Defining Risky Sexual Behaviour in the UK: A Latent Class Approach

Beth Sonkin, Andrew Hinde

Abstract

This paper aims to define risky sexual behaviour in the UK with respect to the two most common bacterial sexually transmitted diseases: chlamydia and gonorrhoea. Using data from the National Survey of Sexual Attitudes and Lifestyles II, a nationally representative survey of sexual behaviour in Britain, this study aims to identify patterns of behaviours associated with increased disease risk by applying latent class techniques. A 3-class solution was obtained, splitting the sample into individuals with no sexual partners in the last year (8%), one sexual partner in the last year (71%) and the risky group, who had two or more sexual partners in the last year (21%). The paper then explores the prevalence of risky behaviour by ethnic group, age group and marital status.

Southampton Statistical Sciences Research Institute
Applications & Policy Working Paper A07/01

DEFINING RISKY SEXUAL BEHAVIOUR IN THE UK: A LATENT CLASS APPROACH

Beth Sonkin and Andrew Hinde

School of Social Sciences,

University of Southampton

Abstract

This paper aims to define risky sexual behaviour in the UK with respect to the two most common bacterial sexually transmitted diseases: chlamydia and gonorrhoea. Using data from the National Survey of Sexual Attitudes and Lifestyles II, a nationally representative survey of sexual behaviour in Britain, this study aims to identify patterns of behaviours associated with increased disease risk by applying latent class techniques. A 3-class solution was obtained, splitting the sample into individuals with no sexual partners in the last year (8%), one sexual partner in the last year (71%) and the risky group, who had two or more sexual partners in the last year (21%). The paper then explores the prevalence of risky behaviour by ethnic group, age group and marital status.

1 Introduction and Background

In the White Paper Choosing Health, published in November 2004 (Department of Health, 2004), the Government highlighted sexual health as one of its key target areas. In an accompanying statement, the then Health Secretary John Reid announced that £130 million would be spent to modernise Genitourinary Medicine clinics, £80 million to roll out a national chlamydia screening program, £50 million on a sexual health advertising campaign aimed at those aged under 25 years and £40 million to upgrade prevention services.

Prevention services and advertising will be aimed at the groups that the government has identified as a particularly “at-risk” due to high incidence of sexually transmitted diseases: young people aged under 25 years and black and ethnic minority populations (Health Protection Agency, 2005). But why are these groups particularly at risk? Is it because their behaviour differs in key ways from other individuals? And are there other groups that are also “at risk” that should be included in targeted campaigns to prevent sexually transmitted disease?

In order to answer these questions, we need to understand which behaviours are risky and how these are distributed in the population. By doing so, we will be able to design more effective public health campaigns. Observational studies can help us to determine which behaviours are associated with increased risk and in which population groups the odds of infection are highest. But it can still be difficult to determine what constitutes risky behaviour. For example, is someone with two partners who never uses condoms behaving in a risky way? What if those partners are not concurrent? Is someone in a new relationship who uses condoms safer? To truly understand and define risky sexual behaviour, we need to examine closely the interrelationships between different examples of sexual behaviours, and between these behaviours and outcomes.

Latent class analysis is a technique that can help to identify groups of individuals who share similar interests, values, characteristics or behaviours (Magidson and Vermunt, 2004a). This study will apply this technique to data from the National Survey of Sexual Attitudes and Lifestyles II (NATSAL II), with the aim of identifying sexual behaviour which puts an individual at risk of contracting a sexually transmitted disease (STD). This information will be used to develop a simple measure of risky sexual behaviour. It may also be used to inform policies aimed at reducing the incidence and prevalence of STDs in the general population.

It has been argued that current behaviour is more relevant to the study of incidence and prevalence rates of bacterial infections than viral infections. "Infections such as gonorrhoeal and chlamydial infection (short duration infections) are in general acquired as a result of recent sexual behaviours whereas infection with HIV and HSV-2 (long duration infections) may be acquired through behaviours that took place decades earlier" (Aral, 2004, p. 10). As NATSAL II is a cross-sectional study which asked individuals about their current behaviours, this paper will concentrate only on the two most common bacterial sexually transmitted diseases: chlamydia and gonorrhoea.

This study aims to define risky sexual behaviour with reference to chlamydia and gonorrhoea.

The study objectives are:

- to review the existing literature on behavioural risk factors associated with the two most commonly diagnosed bacterial STDs (*Chlamydia trachomatis* and *Neisseria gonorrhoeae*) to determine which are associated with increased disease risk in observational studies and which groups have the highest risk of STD infection;
- using latent class analysis, to analyse survey data on sexual behaviour drawn from the general population to determine whether there are clusters of individuals within the data with similar sexual behaviours;
- to use these findings to develop a simple variable to measure risky sexual behaviour;
- to determine the prevalence of risky sexual behaviour in key groups within the study population; and
- to explore any implications of these findings for policies targeted at reducing the incidence/prevalence of bacterial STDs in the UK.

2 Literature Review

2.1 Background

The variables included in a latent class analysis should be known risk factors for STDs. Otherwise, individuals may be allocated to classes for reasons other than whether their behaviour is risky. For example, two distinct groups may differ in their smoking habits. But if smoking is not a known risk factor for any bacterial STD, then the analysis may not be usefully identifying from the data groups engaging in risky sexual behaviour.

Epidemiological studies provide quantitative estimates of the levels of risk at which certain behaviours place individuals of contracting a bacterial STD. A review of the literature was undertaken in order to determine which sexual behaviours have been associated with increased risk of STD infection in previous studies.

2.2 Study selection

2.2.1 Study identification

The search was conducted by reviewing the online databases PubMed, Popline, and the Cochrane Collaboration's Controlled Trials Register. Online searches were also carried out using conventional search engines such as Google, Google Scholar, Yahoo!, etc. As relevant papers were identified, their reference lists were reviewed and followed up.

2.2.2 Eligibility criteria

- Papers must have been published in English. Unpublished studies were not included.
- Study participants must have been drawn from the general population (i.e. not from specific “at risk” groups such as sex workers, gay men, etc.).
- The study must have considered the odds of disease infection for at least one of the diseases of interest (i.e. *Chlamydia trachomatis* or *Neisseria. gonorrhoea*)
- The results must have been disease-specific and clearly identified. Different diseases may have different risk factors and the results of the review might be skewed by including results where the outcome measure was not clear.
- The study must have examined the odds of STD infection for one or more behavioural variables.
- Studies must have reported odds ratios and 95% confidence intervals for their estimates or have provided sufficient data to allow these measures to be calculated.

Systematic reviews were eligible for inclusion but only those studies in the reviews which met the above criteria were included.

2.3 Selected studies

The 24 studies which met the selection criteria are summarised in Tables 1 and 2 below. This included one systematic review which provided data from a further five studies.

Table 1. Epidemiological studies of behavioural risk factors for chlamydia infection

First author and year of publication	Type of study	Study size	Study population
Fenton et al. (2001a)	Cross-sectional	11,161	From NATSAL II
Gershman and Barrow (1996)	Cross-sectional	12,926	Females attending family planning clinics in Colorado
Hart (1992)	Cross-sectional	3,533	Females attending STD clinic in Adelaide, Australia
Hart (1993)	Cross-sectional	7,992	Men attending STD clinic in Adelaide, Australia
Hughes et al. (2000a)	Cross-sectional	18,238	STD clinic patients in London and Sheffield
Jonsson et al. (1995)	Cross-sectional	611	Sample of women living in Umea, Sweden
Latino et al. (2002)	Cross-sectional	3,314	Women in Turin, Italy
Niccolai et al. (2005)	Retrospective	1,455	Medical records from an STD clinic in Connecticut, USA
Radcliffe et al. (2001)	Case-control	1,351	Patients attending STD clinic in Birmingham, UK
Ramstedt et al. (1992)	Cross-sectional	5,274	Women seeking contraceptive advice in Gothenburg, Sweden
Vuylsteke et al. (1999)	Cross-sectional	2,784	Sample of women living in Antwerp, Belgium
Weinstock et al. (1991)	Cross-sectional	1,348	Women seeking contraceptive advice in San Francisco, California
Zenilman et al. (1994)	Cross-sectional	1,155	STD clinic attendees in Baltimore, USA

