogeneous mixing models reflecting the probability of infection per partnership, a constant rate of partner change and the probability that an infective will develop AIDS.

Assumes exponential distributions for the incubation periods of seropositives that will develop AIDS and for the infectious periods of those that will not develop AIDS. Also extends models to include heterogeneity of sexual activity and nonexponential incubation and infectious periods. Considers models for heterosexual populations as well. Includes an extensive bibliography. (I)

IVERSEN, O. and ENGEN, S. (1986). Epidemiology of AIDS—statistical analyses. *J. Epidemiol. Community Health* **41** 55–58.

Suggests that for United States transfusion AIDS data through 1984, the incubation period for cases emerging within 8 years follows a normal distribution truncated at 8 years. Employs standard numerical maximum likelihood procedures to estimate the mean and standard deviation of the distribution. Using this model, derives an integral equation model for the expected number of AIDS cases at time t in the City Clinic Cohort. Also, fits a log-linear model for the number infected by blood transfusion in the United States, with weights obtained from the preceding models, to obtain an estimated doubling time for infection. (II)

JACQUEZ, J. A., SIMON, C. P., KOOPMAN, J., SATTENSPIEL, L. and PERRY, T. (1988). Modeling and analyzing HIV transmission: The effect of contact patterns. *Math. Biosci.* **92** 119–199.

Presents a compartmental model of HIV transmission in a gay population categorized by frequency of sexual contact. Models incubation by a series of compartments, yielding a gamma distribution for the overall period and allowing for variable infectiousness. Assumes preferential mixing patterns whereby each group reserves a fraction of its contacts for members within the group and otherwise engages in proportionate mixing. Generalizes the classic definition of "reproductive number" to include heterogeneous populations. Explores the sensitivity of the model's steady state and dynamic behavior to changes in parameter values. Reports that a slight increase in the interaction between low- and high-contact groups significantly increases the rate and overall extent of HIV spread in the low-contact groups. (I)

JAGER, H. J. C. and RUITENBERG, E. J. (1988). *Statistical Analysis and Mathematical Modelling of AIDS*. Oxford Univ. Press, Oxford.

Contains the proceedings of a workshop conducted by the European Community Work Group held in Bilthoven, the Netherlands, December 15–17, 1986. The papers report early mathematical and statistical attempts to describe, predict and simulate the course of the AIDS epidemic. (I)

JEWELL, N. P. and KALBFLEISCH, J. D. (1989). Marker processes in survival analysis. To appear.

Considers the use of markers (i.e., variables that measure the progress of an individual towards a specified endpoint) in estimation of the incubation distribution. Assumes that such a marker is governed by a Poisson process and uses a simple additive model to relate the overall hazard (of incidence of AIDS, for example) at a point in time to a background hazard and the current state of the marker. Illustrates use of this model with censored follow-up data and in prediction of infection times for prevalent samples of infected individuals. Compares the efficiency of maximum likelihood procedures for estimation of parameters relating to the incubation distribution based on data without knowledge of the marker process to the efficiency of the same procedures with such information. (IIIA)

JEWELL, N. P. and SHIBOSKI, S. (1989). Statistical analysis of HIV infectivity based on partner studies. To appear.

Considers the statistical analysis of data from partner studies which yield information on the infection status of partners of individuals known to be infected with HIV, after a known number of contacts. Develops parametric and nonparametric procedures to examine and estimate the risk of infection after a known number of contacts. Graphical methods and inference techniques which allow evaluation of the assumption of constant infectivity per contact are presented. Considers the data in terms of a binary regression model with complementary log-log link which allows examination of the impact of heterogeneity of infectivity, error in measurement of the number of contacts and regression effects of other covariates. Illustrates methods with data on heterosexual contacts from the California Partners' Study. (IIIB)

JONES, M. E. and SOLOMON, P. J. (1988). A note on the concept of the incubation period in HIV infection. *J. Roy. Statist. Soc. Ser. A* **151** 40–42.

Discusses several problems with methods of back-calculating infection rates from AIDS incidence data and the incubation period. Suggests, for example, that the incubation density is probably not stationary given the rapid mutation of HIV and that neither is HIV incidence in light of changing sexual behavior. Also, urges modelers to consider the seronegative period, i.e., the time following HIV infection prior to production of antibodies. Suggests that the proportion of seronegatives among true infecteds may be surprisingly high, thus posing a potentially substantial threat of transfusion infection under current screening methods. Also suggests that the seronegative period for low dose infections, such as those induced by needlestick injuries, may be long in light of the small viral load transmitted. (II)

KALBFLEISCH, J. D. and LAWLESS, J. F. (1988). Estimating the incubation period for AIDS patients. *Nature* **333** 504–505.

Provides nonparametric analysis of blood transfusion AIDS data previously analyzed by Medley et al. (1987). Indicates identifiability problems resulting from retrospective ascertainment of infection times: the expected number of infections prior to a specific time and the incubation period distribution may be estimated only up to constants of proportionality. For example, transfusion data cannot discriminate between very high infection rates and long incubation times on one hand, and low infection rates and short incubation times on the other. An external estimate of the proportionality constant is necessary for a nonparametric analysis to place an upper bound on the median incubation period for adult and elderly patients. Also reports that, under the parametric assumptions employed by Medley et al., the identifiability problems persist, giving rise to very broad confidence intervals. Medley et al. respond and note that estimates based on an exponential infection rate and Weibull incubation times using more current data agree qualitatively with previous results. (II, IIIA)

KALBFLEISCH, J. D. and LAWLESS, J. F. (1989a). Estimating the incubation time distribution and expected number of cases for transfusion-associated acquired immune deficiency syndrome. *Transfusion*. To appear.

Discusses the important role of the reporting lag in interpreting the U.S. transfusion data when estimating the number of transfusion associated AIDS cases. Uses a similar model to Kalbfleisch and Lawless (1989b), but allows a nonparametric model for the reporting lag distribution. Reports that estimates

of the number of cases increase dramatically when the reporting lag is incorporated. Uses nonparametric and parametric analyses to forecast numbers of new cases. (IIIA)

KALBFLEISCH, J. D. and LAWLESS, J. F. (1989b). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84** 360–372.

Presents methodologies for analyzing data collected by retrospective ascertainment, which are implemented in providing detailed analyses of CDC transfusion related AIDS data. Considers retrospective ascertainment of events generated by a nonhomogeneous Poisson process where the distribution of time from initiating to final event is independent of time of initiating event. Obtains full and conditional likelihood equations. Parametric versions of this model were considered by Medley et al. (1987, 1988).

Performs nonparametric estimation of the intensity of the process of initiating events and of the incubation time distribution. Develops an efficient algorithm for fitting parametric models. Discusses estimability and identifiability issues for both parametric and nonparametric approaches. Explores methods for incorporating covariates in the incubation time distribution, including parametric regression modeling. Examines various generalizations to account for reporting lags, or for mortality of subjects through causes other than AIDS. (IIIA)

KAPLAN, E. H. (1989a). Modeling HIV infectivity: Must sex acts be counted? To appear.

Proposes an approximate maximum likelihood estimator for simple Bernoulli process models of HIV infectivity. Assumes a small, constant per sexual contact (or partner) transmission probability. Applies this approximate maximum likelihood estimator to two data sets: one contains the number of acts of unprotected vaginal intercourse between initially uninfected women and their seropositive husbands; the other records the number of partners for a cohort of homosexual men engaged in unprotected receptive anal intercourse (Grant, Wiley and Winkelstein (1987)). Constructs a nonparametric maximum likelihood estimator of the infectivity process via the pool adjacent violators algorithm of isotonic regression. Through simulations, compares the nonparametric model to the simpler Bernoulli models for the two data sets. Reports that the Bernoulli model appears to adequately describe infectivity per partner, but that such an approach may not be acceptable for modeling infectivity on a per contact basis. Discusses several probabilistic arguments that may explain the apparent inappropriateness of the constant infectivity per sexual act model. (IIIB)

KAPLAN, E. H. (1989b). Needles that kill: Modeling HIV transmission via shared drug injection equipment in shooting galleries. *Rev. Infect. Dis.* **11** 289–298.

Presents a series of models of HIV transmission among IV drug users via the sharing of injection equipment in shooting galleries. Initial assumptions include the following: all addicts select galleries in accordance with a Poisson process at a constant rate; equipment becomes infected when used by an infective, and is cleansed, with constant probability, when used by a susceptible; the usage of infected equipment transmits HIV with constant probability; the addict population size remains stable; and individual infectivity is constant. Derives a system of differential equations embodying these assumptions and examines the resulting dynamics numerically. Finds that the distribution of sterile equipment may delay the epidemic's spread.

Extends the model to account for heterogeneous equipment sharing rates. Numerical simulations reveal that increased heterogeneity escalates the infection's spread while decreasing the ultimate prevalence since the smaller group of more active shooters becomes saturated earlier (this is analogous to May and Anderson's (1987) conclusions regarding heterogeneous sexual behavior). Considers the effect of equipment cleaning and reports that even if addicts do not always bleach their works, such imperfect cleansing could stop the spread of HIV in galleries. Also examines the impact of natural inactivation of HIV over time. (I)

KAPLAN, E. H. (1989c). What are the risks of risky sex? Powerful insights from simple models. *Oper. Res.* **37** 198–209.

Constructs a model for the long-run prevalence of HIV in a population where cell members find partners at the same average rate for risky sex, i.e., unprotected receptive anal intercourse. Assumes random partner selection and constant lifetime infectivity. Also assumes exponentially distributed incubation times, but claims that the models are fairly robust to choice of incubation distribution. Constructs a system of rate balance equations and solves for HIV prevalence. Derives critical infectivity and risky sex rates, below which epidemics will not develop.

Extends model to incorporate variable rates of risky sex, while accounting for the random incidence phenomenon of stochastic processes. Contends that selectivity, the probability of selecting an infected partner, exceeds HIV prevalence, except in a sexually homogeneous population where these fractions are equal. Claims that greater variability in risk behavior facilitates the epidemic's spread, although for given mean risky sex rate and infectivity, greater variability in behavior implies a smaller epidemic.

Examines the impact of an imperfect immunizing vaccine through similar models, deriving a critical vaccine performance level. Illustrates these models through application to the cohort data of the San Francisco Men's Health Study (Winkelstein et al., 1987), and concludes, *inter alia*, that San Francisco gay men may have modified their sexual behavior sufficiently to greatly reduce HIV prevalence. Examines the sensitivity of HIV prevalence and selectivity estimates to assumptions regarding infectivity, risky sex rates, vaccine effectiveness and fraction of HIV infections progressing to AIDS. (I)

KAPLAN, E. H. and ABRAMSON, P. R. (1989). So what if the program ain't perfect? A mathematical model of AIDS education. *Evaluation Rev.* **13** 107–122.