Table 2. Epidemiological studies of behavioural risk factors for gonorrhoea infection

First author and year of publication	Type of study	Study size	Study population
Austin et al. (1984)	Case-control	Not available	STD clinic, USA
Barlow (1977)	Cross-sectional	Not available	STD clinic, UK
Bjekic et al. (1997)	Case-control	800	Hospital patients in Belgrade, Yugoslavia
D'Oro et al. (1994)*	Systematic review	N/A	N/A
Hart (1992)	Cross-sectional	3,533	Females attending STD clinic in Adelaide, Australia
Hart (1993)	Cross-sectional	7,992	Males attending STD clinic in Adelaide, Australia
Hughes et al. (2000)	Cross-sectional	18,238	STD clinic patients in London and Sheffield
Mertz et al. (2000)	Case-control	307	Male STD clinic patients in Newark, USA
Pemberton et al. (1972)	Cross-sectional	Not available	STD clinic Ireland
Rosenberg et al. (1992)	Retrospective	Not available	STD clinic USA
Upchurch et al. (1990)	Cross-sectional	607	STD clinic patients in Baltimore, Maryland

*Provided data from the following studies: Austin, Barlow, Pemberton, Rosenberg.

2.4 Results

Where studies provided results for both males and females, these have been presented separately. This was to explore whether there were important differences between the sexes with respect to risk factors. It was not considered appropriate to combine the study results and present a meta-analysis as the risk factors measured were not consistently defined across studies (Egger et al., 1997). The definitions used in each study are presented in Appendix I. The results presented in Figures 1 and 2 are those following multivariate analysis. This aimed to control for the possible confounding effects of other variables as well as demographic and socioeconomic factors such as age, sex and socioeconomic status. Not all studies included the same variables in the multivariate analysis.

The review found that having multiple partners, not using a condom with all partners and having had a short-term relationship were all associated with increased risk of chlamydia or gonorrhoea infection. The odds of chlamydia infection were also increased in girls who had their first sexual experience before age 16 years. These were the only statistically significant variables found in the majority of studies.

2.4.1 Previous STD infection and alcohol consumption

Unlike the other risk factors, studies were not found which presented odds ratios and confidence intervals for the risk of subsequent chlamydia or gonorrhoea infection if an individual had been previously diagnosed with an STD. However, a number of studies have found a high prevalence of reinfection with either chlamydia or gonorrhoea (Mehta et al., 2003; Whittington et al., 2001; Rietmeijer et al., 2002; Burstein et al., 2001). Moreover, individuals who have had a previous STD are more likely to become infected with chlamydia or gonorrhoea (Fortenberry et al., 1999; Gunn et al., 2000; Hughes et al., 2000b).

Several studies did consider whether individuals who drank alcohol were more at risk than those who were non-drinkers. Although odds ratios and confidence intervals were not presented, these studies did not find any significant difference in the odds of infection with either chlamydia or gonorrhoea (Radcliffe et al., 2001; Vuylsteke et al., 1999; Bjekic et al., 1997; Zenilman et al., 1994).

Figure 1. Odds of chlamydia infection

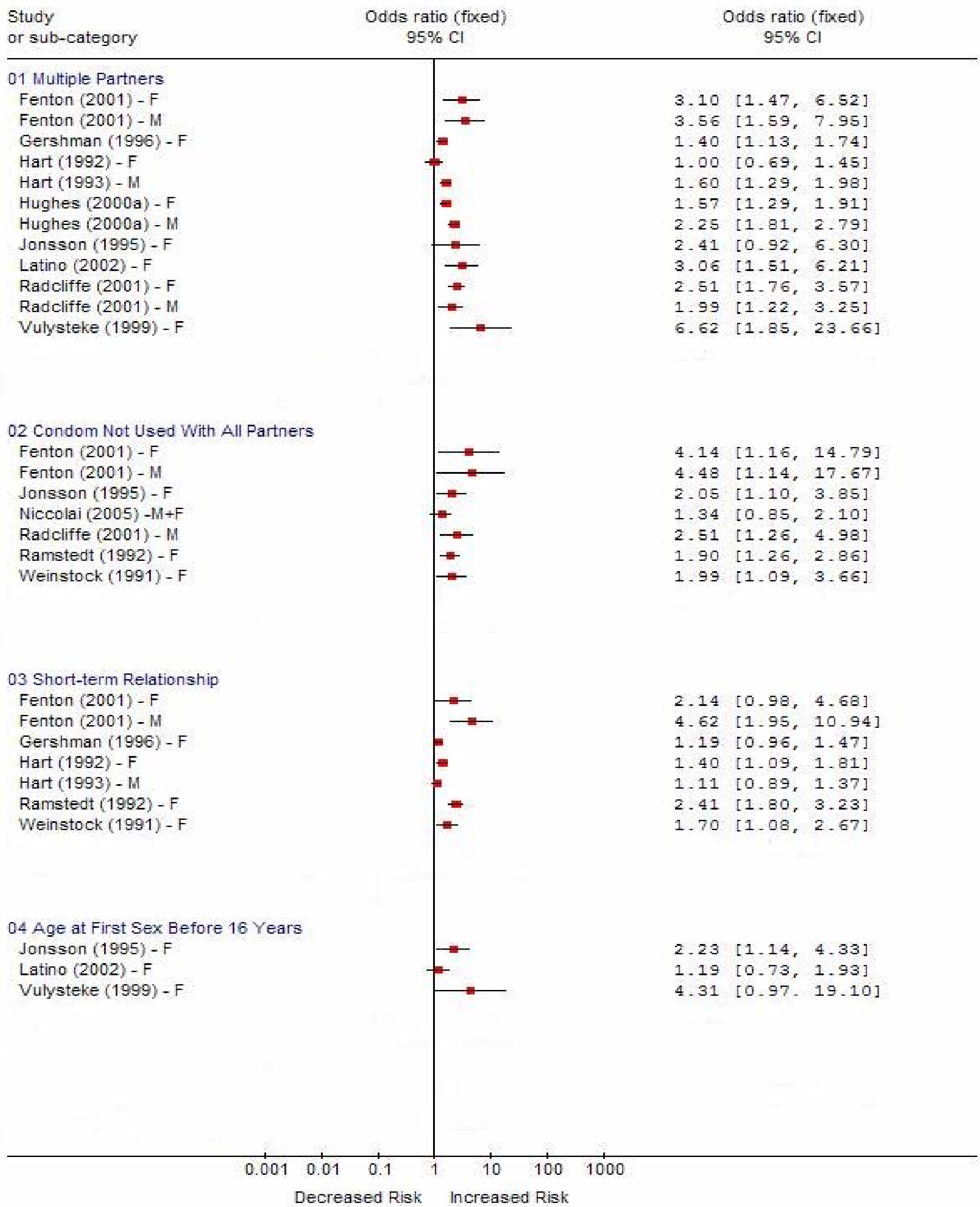
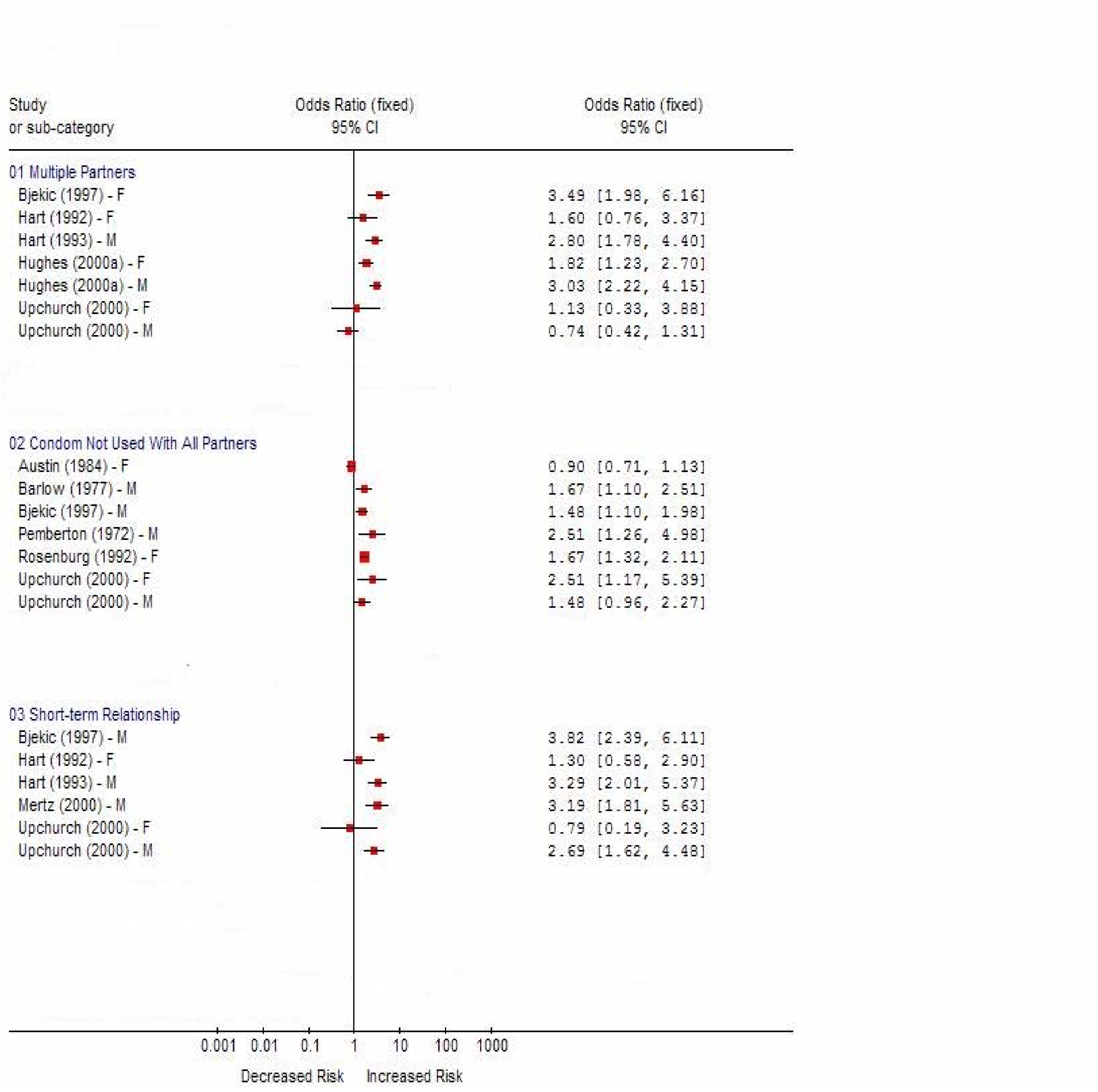