Presents a simple deterministic HIV transmission model to examine the effects of an education program that causes some participants to abstain from high-risk behavior, such as unprotected receptive anal intercourse, for a time only to eventually recidivate to the same level of risky behavior. Assumes homogeneous rates of risky sex, a constant probability of transmission per contact and a stable population size. Presents numerical examples illustrating that even an imperfect education program can in the long run significantly retard HIV transmission. Suggests extending the model to incorporate the ancillary benefits in combating AIDS of reducing other sexually transmitted diseases. (I)

KAPLAN, J. E., SCHONBERGER, L. B. and LUI, K.-J. (1986). The use of the binomial distribution in establishing an association between high-risk donors and transfusion-associated AIDS. *Statistics in Medicine* **5** 355–362.

In the first seven cases of transfusion-associated AIDS, there was at least one high-risk donor in each donor set. Tests the hypothesis that there was an association between the high-risk donors and the AIDS cases by calculating the probability of finding that many high-risk donors by chance in the donor set, via the binomial distribution where p was the estimated prevalence of such donors in the overall donor population. (IIID)

KIESSLLING, D., STANNAT, S., SCHEDEL, I. and DEICHER, H. (1986). Überlegungen und Hochrechnungen zur Epidemiologie des Acquired Immunodeficiency Syndrome in der Bundesrepublik Deutschland. *Infection* **14** 217. (In German, with English summary.)

Describes a computer simulation of the morbidity and mortality rates of AIDS in the Federal Republic of Germany, where the population was divided into six groups of differing sexual behaviors. (I)

KNOX, E. G. (1986). A transmission model for AIDS. *Eur. J. Epidemiol.* **2** 165–177.

Presents a series of equilibrium models culminating in a system that divides the population into 12 classes based on gender, sexual preference and type of sexual behavior. Assumes risk of infection depends upon new partner acquisitions rather than on individual sexual contacts. Argues that the preferences of each sexual class create a market of supply and demand; thus, uses a computer model that reconciles the preferences of each group in estimating the frequency of partner changes and relative preferences between groups. Using these parameters, employs a computer simulation to calculate incidence and prevalence of HIV at epidemic equilibrium. Also examines the effect of global and selective behavioral changes and vaccination. (I)

KOOPMAN, J., SIMON, C., JACQUEZ, J., JOSEPH, J., SATTENSPIEL, L. and PARK, T. (1988). Sexual partner selectiveness effects on homosexual HIV transmission dynamics. *J. Acquired Immune Deficiency Syndromes* **1** 486–504.

Develops deterministic models to examine the impact of partner selectiveness on HIV transmission through a gay population divided into five groups based upon frequency of new sexual partnership formation. Separates the infection into five stages, thereby accommodating time-varying infectivity. Bases per partnership transmission probabilities upon the following subgroup-specific factors: per sexual contact transmission probability; average partnership duration; background and AIDS death rates; and sexual contact rate. Does not distinguish between types of sexual contacts, suggesting that the transmission probabilities be viewed as weighted averages across activities presenting differing risks. Models partnership selectivity assuming that individuals reserve a proportion of their partnerships for members of their own groups and otherwise mix randomly. Implements the model through the STELLA program for Apple Macintosh microcomputers and offers to supply copies of the model upon request.

Numerical simulations reveal that even modest contact with high-risk groups can dramatically accelerate the epidemic in groups with low rates of contact and partnership formation. Suggests that mixing with high-risk groups may result in relatively high infection rates in low-risk groups that would experience no epidemic if isolated. Simulations using data from the Coping and Change Study, conducted in collaboration with the Chicago MACS investigation, indicate that a second epidemic may appear in the population not engaged in casual and anonymous sex sometime after the epidemic in the higher-risk group has waned, with longer delays resulting from limited mixing. (I)

LAGAKOS, S. W., BARRAJ, L. M. and DEGRUTTOLA, V. (1988). Nonparametric analysis of truncated survival data with applications to AIDS. *Biometrika* **75** 515–523.

Determining the distribution of the incubation period from transfusion or pediatric AIDS data is complicated by the right truncation of the data, i.e., only those who develop AIDS by a specific time are counted. Proposes, therefore, considering infection and disease processes in reverse time, employing modified survival methods for data left truncated in internal time and then transforming results back to forward time. Modifies nonparametric maximum likelihood estimation methods for identifiable aspects of the incubation distribution. Also, adapts a weighted log rank test for comparing identifiable aspects of the incubation distributions of two groups. Applies these methods to transfusion data analyzed by Lui et al. (1986). Suggests ready extensions to data categorized by type of AIDS manifestation and to comparison of more than two groups. (IIIA)

LAGAKOS, S. W. and DEGRUTTOLA, V. (1989). The conditional latency distribution of AIDS for persons infected by blood transfusion. *J. Acquired Immune Deficiency Syndromes* **2** 84–87

Applies the nonparametric maximum likelihood estimation methods reviewed in Lagakos, Barraj and DeGruttola (1988) to CDC transfusion-associated AIDS data. For several subgroups, compares the estimated conditional incubation distribution for infecteds that develop AIDS within the observation period. Reports that the estimated distribution for infants is concave for 2 years after infection; suggests that since unimodal parametric models cannot capture this shape, a lack of fit may underlie their smaller estimates. As an alternative to making parametric assumptions about its shape, proposes estimating the unconditional distribution by combining the nonparametric estimate of the conditional distribution with a plausible range of values for the proportion of infecteds that develop AIDS within the observation period. (IIIA)

LONGINI, I. M., JR., CLARK, W. S., BYERS, R. H., LEMP, G. F., WARD, J. W., DARROW, W. W. and HETHCOTE, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine.* To appear.

Fits a five-stage, time-homogeneous Markov model by numerical maximum likelihood techniques to left, right and interval censored data from the San Francisco Hepatitis B cohort and cohorts of hemophiliacs and recipients of HIV-infected transfusions. Based on this model, estimates the incubation period density and survival times from each stage of infection. Compares these estimates to those reported by other investigators, discussing the differences in biases and censoring. (IIIA)

LUI, K.-J., DARROW, W. W. and RUTHERFORD, G. W. III (1988). A model-based estimate of the mean incubation period for AIDS in homosexual men. *Science* **240** 1333–1335.

Estimates the mean incubation period for AIDS in a cohort of 84 HIV infected gay and bisexual men from the San Francisco City Clinic cohort. Presents a likelihood equation for the time from seroconversion to AIDS diagnosis, assuming that this period follows a Weibull distribution and assuming that only a proportion, p, of infected gay men will develop AIDS. Calculates maximum likelihood estimates for the Weibull parameters and for p. Also provides results for a log-logistic incubation distribution. Compares these results to incubation period estimates based on transfusion-associated AIDS data. (IIIA)

LUI, K.-J., LAWRENCE, D. N., MORGAN, W. M., PETERMAN, T. A., HAVERKOS, H. W. and BREGMAN, D. J. (1986). A model-based approach for estimating the mean incubation period of

transfusion-associated acquired immunodeficiency syndrome. *Proc. Nat. Acad. Sci. U.S.A.* **83** 3051–3055.

Suggests that the traditional estimate of the mean incubation period, i.e., the simple average of observed incubation periods, is length-biased, due to right truncation, and thus tends to underestimate the true mean. Proposes a parametric model-based approach, truncating the right side of the distribution to remove observation bias resulting from cases with such long incubation periods that they have not yet been diagnosed. Then truncates the left side since some transfusion-AIDS cases may not have been properly diagnosed prior to 1982. Assumes that the incubation period follows a Weibull distribution and obtains maximum likelihood estimates of its parameters. Also provides estimates assuming log-logistic and Gompertz distributions. (IIIA)

LUI, K.-J., PETERMAN, T. A., LAWRENCE, D. N. and ALLEN, J. R. (1988). A model-based approach to characterize the incubation period of paediatric transfusion-associated acquired immunodeficiency syndrome. *Statistics in Medicine* **7** 395–401.

Estimates the mean incubation period for pediatric transfusion-associated AIDS cases through the truncated parametric model-based approach set out in Lui et al. (1986). Assumes the Weibull distribution as the basic parametric model for the incubation period distribution. (IIIA)

MAKUCH, R. W. and PARKS, W. P. (1988). Statistical methods for the analysis of HIV-1 core polypeptide antigen data in clinical studies. *AIDS Research and Human Retroviruses* **4** 305–316.

Describes statistical methods for analyzing dichotomized or continuous HIV-1 p24 antigen data from participants in a randomized clinical trial. Recommends standard nonparametric methods for analyzing within and between group differences based on p24 antigen serum levels. An alternative measure of HIV-1 virus expression is the number of days a patient's lymphocytes must culture before exhibiting virus positivity. Such data is subject to censoring when the culture requires a longer observation period than is practicable in the lab or when contamination occurs prior to observing positivity. Considers nonparametric tests for assessing within and between group differences in such censored data based on Prentice-Wilcoxon scores. Applies these methods to the results of a prospectively randomized, placebo-controlled study of AZT. Offers PC software written in TURBO BASIC that performs both large sample and exact permutation versions of these tests. (V)

MAY, R. M. and ANDERSON, R. M. (1987). Transmission dynamics of HIV infection. *Nature* **326** 137–142.

Presents a discussion of transmission dynamics similar to that of Anderson et al. (1986), but also considers heterosexual transmission. (I)

MAY, R. M., ANDERSON, R. M. and JOHNSON, A. M. (1988). The influence of temporal variation in the infectiousness of infected individuals on the transmission dynamics of HIV. In *Proc. International Symposium in Mathematical Approaches to Ecological and Environmental Problem Solving* (C. Castillo-Chavez, S. A. Levin and C. Shoemaker, eds.). *Lecture Notes in Biomath.* Springer, New York. (I)

MAY, R. M., ANDERSON, R. M. and MCLEAN, A. R. (1988a). Possible demographic consequences of HIV/AIDS epidemics. I. Assuming HIV infection always leads to AIDS. *Math. Biosci.* **90** 475–505.