Figure 2. Odds of gonorrhoea infection



2.5 Conclusions

The literature review found the following behavioural risk factors associated with chlamydia and gonorrhoea infection:

- multiple partners,
- short term partnerships,
- non-use of condoms,
- age at first sex before 16 years old, and
- previous STD diagnosis.

These variables were taken forward and considered for inclusion in the latent class model.

3 Data and Methodology

3.1 Data source

The data used in this study were drawn from the National Survey of Sexual Attitudes and Lifestyles II (NATSAL II). NATSAL II is a nationally representative survey of sexual behaviour in Britain. Modelled on the first NATSAL survey conducted in 1990-1991, NATSAL II aims to provide a detailed understanding of the sexual behaviour patterns.

Using a combination of computer assisted personal interview (CAPI) and computer assisted self-interview (CASI), NATSAL II gathered data on sexual attitudes and behaviours from 12,110 individuals aged 16-44 years (11,161 from the general population and 949 from an ethnic minority boost sample) (Erens et al., 2001). Interviews began in May 1999 and were fully completed in February 2001. The general population sample was drawn using a multi-stage stratified probability sampling method. However, it was necessary to oversample in inner and outer London to compensate for predicted lower response rates and because NATSAL I showed a higher prevalence of HIV risk behaviours in London than elsewhere in Britain. It was thought that oversampling these areas would increase the precision of HIV prevalence estimates (Erens et al., 2001).

A sub-sample of individuals was asked to provide a urine sample to test for *Chlamydia trachomitis*. Half of the addresses at all sample points were selected for participation. Only those aged 18-44 years were eligible to participate. Approximately 70% did so, providing a sample of 3,608 individuals (Erens et al., 2001).

The ethnic minority boost sample was also selected using a multi-stage process. To ensure adequate numbers for analysis, selection was based on a combination of full screening and focused enumeration in areas identified in the 1991 census where at least 6% of the population were ethnic minorities (Erens et al., 2001).

Further details of the NATSAL II sampling methodology can be found in the survey's technical report (Erens et al., 2001). A response rate of 63.9% was achieved overall in the general population sample and 59% in the ethnic minority boost sample. This was slightly below the 64.7% response rate for NATSAL I.

The NATSAL II sample was compared with the mid-1999 population estimates on age, sex and government office region. In spite of oversampling in London, London residents were still underrepresented, as were men aged 25-29 years. It was determined that additional weightings were required as these differences might have been due to differential non-response. Following the application of all relevant weightings, the characteristics of the NATSAL II sample closely reflected those of the general population (Erens et al., 2001).

3.2 Latent class analysis

Sometimes we cannot directly observe the construct in which we are interested. Just imagine the responses you would get to the question “Do you engage in risky sexual behaviour?”. However, we can measure variables which we believe are characteristic of risky sexual behaviour. For example, we might expect people with risky sexual behaviour to have more partners, not to use condoms, to have previously had an STD, and so on. Since these observable, or “manifest”, variables are caused by the underlying, or “latent” variable, we expect a high degree of covariation among them (McCutcheon, 1987).

Latent class analysis studies the interrelationships between these manifest variables to help us to understand the latent variable. It can help us to identify classes of people who share similar interests, values, characteristics or behaviours (Magidson and Vermunt, 2003). It can also help us to highlight which behaviours differ between groupings and hence which are key to understanding risky sexual behaviour.

3.2.1 Model formulation

The calculations that underlie latent class analysis are based the principle of conditional independence, i.e. in a correctly specified latent class model, all the covariation between the observed variables will be explained by the latent variable. Within each latent class that is identified, the manifest variables are all assumed to be statistically independent of one another (Uebersax, 2001).

The latent class model is a simple parametric one. It uses the observed data to estimate two sets of parameters: the conditional response probabilities and the latent class prevalences.

The conditional response probabilities give the probability that in a particular latent class, for a given manifest variable, a randomly selected member of that class will give a particular response, for example, the probability that an individual in latent class 1 would have more than one partner (Uebersax, 2001). Comparing the response probabilities allows the examination of how latent classes differ from one another. If, for example, there is no difference between the probabilities of condom use between those in latent class 1 and those in latent classes 2 or 3, then condom use is probably not a key differentiating feature between people who engage in risky behaviour and those who do not.

The other parameters, the latent class prevalences, tell us the proportion of the population which falls into each latent class. They tell us how common certain groupings are in the study population.

Using these two sets of parameters, the probability of obtaining a specific response pattern can be expressed as the product of the conditional probabilities and the latent class prevalence. For example, if we have three manifest variables (or items) A, B and C, then the probability that a person who gave response i to item A, response j to item B and response k to item C will be in latent class t is $\Pi_{ijkt}^{ABCX} = \Pi_{it}^{A|X} . \Pi_{jt}^{B|X} . \Pi_{kt}^{C|X} . \Pi_{t}^{X}$, where X is the latent variable, t indexes the classes of the latent variable X, Π_{t}^{X} is the probability of a randomly selected case being at level t of the latent variable X and $\Pi_{it}^{A|X}$, $\Pi_{it}^{B|X}$ and $\Pi_{it}^{C|X}$ are the conditional probabilities of obtaining the ith, jth and kth responses to items A, B and C respectively from members of class t (Magidson and Vermunt, 2004b).

3.2.2 Parameter estimation

The parameters in the latent class model are estimated by the maximum likelihood (ML) criterion. The ML estimates are the ones most likely to have occurred given the observed data. Estimation requires iterative computation, and is usually undertaken using a computer program.

Several methods are available for calculating the ML estimator. The Expectation-Maximization (EM) Algorithm was derived by Goodman (1974). It considerably simplified the process which had previously been achieved through matrix manipulation and the calculation of solutions to simultaneous linear equations (Uebersax, 2001; McCutcheon, 1987). Although it can be slower than some of the more recently developed methods, the EM method is very stable and works well with sparse or incomplete data (Vermunt, 1997). As such, this is the method employed by most available latent class analysis programs including LEM, the program used in this analysis (Vermunt, 1997).

If the likelihood does not have a single local maximum, the results may depend upon the starting value selected. Magidson and Vermunt argue that the best way to proceed in this case is to estimate the model with different sets of random starting values. “Typically, several sets converge to the same highest log-likelihood value, which can then be assumed to be the ML solution” (Magidson and Vermunt, 2004a, p. 5).

4 Latent Class Analysis

4.1 Selecting manifest variables

Chlamydia and gonorrhoea are both treatable infections of short duration. As such, recent behaviours are likely to be more relevant to disease risk than those that may have taken place years before. It was decided to exclude behaviours that may have taken place many years earlier (such as age at first sexual experience) and to concentrate on those that have occurred in the last year. The exception to this was “previous STD diagnosis”, as a previous diagnosis may still be affecting an individual’s behaviour, perhaps making him/her more cautious either to avoid another infection or to avoid infecting a partner.

Based on the results of the literature review, five variables were selected from NATSAL II as possible manifest variables for the latent variable “risky sexual behaviour”. These variables were checked for association with self-reported incidence of chlamydia and gonorrhoea in the last year in NATSAL II. Because only nine individuals reported a gonorrhoea diagnosis in the last year before the survey, we also considered a diagnosis in the last five years. The p-values for the univariate associations are reported in Table 3 below, showing that at the 5% level, all the variables identified by the literature review were associated with both chlamydia and gonorrhoea diagnosis. In addition, concurrent relationships in the last year seemed to be a possible risk factor for chlamydia diagnosis and so this variable was included.

4.2 Selecting covariates

Some groups within the UK population have a higher observed risk of chlamydia or gonorrhoea infection than others. In 2005, the Health Protection Agency identified higher incidence of both chlamydia and gonorrhoea in black ethnic minority groups and people under 25 (Health Protection Agency, 2005). Previous studies have also found that Black Africans and Black Caribbeans have higher odds of infection when compared to Whites and Asian groups. Married people have been observed to be less at risk than their single counterparts and younger people have much higher odds of disease than older age groups (Winter et al., 2000; Fenton et al., 2005; Lacey et al., 1997; Low et al., 2001; Fenton et al., 2001a; Radcliffe et al., 2001).

Table 3. Univariate association between five possible risk factors and self-reported chlamydia and gonorrhoea diagnosis in last one and five years

Variable	p-value for chlamydia last year	p-value for chlamydia last five years	p-value for gonorrhoea last year	p-value for gonorrhoea last five years
Number of sexual partners in the last year	<0.0001	<0.0001	<0.0001	<0.0001
Ever diagnosed with an STI	<0.0001	<0.0001	<0.0001	<0.0001
Concurrent relationships	0.0722	<0.0001	0.2943	0.5217
New partner	0.0003	0.0001	0.0004	0.0066
Sex without a condom	0.0006	<0.0001	0.0051	0.0367

These variables are therefore included in the latent class analysis as covariates. By analysing the data for the population stratified by these variables, the latent class analysis can help us to identify any differences in the prevalence of risky behaviour.

Table 4 summarises the breakdown of the sample population by age group and marital status. About half of the single people were in the youngest age group and only 16% were in the oldest group. Marriage, and widowhood, separation and divorce (respondents having experienced the last three and not having remarried being combined into a “previously married” group for convenience) are more common in the older age groups. About half of all people who were cohabiting were in the middle age group, 25-34 years.

Table 4. Age composition of different marital statuses

Marital status	Age group		
	16-24 years	25-34 years	35-44 years
Married	3.44%	40.54%	56.03%
Cohabiting	19.95%	50.49%	29.55%
Single	50.57%	33.65%	15.79%
Previously Married	2.02%	34.89%	63.09%

NATSALII asked respondents to identify their ethnic group. The variable derived from this information identified the following groups: Black, White, Indian, Pakistani, Bangladeshi, Chinese, Other Asian and Other. The Bangladeshi, Chinese and Other Asian groups were too small to be used in further analysis. The Other group was also disregarded as it was unclear what the ethnic origin was of individuals who had been allocated to this group, except that it was not one of the ones listed. We therefore included four ethnic groups in the analysis: Black, White, Indian and Pakistani.

The age distributions were fairly similar across all four ethnic groups. The Pakistani group was slightly younger than the others with 25% in the 16-24 year age group, compared with 17-18% of the Indian and Black group and 21% of the White group. The largest age group among Blacks was 35-44 years (44% of Blacks were in this age group); in the other ethnic groups where the largest age group was 25-34 years (Table 5).

Table 5. Distribution of ethnic group by age group

Ethnic group	Age group		
	16-24 years	25-34 years	35-44 years
White	20.82%	40.10%	39.09%
Black	17.74%	38.33%	43.93%
Indian	16.94%	45.18%	37.87%
Pakistani	25.31%	50.61%	24.08%

Unlike the age distribution, the marital status distribution differed substantially between ethnic groups (Table 6). The Black group had the highest percentage single (49%) whilst the Pakistani group had the lowest (18%). On the other hand, 61% of Indians and 66% of Pakistanis were married, which was higher than in the other groups, with Blacks having the lowest proportion married at only 28%. Cohabitation was most prevalent amongst the white group (17%) and rare amongst Indians and Pakistanis.

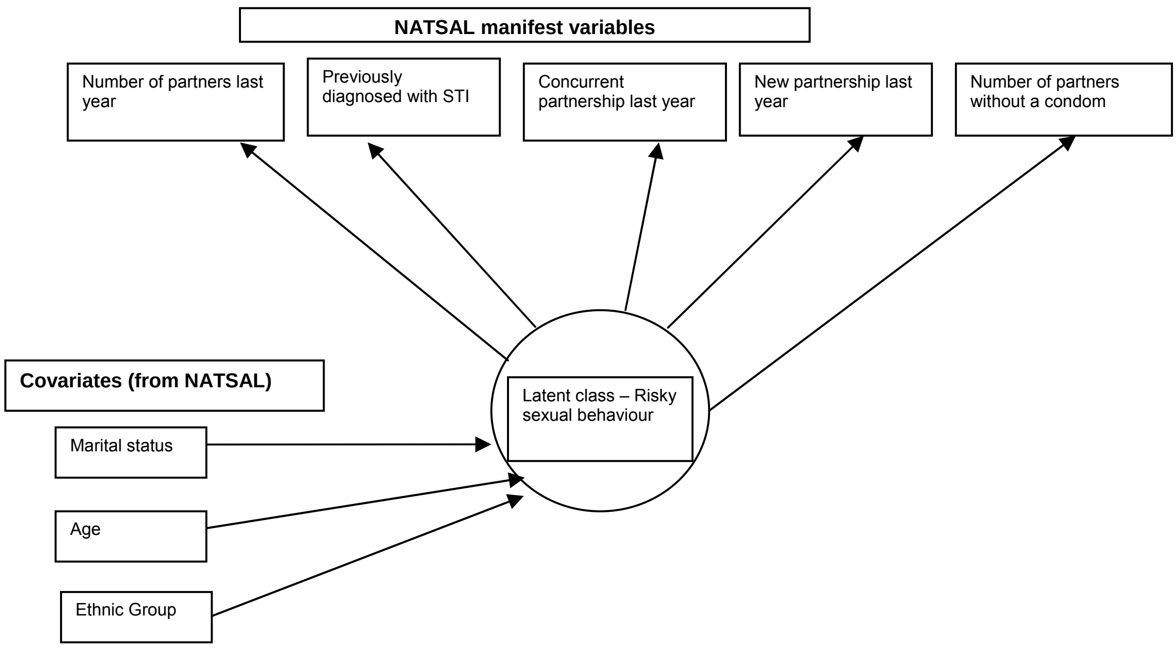
Table 6. Distribution of marital status by ethnic group

Ethnic group	Marital Status			
	Married	Cohabiting	Single	Previously married
White	38.57%	16.56%	35.18%	9.69%
Black	28.48%	10.18%	48.61%	12.73%
Indian	61.46%	2.66%	28.90%	6.98%
Pakistani	65.98%	2.46%	18.44%	13.11%

4.3 The model

The model proposed for latent class analysis is illustrated in Figure 3.