Points out that most dynamic models of HIV transmission do not systematically explore the demographic implications of AIDS since they assume a constant population size except as diminished by AIDS-related mortality. Thus, present a series of deliberately oversimplified models to examine qualitatively the demographic impact of adult mortality and depressed birth rates resulting from AIDS. Begins with a simple model assuming constant infectiveness, constant disease progression rates and an identical average rate of new heterosexual partner acquisition. Provides a general analytic solution for this model. Qualitatively describes the dynamic properties of this and subsequent models through phase plane techniques and asymptotic analyses. Presents more refined models incorporating asymmetric heterosexual transmission, heterogeneous partner acquisition rates and population age structure, thus allowing investigation of AIDS' effect on age profiles. In each model, assumes that all infecteds eventually develop AIDS, absent prior death from other causes. (I)

MAY, R. M., ANDERSON, R. M. and MCLEAN, A. R. (1988b). Possible demographic consequences of HIV/AIDS epidemics. II. Assuming HIV infection does not necessarily lead to AIDS. In *Proc. International Symposium in Mathematical Approaches to Ecological and Environmental Problem Solving* (C. Castillo-Chavez, S. A. Levin and C. Shoemaker, eds.). *Lecture Notes in Biomath.* Springer, New York.

Cited in May, Anderson and McLean (1988a), as providing a largely numerical study of similar dynamic models combining demography with epidemiology, while allowing some infecteds to remain asymptomatic for life. (I)

MEDLEY, G. F., ANDERSON, R. M., COX, D. R. and BILLARD, L. (1987). Incubation period of AIDS in patients infected via blood transfusion. *Nature* **328** 719–721.

Analyzes transfusion-related AIDS data to determine distribution of incubation period. Postulates linear or exponential forms for growth in the number of infected who will develop AIDS and Weibull or gamma distributions for the incubation period. Also includes time-dependent probability of diagnosis in probit form. Obtains an explicit expression for the log-likelihood equation (not reported) and maximizes it numerically. Concludes that exponential growth with a doubling time of approximately 1 year fits the data better than a linear model. Finds that Weibull and gamma distributions produce equally good fits. Reports age- and sex-specific mean (and median) incubation periods based on Weibull and gamma models. (IIIA)

MEDLEY, G. F., BILLARD, L., COX, D. R. and ANDERSON, R. M. (1988). The distribution of the incubation period for the acquired immunodeficiency syndrome (AIDS). *Proc. Roy. Soc. London Ser. B* **233** 267–277.

Describes the model and methods employed to analyze U.S. AIDS transfusion data as summarized in Medley et al. (1987). Postulates a parametric nonhomogeneous Poisson model for new infections caused by blood transfusions and, independently, a parametric model for the incubation time distribution. A probit model specifies the probability of diagnosis as a function of time. Provides an explicit form for the likelihood function and results for estimation under linear and exponential forms of the Poisson rate and Weibull and gamma incubation time models. Introduces subgroups explicitly and tests for differences between them. Discusses limitations associated with the transfusion data and this method of analysis. (IIIA)

MORGAN, W. M. and CURRAN, J. W. (1986). Acquired Immunodeficiency Syndrome: current and future trends. *Public Health Reports* **101** 459–465.

Uses empirical models, based on U.S. CDC surveillance data for June 1981 through May 1986, to project the number and

distribution of AIDS cases by risk group, geography, gender, race, and age. Adjusts reported cases to estimate actual diagnosed cases. Fits a quadratic polynomial to these adjusted counts transformed by a modified Box-Cox method to provide homoscedastic weighted residuals. Projects these results to 1991 assuming the trends remain unchanged over time.

To estimate AIDS associated deaths, calculates survival times by the Kaplan–Meier method. Projects AIDS mortality assuming the cumulative survival times follow a negative exponential distribution. Tests changes over time of the distribution of diagnosed cases using chi-square tests for linear trends when examining proportions and Spearman rank correlation for continuous variables. Projects future trends in case distribution by applying weighted linear regression to the logit of these proportions. (II)

MUÑOZ, A., CAREY, V., SAAH, A. J., PHAIR, J. P., KINGSLEY, L. A., FAHEY, J. L., GINZBURG, H. M. and POLK, B. F. (1988). Predictors of decline in CD4 lymphocytes in a cohort of homosexual men infected with human immunodeficiency virus. *J. Acquired Immune Deficiency Syndromes* **1** 396–404.

Employs an autoregressive model relating CD4 (T-lymphocyte helper) cell counts to fixed and time varying predictor variables, while adjusting for previous CD4 counts, in order to identify predictors of CD4 cell decline. Analyzes longitudinal data from four semi-annual examinations of a cohort of homosexual seroconvertors in the MACS study. (IIIC)

MUÑOZ, A., WANG, M.-C., BASS, S., TAYLOR, J. M. G., KINGSLEY, L. A., CHMIEL, J. S., POLK, B. F., and THE MULTICENTER AIDS COHORT STUDY GROUP (1989). AIDS-free time after HIV-1 Seroconversion in homosexual men. *Am. J. Epidemiol.* To appear.

Estimates distribution of AIDS-free time following seroconversion by combining data from seroconverters and seroprevalent participants in the Multicenter AIDS Cohort Study (MACS), an investigation of nearly 5000 gay men without prior diagnosis of AIDS. Using data from seroconverters, models time since seroconversion by Weibull regression methods, conditional upon levels of time-varying hematological measures. Identifies these variables by graphical procedures based upon a locally weighted scatterplot smoother. Tests goodness of fit by examining whether the residuals follow a unit exponential distribution. Then, for the seroprevalent participants, imputes the unknown time from seroconversion to study entry using their baseline hematological measures, given the Weibull model developed from seroconverter data.