Figure 3. Latent Class Analysis Model



We started by fitting a 1-class model and continued adding one extra class at a time, considering all elements of model fit until a suitable model was found. We decided not to fit more than five classes. One of the aims of this study was to develop a simple measure which would aid in the understanding and analysis of risky sexual behaviour. If we needed more latent classes than we had manifest variables in order to explain risky sexual behaviour, then it was deemed that the latent class analysis was not helpful and another technique should be considered.

The data were cleaned to eliminate 172 individuals who had not provided any responses to any of the five manifest variables under consideration. Any individuals who had not had a sexual experience at the time of the survey were excluded as they would not have been exposed to the risk of contracting a sexually transmitted disease. This removed a further 706 individuals. The final sample size was 11,232. A further 236 individuals were identified as having given inconsistent answers (or example, they claimed only one partner during the last year but indicated two or more partners without a condom during the same period). The latent class analysis can deal with these inconsistencies and allocates these individuals to the latent class in which they have the highest posterior membership probability (Vermunt, 1997).

Missing data are assumed to be missing at random and the class allocation is made by calculating the posterior membership probability using the data which is available (Vermunt, 1997).

The program used for this analysis was LEM, developed by JK Vermunt specifically for the analysis of categorical data. The maximum likelihood estimates are computed using the Expectation Maximisation algorithm (Vermunt, 1997).

4.4 Results

4.4.1 Number of latent classes

There is no single statistical test to determine the number of latent classes a model should have. Selecting the "best" model requires the consideration of statistical measures of model fit and the substantive interpretation of model usefulness. For example, statistical model fit is often improved by adding an additional latent class; but the additional class may not improve our understanding of the characteristics of the underlying variable and may make comparing the conditional response probabilities more difficult (Storr et al., 2004)

The most common methods of selecting a model are as follows:

- comparing the model fit to the observed data using a chi-squared test,
- finding the simplest model using parsimony indices,
- comparing to a baseline model, and
- considering the level of classification error.

4.4.1.1 Chi-squared test statistic

Probably the most common and most familiar method of assessing model fit is the likelihood ratio chi-squared test statistic. This compares the observed data to the frequencies expected by the model. The test statistic is taken from the chi-squared distribution with a number of degrees of freedom equal to the number of different response patterns minus the number of estimated parameters. A significant result on the chi-squared test indicates that the model fits the data well (Uebersax, 2001). However, in latent class models with sparse data, the likelihood ratio does not always conform to the chi-squared distribution and the resulting test statistic becomes a less reliable measure (Storr et al., 2004; Magidson and Vermunt, 2004a). As a result, the chi-squared test statistic alone is often not enough.

The p-value for a 1-class model, as calculated by LEM, was not significant. However for the 2-, 3-, 4- and 5-class solutions, the chi-squared test statistic had a p-value of $p<0.0001$. This means that potentially any of these solutions provide a good fit to the observed data. However, with five manifest variables and several categories of response to each, the data may well have been sparse in some response cells. Therefore, this measure was not considered to be reliable

4.4.1.2 Parsimony indices

Instead of looking at the way that the model fits the observed data, we might consider which model (2-class, 3-class, etc) can most simply model the data – a sort of mathematical approach to Occam's razor. The Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) are indices which measure the number of estimated parameters required to fit the model. Models with lower AIC and BIC values are preferred.

Table 7 shows the BIC and AIC values for the models of risky sexual behaviour. The BIC and AIC both fall as additional latent classes are added until we reach four latent classes. As we increase from four to five latent classes, the BIC rises again, though the AIC continues to fall slightly. The parsimony indices suggest that the 4-class solution is the simplest and hence most acceptable.

However, the change from a 3-class model to a 4-class model is less than 1%, as it is from a 4-class model to a 5-class model. Since the differences are so small, it is worthwhile considering other measures of model fit before selecting a model.

Table 7. Information Criteria values for the models

Number of latent classes	BIC	AIC
2	71,209	70,791
3	63,868	63,348
4	63,720	63,098
5	63,810	63,084

4.4.1.3 Comparing to a baseline model

Adding latent classes complicates the model and its interpretation. It is worthwhile only if it adds to our total understanding of the latent variable and helps to explain the total association between the latent and manifest variables. Comparing to a baseline model gives an indication of how much of the total association is explained by adding another latent class. “In covariance structure modelling, a common choice of baseline model is a model imposing independence among the response variables” (Skrondal and Rabe-Hesketh, 2004, p. 270). Since a 1-class solution means that all the manifest variables are independent of one another, this is usually chosen as the baseline (Magidson and Vermunt, 2004a).

As shown in Table 8 below, moving from two to three latent classes explains an extra 25% of the association. But the addition of a fourth and a fifth latent class adds less than 1% each time.

Table 8. Proportion of total association accounted for by the model

Number of latent classes	Percentage of association explained
2	53.9%
3	81.4%
4	82.4%
5	82.6%

4.4.1.4 Classification error

When classes are well-differentiated, it is not difficult to determine in which latent class an individual belongs. However, when two or more latent classes have similar response probabilities, it can be difficult to determine where to allocate an individual (Nyland, 2005). For analytical purposes, it is useful to have a model with clearly defined classes and hence a low level of possible misclassification. For a full discussion of how the level of misclassification is determined, see Skrondal and Rabe-Hesketh (2004).

In the 2- and 3-class models the classification accuracy was very high and thus the classes were well-differentiated. This deteriorated with the addition of further latent classes. Under the 4- and 5-class models, approximately 15% of people were subject to potential misclassification.

Table 9. Classification error

Number of latent classes	Percentage of sample correctly classified
2	99,94%
3	99.97%
4	85.48%
5	83.42%

4.4.1.5 Conclusion

The parsimony indices seemed to indicate that the 4-class model might be the best solution as it is the simplest. However, taking all the measures into account, it was determined that, on balance, a 3-class model was preferable. It offered intuitive clarity, allowing us to classify people as “risky” or “faithful” or “alone” (see below). Although it had a slightly higher AIC and BIC than the 4-class model, the difference was negligible (about 1%). It also explained approximately the same amount of the total association and had a lower level of classification error. Furthermore, a 4-class model did not offer any additional insight into the “risky” group. Rather it further subdivided the “faithful” group based on whether they used condoms with their partner. Whilst this is an interesting insight, it was not deemed to be helpful in furthering our understanding of risky behaviour. Therefore a 3-class model was selected.

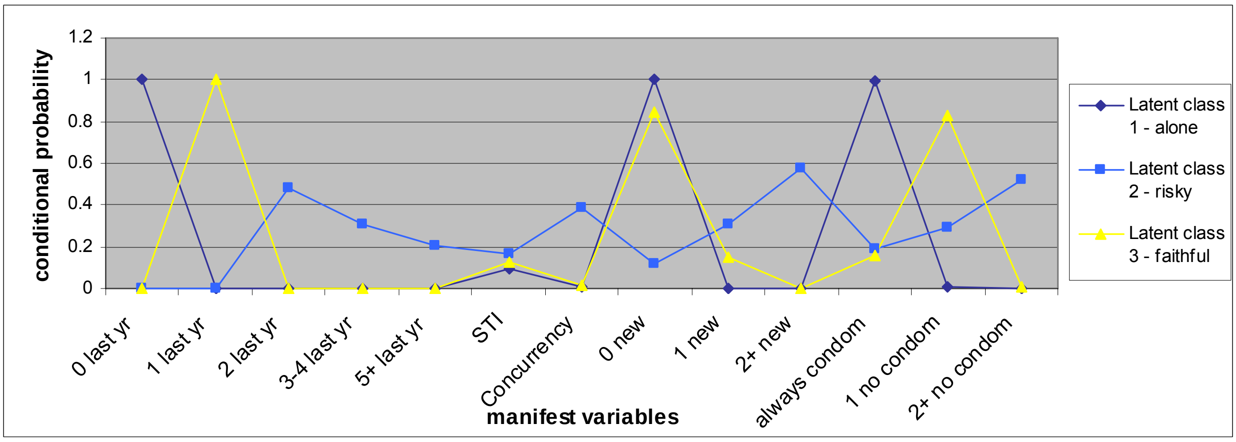
4.4.2 Class description

4.2.1.1 Three-class model – total population

In the 3-class model, 8% of the study population were allocated to latent class 1, 21% to latent class 2 and 71% to latent class 3.

Figure 4 shows the conditional probabilities based on a positive response to one of the key variables. The full list of conditional probabilities is shown in Table 10. By comparing the differences between the conditional probabilities in the three latent classes, we can explore the features of each latent class and how their behaviours differ.

Figure 4. Comparison of Latent Class 1, Latent Class 2 and Latent Class 3 on responses to key manifest variables



Number of partners in the last year seems to be the key differentiating feature between the classes. Individuals in latent class 1 universally had no sexual partners in the last year, although they had been sexually active previously and individuals in latent class 3 all had one sexual partner in the last year.

Individuals in latent class 2 had at least two sexual partners in this period, with 20.7% claiming five or more partners in the last year.

Since latent class allocation is based on the number of partners in the last year, the conditional probabilities on the other variables follow from this result. It is only those individuals in latent class 2 who have had concurrent partnerships, multiple partners, multiple new partners and multiple partners without a

Table 10. Comparison of Latent Class 1, Latent Class 2 and Latent Class 3 on responses to key manifest variables

Variable	Latent Class 1 (alone)	Latent Class 2 (risky)	Latent Class 3 (faithful)
Total number of sexual partners last year			
• 0	1.000	0.000	0.000
• 1	0.000	0.001	1.000
• 2	0.000	0.485	0.000
• 3-4	0.000	0.307	0.000
• 5+	0.000	0.207	0.000
Ever diagnosed with an STD (excluding thrush)			
• No	0.906	0.833	0.873
• Yes	0.094	0.168	0.127
Concurrent partnership in last year			
• No	0.002	0.384	0.956
• Yes	0.009	0.388	0.020
• 2+ partners but unknown concurrency	0.000	0.207	0.000
• Not applicable	0.879	0.000	0.000
• Not answered	0.110	0.022	0.025
Number of new partners last year			
• 0	1.000	0.120	0.847
• 1	0.000	0.307	0.153
• 2+	0.000	0.573	0.000
Number of partners without a condom			
• 0	0.994	0.188	0.157
• 1	0.006	0.291	0.832
• 2+	0.000	0.522	0.010

condom. They also had the highest rates of previous STD diagnoses, 17%, compared to 13% in latent class 3 and 9% in latent class 1. As a result, latent class 2 has been named the “risky” class. Since latent class 1 exclusively includes those with no partners over the period, we have named them the “alone” group. Following a similar approach, latent class 3 has been named the “faithful” group. These names are used in the rest of the paper for ease of reference.

4.2.1.2 Three Class model – stratified by age, marital status and ethnic group

As Table 11 shows, individuals aged under 25 years were more than twice as likely to be allocated to the “risky” group than those in the older age groups. The prevalence of risky behaviour falls as age rises. It is unclear whether this is an age effect (are younger people always more risk-seeking than older people?) or a cohort effect (are younger people now more risk-seeking than young people used to be?). The probability of being allocated to the “faithful” group rises as age increases, as does allocation to the “alone” group, perhaps indicating the rise in divorce and widowhood with age.

The latent class prevalences by marital status are shown in Table 12. Single people were most likely to be allocated to the “risky” group with almost 40% in this class. The prevalence of risky behaviour was much lower amongst married and cohabiting individuals (5.7% and 13.1% respectively) perhaps reflecting their more stable partnerships.

Table 11. Latent class probability by age

Age group	Probability of being in “alone” class	Probability of being in “risky” class	Probability of being in “faithful” class
16-25 (N=2331)	5.71%	38.35%	55.93%
25-34 (N=4543)	6.44%	20.45%	73.11%
35-44 (N=4358)	9.80%	12.78%	77.41%

The previously married group resembles the single group more than the married or cohabiting groups; 31% of them fell into the “risky” category and previously married individuals who were not allocated to the “risky” group were much more likely than any other group to be “alone”.

Table 12. Latent class probability by marital status

Martial status*	Probability of being in “alone” class	Probability of being in “risky” class	Probability of being in “faithful” class
Married (N=4,366)	1.12%	5.69%	93.19%
Cohabiting (N=1,703)	0.45%	13.12%	86.43%
Single (N=4,027)	13.99%	38.70%	47.31%
Previously Married (N=1,115)	20.30%	30.90%	48.81%

*The sum of the Ns does not equal 11,232 as 21 individuals did not provide details of their marital status

The latent class prevalences by ethnic group are shown in Table 13 below. The highest probability of being in the “risky” class is among the Black ethnic group at 25%, followed by the White ethnic group at 21%. The corresponding probability in the Indian and Pakistani groups is much lower with 14% and 13%

respectively. The White group had the lowest probability of being in the “alone” class whilst the Black ethnic group were the least likely to be in the “faithful” class.

Table 13. Latent class probability by ethnic group

Ethnic group	Probability of being in “alone” class	Probability of being in “risky” class	Probability of being in “faithful” class
White (N=9,301)	7.03%	21.01%	71.96%
Black (N=826)	11.87%	25.49%	62.63%
Indian (N=301)	10.48%	13.76%	75.76%
Pakistani (N=245)	11.35%	12.64%	76.01%

4.2.1.3 Standardisation

The results above tell us that the highest prevalence of risky behaviour is found amongst the Black ethnic group, individuals aged 16-24 years and single people. However, to isolate the independent effect of age, ethnic group and marital status, we need to control for the possible confounding effects of the other covariates. For example, most individuals aged 16-24 years are single so is the high prevalence of risky behaviour in this age group in part explained by their single status?

Direct standardisation allows us to control for possible confounding effects by comparing the observed prevalence of risky behaviour for a given covariate with the results we would expect if the prevalence were determined purely by the potentially confounding covariates. Using the simple example above, direct standardisation would compare the observed prevalence of risky behaviour in the 16-24 year age group with the prevalence we should expect if risky behaviour in this age group were determined not by age but by marital status only. If the observed value is very close to the expected value, then the prevalence of risky behaviour is largely dependent on marital status, not age.

The standardised results are presented in Table 14 below. Whilst married and cohabiting people seem to behave in a way that is less risky than predicted by their age and ethnic group profiles, single and previously married people behave in a way that is more risky. Married and previously married people have a very similar age and ethnic group distribution so their expected prevalence of risky behaviour is also similar. However, the actual prevalence shows a large gap, indicating that not being married any more has a very large effect on risky behaviour, independent of age and ethnic group effects.

Young people are slightly riskier in their behaviour than we would predict from their marital status and ethnic group profiles, whilst those aged 35-44 years are slightly less risky. Risky behaviour decreases with age even after controlling for marital status and ethnic group. This implies that the prevalence of risky behaviour is not just decreasing, for example, because as people get older they are more likely to settle down into stable partnerships. There is a further effect that is related to age, though it is still not clear whether this is a cohort effect or an age effect.

For the Black and White ethnic groups, the prevalence of risky behaviour is almost exactly as we would predict given their age and marital status profiles.

This means that the higher prevalence of risky behaviour amongst Blacks and Whites can be explained by their marital status and age distributions. The Indian and Pakistani groups, however, do show an effect of ethnic group with the actual prevalence of risky behaviour about 5% lower than the prevalence predicted by the age and marital status profiles.