Uses a truncated product limit estimator to estimate the distribution of AIDS-free time following seroconversion for the seroprevalent cohort. This method extends the standard Kaplan–Meier estimator to account for truncation due to imputing time since seroconversion and for censoring resulting from loss to follow-up, other deaths and time of analysis. Assesses the precision of this estimate through bootstrap procedures. Compares to survival curve estimates produced by the Kaplan–Meier technique. Calculates the incidence rate from the estimated cumulative hazard rate and examines the median AIDS-free period. (IIIA)

PANJER, H. H. (1988). AIDS: Survival analysis of persons testing HIV+ (with discussion). *Transac. Soc. Actuaries.* **40** 517–542.

Describes a survival analysis of the passage of at risk and seropositive individuals through the Walter Reed Stages. Models this progression as a continuous time Markov chain and assumes constant hazard functions for passing from one stage to the next. Produces numerical maximum likelihood estimates of the hazard function and expected time to next stage, under both midpoint departure and random censoring assumptions. Applies these methods to life-table data from a German longitudinal study previously analyzed in Cowell, M. J. and Hoskins, W. H. (1988). AIDS, HIV mortality and life insurance. In *The Impact of AIDS on Life and Health Insurance Companies: A Guide for Practicing Actuaries, Report of the Society of Actuaries Task Force on AIDS.* Itasca, Ill. (Task Force report also appears in *Trans. Soc. Actuaries* **40**.) (IIIC)

PANJER, H. H. (1989). AIDS: Some aspects of modelling the insurance risk. *Trans. Soc. Actuaries* **41**.

Develops a Poisson process model for estimating the number of AIDS cases that have not yet emerged as death or health benefit claims, thus allowing actuaries to adjust insurance company reserves. Uses an exponential Poisson rate function to model new infectives in the population and a specified gamma or stage model (Panjer, 1988) for the distribution of incubation times. From the resulting Poisson distribution for new AIDS diagnoses, estimates parameters of the rate function using surveillance data from U.S., Canada and Australia. Compares estimates of HIV infecteds for four models of the incubation distribution. (II, IIIC)

PETERMAN, T. A., LUI, K.-J., LAWRENCE, D. N. and ALLEN, J. R. (1988). Estimating the risks of transfusion-associated acquired immune deficiency syndrome and human immunodeficiency virus infection. *Transfusion* **27** 371–374.

Calculates the maximum likelihood estimate of the total U.S. AIDS cases resulting from HIV-infected transfusions received during 1978–84, based on cases already reported and assuming a Weibull incubation density. The method is that of Lui et al. (1986). Estimates the overall number of transfusion infections based on these estimated cases and the prevalence of HIV in 1985 blood donors. (II)

PETO, J. (1986). AIDS and promiscuity. *Lancet* **2** 979.

Describes a series of speculative simulations assuming an arbitrary risk of infection per contact with an infected. Suggests that a promiscuous minority probably played a major role in the emergence of the AIDS epidemic in homosexuals, and that a similar epidemic could develop among single heterosexuals practicing serial monogamy. (II)

PICKERING, J., WILEY, J., PADIAN, N. S., LIEB, L. E., ECHENBERG, D. F. and WALKER, J. (1986). Modeling the incidence of acquired immunodeficiency syndrome (AIDS) in San Francisco, Los Angeles, and New York. *Math. Modelling* **7** 661–688.

Proposes a discrete, nonlinear model to predict new AIDS cases from previous diagnoses. Frames the model as a difference equation in discrete time units and iterates deterministically. As a result, the deviation between estimates under this model and actual incidence has no known probability structure. Model parameters reflect length of time from exposure to diagnosis (assumed constant), length of infectious period (assuming diagnosed cases are not infectious) and proportion developing immunity among the at-risk population (assumed of constant size). Employs incidence data for anal/rectal gonorrhea to estimate a time-varying sexual contact rate, via an auxiliary model. Considers estimates of the incubation period and length of infectious period derived from parameter values producing the best fit to incidence data from Los Angeles, New York and San Francisco. See Jager and Ruitenberg (1988), 38–51, where Pickering et al. update this model to reflect more recent AIDS incidence data, restrict the range of the saturation and incubation period parameters and conduct sensitivity analyses. (I, II)

REES, M. (1987). The somber view of AIDS. *Nature* **326** 343–345.

Fits a normal distribution to the incubation period for United States transfusion-associated AIDS cases. Predicts the number of HIV infections in the United States by the end of 1984 that will result in AIDS cases over the next 30 years.

A long series of commentaries and rejoinders ensued:

BARTON, D. E. (1987). Striking the balance on AIDS. *Nature* **326** 734.

REES, M. (1987). Describing the AIDS epidemic. *Lancet* **2** 98–99.

DAGPUNAR, J. S., LAWSON, A. and SENN, S. J. (1987). Sober view of AIDS. *Nature* **328** 10.

COSTAGLIOLA, D. and DOWNS, A. M. (1987). Incubation time for AIDS. *Nature* **328** 582.

BEALE, S. (1987). On the somber view of AIDS. *Nature* **328** 673.