Table 14. Standardised and observed percentages in "risky" class

	Percentages expected	Percentages observed
Marital status		
Married	16%	6%
Cohabiting	26%	13%
Single	27%	39%
Previously married	16%	31%
Age group		
16 – 24 years	33%	38%
25 – 34 years	20%	20%
35 – 44 years	17%	13%
Ethnic group		
Black	24%	25%
White	21%	21%
Indian	17%	14%
Pakistani	17%	13%

4.3 Testing the results – logistic regression

The analysis above suggests that the key factor differentiating those at risk and those not at risk of chlamydia and gonorrhoea is likely to be the number of partners in the last year. NATSAL II included a urine sample to test for chlamydia, which gives us an independent outcome measure. If number of partners is the key determinant of risky behaviour then a logistic regression model using this as the only explanatory variable should be as good, or nearly as good, at predicting the outcome as a model into which we introduce all the other risk factors as variables.

The baseline model for comparison is the null model. This is the model only including the outcome variable, the chlamydia test results. It is hypothesised that, based on the results of the latent class analysis, adding the variable "number of partners" to the model should have a substantial effect on the log-likelihood.

The null model had a log-likelihood of -317.4. Adding the total number of partners in the last year to the model significantly increased the log-likelihood to -273.5, ($p<0.0001$ in a likelihood ratio test).

However, adding further variables should have little effect. An additional variable was considered to have significantly improved the model if the likelihood ratio test result was significant. Table 15 shows the results of the modelling exercise.

As predicted, the largest change in the log-likelihood occurs when we move from the null model to the model including number of partners. Adding further variables does not significantly improve the model. The exception to this is condom use, which did generate a significant log-likelihood ratio test result. Therefore we also looked at a model that included condom use but not number

Table 15. Results of logistic regression on chlamydia urine test results

Model	Log-likelihood	Likelihood ratio test result comparing to model with number of partners only
Null model	-317.4	
Model with number of partners	-273.5	
Model with number of partners and new partnership	-270.4	0.05
Model with number of partners and STD diagnosis	-271.5	0.05
Model with number of partners and concurrency	-272.9	0.57
Model with number of partners and condom use	-268.1	<0.01

of partners. The log-likelihood of such a model was -296.8. The change from the null model was also highly significant ($p<0.0001$), however, the effect was less than the model with number of partners only, where the log-likelihood fell to -273.5. This indicates that the largest effect on the model is the addition of the variable measuring number of partners but that adding a variable measuring condom use does further improve our ability to predict whether an individual will contract chlamydia. It seems that using total number of partners as a simple measure holds up relatively well when applied to real data on chlamydia test results. But including data on condom use, where available, will provide even better predictions.

5. Discussion

5.1 Main findings

The results of the latent class analysis showed that the key factor in determining whether an individual engages in risky sexual behaviour is the number of partners he or she has had in the last year. Approximately 21% of the study population fell into this “risky” category having had two or more partners in the last year (suggesting that risky behaviour is relatively prevalent in the general population), 8% had not had any sexual partners in the last year, whilst 71% had one partner.

On further analysis by age group, risky behaviour was more prevalent in the youngest age group, 16-24 years, than in the older age groups of 25-34 years and 35-44 years. The prevalence of risky behaviour decreased with age from 38% in the youngest group to 20% in the middle group and 13% in the oldest group. This trend remained even after controlling for ethnic group and marital status, although it could not be determined whether this was an age effect or a cohort effect.

Single people had the highest prevalence of risky behaviour (39%) but were closely followed by those who had been previously married (31%). Married and cohabiting individuals were well below the population prevalence of 21% with

6% and 13% respectively. After controlling for the effects of age and ethnic group, this strong effect of marital status remained. Married and cohabiting people had a much lower prevalence of risky behaviour than would be predicted by their age/ethnic group distribution. In contrast, single and previously married people had a much higher prevalence of risky behaviour than their age/ethnic group distribution would predict.

Amongst the four ethnic groups identified in the study, the highest prevalence of risky behaviour was in the Black ethnic group (25%). This was closely followed by the White ethnic group (21%). The prevalence in the Indian and Pakistani groups was much lower, 14% and 13% respectively. The chance of falling into the “risky” class in the Black and White groups could be predicted almost exactly using their age and marital status distributions. This suggests that for the White and Black ethnic groups, ethnicity may not be a key factor in predicting risky sexual behaviour. For the Indian and Pakistani groups, however, the actual prevalence of risky sexual behaviour was lower than would have been predicted from their age and marital status distributions. For these groups, there may be something about their ethnicity which is protective.

5.2 Other studies

The literature review presented in Section 3 identified a number of studies which had found that having multiple sexual partners was an important risk factor for bacterial STD transmission, although no studies were found which had applied latent class methods to arrive at this conclusion.

Although this study agrees with those results, it would actually go further and argue that not only is number of sexual partners in the last year an important variable, it is the key variable in differentiating between those at risk and those not at risk in the UK. If we know how many partners an individual has had in the last year, we do not need any additional information to decide whether the behaviour puts them at risk of chlamydia or gonorrhoea infection, although having additional information on condom use does improve the accuracy of predicting whether an individual will actually go on to test positive for chlamydia..

In the primary analysis of the NATSAL I data, Johnson et al.(1994) reported that the highest prevalence of “unsafe sex” was found in the group of widowed, divorced and separated individuals when compared to other marital status groups, with the previously married individuals six times more likely to report unsafe sex than those who are married. They defined having unsafe sex as having two or more partners in the last year but never using a condom in that time. This definition included condom use as a variable, which the results of the latent class analysis do not. However, it arrived at similar conclusions regarding the increased risk of the previously married group.

5.3 Further research

The aim of this study was to define risky sexual behaviour with reference to chlamydia and gonorrhoea in the UK. An obvious area for further research would be to extend this work to look at risky sexual behaviour in the context of other diseases and other countries. For instance, it might be interesting to explore whether the differences in transmission and duration of viral STDs translate into a different risky behaviour profile to the one that we have found for bacterial STDs. The picture might also look different if we were looking at a country other than the UK. In developing countries where HIV has become endemic, condom use might emerge as far more important than the number of partners.

For this study, data were only available on the behaviour of individual respondents to the survey. However, it might be interesting for another study to explore the effect of partnership networks on STD risk. An individual may be engaging in what they think is safe behaviour because they think that their partner is safe. However, if the partner is engaging in risky sex, then by only measuring the individual's sexual behaviour, we would be underestimating their disease risk.

This study seems to highlight a large discrepancy between married and previously married people in the same age group and ethnic group. There seems to be something about not being married anymore which is associated with riskiness. Is it because divorced people suddenly find themselves free and single again? Is it because in their efforts to find a new partner, they feel too unsure of themselves to negotiate safe sex? Or is it their risky behaviour which prompted the divorce in the first place? Qualitative work to explore the effect of the transition from being married to being divorced and its effects on behaviour could shed light on the risky behaviours of a group that has not previously been targeted by interventions to reduce risky behaviour.

It was noted above that although the prevalence of risky behaviour seems to decrease with age, it was not clear whether this was an age or a cohort effect. It is hoped that time series data will become available which will allow further analysis in the future. Another round of NATSAL is planned in 2010 and perhaps that will allow us to begin looking at trends over the 20 years since NATSAL I in 1990.

5.4 Data limitations

5.4.1 Participation bias

Because sexual behaviour requires the provision of personal and often intimate information, some people may be more willing to participate in the research than others. This can lead to participation bias if there are systematic differences, for example in age, sex or social class, between those who agree to participate and those who do not (Fenton et al., 2001b).

In NATSAL II there were more female than male respondents, with males in the 25-29 age group particularly under-represented. However, this group generally tend to be under-represented in surveys, and also in the UK census (Office for National Statistics, 2001). And in spite of efforts to over-sample for predicted non-response in London, London residents were still under-represented (Erens et al., 2001).

The studies used in Section 3 were also subject to participation bias, as the majority of them were carried out in sexually transmitted disease clinics. People will generally attend an STD clinic if they think that they have an STD. Thus this group may have a higher prevalence of risky behaviours than the general population and also may differ in important socio-demographic ways. As a result, the findings might not be representative of the wider target population (Fenton et al., 2001b; Aral, 2004).