LUI, K.-J., PETERMAN, T. A. and LAWRENCE, D. N. (1987). Comments on the somber view of AIDS. *Nature* **329** 207.

REES, M. (1987). Incubation period for AIDS. *Nature* **330** 427–428. (II, IIIA)

ROSEN, G. (1988). HIV spread in the San Francisco cohort: Scaling of the effective logistic rate for seropositivity. *Bull. Math. Biol.* **50** 345–349.

Modifies the May and Anderson (1987) homogeneous mixing model to account for changes in transmission efficiency and the mean number of sexual partners over time. Reports that for San Francisco City Clinic cohort seroprevalence data from 1978 through 1985, the effective logistic rate of increase in the proportion of seropositives manifests a simple scaling (semigroup) property that incorporates the impact of these changes. (I)

STRUTHERS, C. A. and FAREWELL, V. T. (1989). A mixture model for time to AIDS data with left truncation and uncertain origin. *Biometrika* **76**. To appear.

Formulates a mixture model for time to AIDS that employs a logistic model for the probability of developing AIDS following HIV infection and assumes a Weibull distribution for incubation time. Investigates the impact on maximum likelihood estimation of assuming that infection time is uniformly distributed over a known interval and that the data are left truncated and right censored. Indicates that, except for severe censoring, maximum likelihood estimates of the probability of developing AIDS and of the Weibull parameters are reasonably efficient compared to estimates based on observed infection times, for interval widths of up to half the mean of the Weibull distribution. Applies these methods to data from a cohort of seropositive homosexual/bisexual males known to have had sexual contact with diagnosed AIDS or ARC cases. (IIIA)

TAN, W. Y. and HSU, H. (1989). Some stochastic models of AIDS spread. *Statistics in Medicine* **8** 121–136.

Proposes a stochastic model for HIV transmission through a gay population. Postulates a latency period following infection during which an HIV recipient is not infectious. Thus, divides the population into suceptibles, latent carriers, infectives and AIDS cases. Assumes homogeneous mixing between susceptibles and infectives. Shows that the probability generating function for the numbers in each category satisfies the Kolmogorov forward equation. Derives equations for the expected value and the variance and covariance of these numbers. Indicates that for a large population of susceptibles, the deterministic approach is equivalent to working with the expected numbers in the stochastic model. Numerically explores the effects of initial infective population sizes, changes in contact rates and the AIDS mortality rate. (I)

TAYLOR, J. M. G. (1989). Models for the HIV infection and AIDS epidemic in the United States. *Statistics in Medicine* **8** 45–58.

Employs a variant of Brookmeyer and Gail's back calculation method to estimate the minimum size of the AIDS epidemic. Models the number of cases developing in different chronological periods as multinomial random variables whose cell probabilities are the convolution of the incidence distr. and the distribution of times from HIV infection to AIDS. Employs five different models for incidence of infection: double exponential, root exponential, logistic, logistic prevalence and quadratic. Examines 21 nonparametric distributions for the time from HIV infection to AIDS, based on data from cohorts of seroconvertors (compare Brookmeyer and Gail (1986) who assume a Weibull incubation period estimated from retrospective transfusion-related AIDS incidence data).

Fits these 105 models by maximum likelihood methods to CDC AIDS incidence data from 1978 to 1987 adjusted for reporting delays and examines the goodness of their fits via likelihood ratio chi-squared statistics. Finds that the double exponential incidence distribution generally provides the best fit, but observes no obvious pattern over the time to AIDS distributions.

Estimates seroprevalence, but points out that traditional methods for calculating the standard error of this estimate capture only sampling variability. Attempts to incorporate model selection uncertainty in the standard error estimate through a Bayesian approach, calculating the predictive standard error over all 105 models and assigning greater weights to those providing better fits. Acknowledges, however, that this method still does not account for uncertainty in AIDS surveillance data or reporting delay assumptions and is also dependent upon how one choses the prior set of distributions and infection models. (II)

TAYLOR, J. M. G., SCHWARTZ, K. and DETELS, R. (1986). The time from infection with human immunodeficiency virus (HIV) to the onset of AIDS. *J. Infect. Dis.* **154** 694–697.

Claims that examining the incubation time distribution is more appropriate than calculating the mean incubation period given the uncertainty about long-term AIDS incidence. Employs the method of multiple imputations to obtain Kaplan–Meier estimates of the incubation period distribution with standard errors, based on estimates of the earliest and latest possible dates of infection. Predicts number of AIDS cases resulting from pre-1985 infection using the Kaplan–Meier estimate for number of infections. Also uses multiple imputations to obtain Kaplan–Meier estimates of time to appreciable immune deficiency as evidenced by T-helper cell counts below 250/ml. (IIIA)

TENNISON, B. R. and HAGARD, S. (1988). AIDS: Predicting cases nationally and locally. *Brit. Med. J.* **297** 711–713.

Predicts short-term United Kingdom AIDS incidence through trend extrapolation. Employs a Box-Cox analysis to determine the appropriate power transformation for weighted logarithmic linear and quadratic regression models. Emphasizes recent data by using weights that exponentially decrease into the past with a common ratio. Produces local predictions by scaling down national estimates by the proportion of cases observed locally. (II)

THOMPSON, J. R. (1989). A simple model of AIDS. *Empirical Model Building*. Wiley, New York. (Chapter 2, Section 6, 79–91.)

Examines a simple deterministic model of HIV transmission through a gay population. Divides the population into two compartments with different contact rates and assumes population-wide random mixing. Examines the equilibrium values and finds that the presence of a small promiscuous subgroup substantially enhances the epidemic's sustainability. Contends that given the apparently low infectivity of HIV, the AIDS epidemic is rather fragile. Thus, suggests that the epidemic's current virulence may have resulted from the failure of public health authorities to close establishments that facilitated high anal sex contact rates. (I)

VAN DRUTEN, J. A. M., DE BOO, TH., JAGER, J. C., HEISTERKAMP, S. H., COUTINHO, R. A. and RUITENBERG, E. J. (1986). AIDS prediction and intervention. *Lancet* **1** 852–853.

Reports preliminary modeling results for seroprevalence and AIDS incidence from San Francisco CDC data, assuming one potentially effective sexual contact per person annually, an exponentially distributed infectious period and a 10% progression rate to AIDS after an average incubation period of the same length as the infectious period. (I)

WEYER, J., SCHMIDT, B. C. and KORNER, B. (1988). Ein Mehrgruppenmodell zur Simulation der epidemischen Dynamik von AIDS. (A multi-group model for the simulation of the epidemic dynamics of AIDS.) *AIDS-FORSCHUNG.* Part 1. March 154–156. (In German, with English summary.) Part 2. April 206–220. (In German.)

Simulates the dynamics of HIV infection in a city of one million inhabitants through a high-dimensional system of differential equations, accounting for demographic, sociologic and medical parameters. Starting from empirical values, the parameters are balanced to obtain an equilibrium between the supply and demand of sexual and drug contacts. Balanced parameters are further adjusted to reproduce data on the spread of AIDS. Computer simulations predict that AIDS incidence will cascade in gay men and IV drug users, ultimately spreading to the general population. Reports that the simulations reveal that changes in sexual behavior only benefit particular groups or temporarily delay the epidemic's spread. (I)

WILEY, J. A. and HERSCHKORN, S. J. (1988). The perils of promiscuity. *J. Infect. Dis.* **158** 500–501.

Examines the risks associated with having a fixed total number of sexual contacts in a period with various numbers of randomly selected partners. Assumes that all infectives are equally infectious, that the risk of encountering an infective is approximately constant through the period, and that each person's contacts are equally divided among his or her partners. Finds that when the probability of transmission by a single exposure is of the order of that probability estimated for anal intercourse between gay men, the risk increases significantly with multiple partners. Indicates that, under these assumptions, for the low levels of infectivity presently estimated for heterosexual intercourse, the number of partners has little effect on risk.

Points out, however, that the presence of lesions caused by venereal diseases, which are more readily transmitted through heterosexual intercourse than HIV, may increase the risk of HIV infection, thus placing promiscuous heterosexuals at greater risk. Also cautions that if infectiousness varies, promiscuous individuals bear an increased risk since their behavior presents greater opportunities for encountering the more infectious partners. (IIIB)

WILEY, J. A., HERSCHKORN, S. J. and PADIAN, N. S. (1989). Heterogeneity in the probability of HIV transmission per sexual contact: The case of male-to-female transmission in penile-vaginal intercourse. *Statistics in Medicine* **8** 93–102.

Represents heterogeneous infectivity among heterosexual couples by modeling infectivity as a random variable with either a beta or a two-point discrete distribution. Fits these models to two cohorts by constrained numerical maximum likelihood estimation methods: in one cohort, the index cases were transfusion-associated infecteds; in the other, they were largely bisexual males or hemophiliacs. Comparison with a homogeneous infectivity model suggests the existence of extremely heterogeneous infective rates and thus that the transmission risk may depend mainly on the number of sexual partnerships rather than on the contact frequency within each partnership. Cautions, however, that these results stem from studies of relatively modest sample sizes and that these models do not account for temporally heterogeneous infectivity. (IIIB)

WILKIE, A. D. (1988). An actuarial model for AIDS. *J. Roy. Statist. Soc. Ser. A* **151** 35–39.

Discusses the modeling requirements for actuarial applications. Describes a Markov stochastic process model, where transition intensities may vary by age, calendar year and duration in previous state. Points out that since intensity of infectivity varies with age, this model represents that level of sexual activity that is age-specific. Models incubation intensity by a Gompertz formula. Proposes numerical solution of the Kolmogorov differential equations. (I)

ZEGER, S. L., SEE, L-C. and DIGGLE, P. J. (1989). Statistical methods for monitoring the AIDS epidemic. *Statistics in Medicine* **8** 3–21.

Presents log-linear models for estimating the size and growth rate of the AIDS epidemic for different risk groups by region. Arranges the cases for a specific subgroup in a two-way table by date of diagnosis and length of reporting delay. Models the logarithm of the expected cell counts as the sum of a cubic spline function of diagnosis time and a nonparametric function of delay length. Employs Poisson regression to predict values for missing cells caused by reporting delays.

Develops an empirical Bayes procedure to improve the precision of trend estimates for subgroups with low incidence by borrowing strength across subgroups, relying most heavily on data for the same risk group in other regions and other risk groups in the same region. Thus, shrinks the log-linear model coefficients for the time margin from one subgroup towards the average values over all groups. Employs the empirical Bayes trend estimates to produce short term incidence projections for each subgroup. (II)