5.4.2 Item response bias

Even amongst those who agree to participate in a study, not all questions will be answered. Item response bias can arise where the people who choose not to answer a question have risk behaviours which are systematically different to the behaviour of those who elect to answer it (Fenton et al., 2001b).

A detailed study of the NATSAL I responses showed that those who were older, had problems with comprehension and came from certain ethnic groups were more likely to skip the more intimate questions. However, these groups were also more likely to be engaged in lower risk behaviours (Copas et al., 1997).

No study has been done to determine whether, or to what extent, the questions asked in the studies in the literature review suffer from item response bias.

Although it is impossible to estimate how they might have been affected by item response bias, it is likely that to some extent they do. Where responses were sought in face to face interviews rather than using questionnaires or CASI, it is possible that there may have been increased bias and a decreased tendency to disclose personal information.

5.4.3 Recall bias

Cross-sectional surveys, such as NATSAL II and the studies included in the review, ask people to recall past their recent behaviours. The reliability of the responses received can vary between people in important ways. Previous studies have found that the accuracy of recall varied by age, number of partners, ethnicity, number of sexual partners and how far back participants were asked to remember (Fenton et al., 2001b)

A particular problem has been identified in the recall of condom use. Individuals often struggle to recall, except over very short intervals, how often they used a condom with their partners and whether a condom was used with all partners. Questions on condom use triggered the largest numbers of inconsistencies in the NATSAL data, where for example individuals reported no condom use in the last year but then did report condom use with an individual partner. Zenilman et al. (1995) noted that not only do individuals struggle to recall condom use accurately but they also may only report on “use” rather than “correct use”. So condom breakages or slippages, for instance, which would increase STD risk would not be reported and the strength of any association diluted.

5.4.4 Publication bias

An additional source of bias in the literature review is publication bias. Researchers who find significant associations are more likely to pursue publication and possibly to be published. Thus it is possible that studies which find increased or decreased risk are not being balanced out by those that indicate no change in the level of risk. This would lead us to believe that there is stronger evidence for an association than may actually be the case.

5.4.5 Implications for results

None of these potential forms of bias will affect the response patterns uncovered by the latent class analysis. However, participation and item-response bias might affect the generalisability of the latent class prevalences to the general population if a study was not deemed to be representative.

Every effort was made to reduce participation bias in NATSAL II through methods to increase the response rate. For example, advance letters were sent to all homes, interviewers made repeated calls, small rewards offered for participation. Ultimately NATSAL II achieved a response rate of 64% and a sample that was broadly representative of the British population as compared to 1999 census population estimates.

Methods were also employed in NATSAL II to improve item response rates. As noted in Section 3 above, the implementation of CASI improved data quality and reduced the number of skipped questions.

Whilst there is no way to be certain that individuals have accurately reported their past experiences, the survey questions were carefully designed and piloted in order to maximise reliability of responses. NATSAL included 158 internal consistency checks to help researchers assess the reliability of responses received. These checks have shown that respondents tended to complete questions consistently. Around 70% of respondents had no inconsistencies.

Even if a small amount of uncertainty remains about the generalisability of the prevalence estimates or the reliability of the information received, NATSAL II is still an extremely useful tool. It is one of the only sources of information on sexual behaviour designed as a probability sample survey of the general population. Whilst it is important to be aware of any biases that may arise in using it, efforts have been made throughout the design process to address potential sources of bias and issues regarding reliability.

The results of any systematic review are only as good as the studies from which they are drawn. Every effort was made only to select high quality studies published in peer-reviewed journals. Any bias in the original work, however, will have made its way into the results of this review. There is no way to correct for this at the review stage and it must simply be acknowledged that there are some threats to the generalisability and reliability to consider when looking at the results. Similarly, there is no way to predict how or to what extent the review is subject to publication bias.

5.5 Methodological limitations

"As data on sexual behaviour accumulate, the interdependencies among specific sexual behaviours and between epidemiological parameters and behaviours become increasingly clear." (Aral, 2004, p. 10) One of the assumptions made by latent class analysis is that the manifest variables are conditionally independent – that all their covariance is explained by the underlying risky sexual behaviour variable. But there seems to be increasing evidence that there may be a degree of interdependence between manifest variables which cannot be accounted for solely through the latent variable. For example, individuals with more sexual partners may be more likely to use a condom. People who have previously been diagnosed with herpes may use a condom with all partners to prevent transmission. It is possible that there may be local dependence.

Generally, we can account for local dependence by increasing the number of latent classes until conditional independence holds. However, it is sometimes necessary to incorporate local dependence explicitly in the model. If you do not, the model fit statistics will be too high and you will end up adding latent classes when you do not really need them. This means that the extra latent classes do not reflect genuine subgroups.

There are diagnostic and modelling techniques specifically to address this issue but they have not been used here. It was decided that since a three-class model fit the data well, and the “risky” class is highly stable, there was little danger that superfluous latent classes were being added simply to satisfy the conditional independence criteria. It might be interesting for a future study to explore the extent to which there is local dependence amongst the manifest variables and which ones are affected. It could then employ appropriate techniques to see whether controlling for these relationships alters the latent class results.

As discussed in Section 3, a number of weights were applied to the NATSAL study population to control for the under- or over-representation of certain groups. However, it was not possible to apply these weights to the data in the latent class analysis. Although this would not have had an effect on the specification of the classes and the conditional probabilities of class membership, it might have affected the latent class prevalences, though it is not possible to tell in which direction.

5.6 Implications of the results

This paper has found that the key to determining whether an individual engages in risky sexual behaviour is the number of partners that they have had in the last year. This has important implications for how researchers interested in bacterial STDs conduct future studies.

For some categorical variables, there is a clinical guidance that helps us to decide how to define the categories. For example, hypertension is a diastolic blood pressure reading above 90 mm/hg and a systolic pressure reading of greater than 140 mm/hg (Carretero and Opartil, 2000). The threshold for obesity starts from a Body Mass Index (BMI) of 30, whilst a BMI of 25 or more means a person is overweight (World Health Organisation, 2000). Of course this does not mean that there is no debate about these definitions but they are generally held to be clear guidelines and a study that chooses not to use these measurements will generally justify this decision.

Things are less clear for non-clinical variables. What is a risky number of partners – is it more than one or more than three? Different studies have used very different definitions (see Appendix) and this can make comparisons between studies difficult. What the latent class analysis in this study makes clear is that risky behaviour is defined by having more than one partner. Adopting this definition, as we have done with BMI or blood pressure, could ensure that when researchers talk about risky behaviour, they are all talking about the same measure.

Being able confidently to use this single measure rather than a combination of measures would also make life easier for researchers and participants, ensuring that fewer and less personal sexual behaviour questions have to be asked. Intrusion into personal lives is really only ethical if it adds substantially to our understanding of risk behaviours. This study suggests that it does not and that by simply asking people “How many sexual partners have you had in the last year?” we can predict their STD risk almost as well as if we probed further into condom use, concurrency, etc.

As a measure, any variable is useful only to the extent that it is accurately reported. It may seem to be a key variable in a latent class analysis but if it is not a valid or reliable measure then it is not a useful indicator. Recall of the number of partners in the last year is generally good. “Test-retest” studies have investigated whether people are able consistently to give the same response on different occasions. These have found that a high percentage of people are consistent in their responses about the number of partners they have had, especially if they have had one partner or no partners (Van Duynhoven et al., 1999; Jaccard et al., 2004).

The usefulness of this study extends beyond its call to adopt a simple, uniform measure for risky sexual behaviour. It also expands our understanding of the distribution of risky sexual behaviour within key groups, which can in turn inform efforts to reduce STD prevalence or incidence through public policy.

Current government policy with respect to STDs includes measures to specifically target groups which they have identified as “at risk” especially young people and black and ethnic minority groups (Health Protection Agency, 2005). This study has shown that young people are indeed a key group with a higher prevalence of risky behaviour than their older counterparts.

The story is quite different for Black and ethnic minority groups. The prevalence of risky behaviour in the Black ethnic minority group was slightly higher than in the White group but this prevalence could be predicted by their age and marital status alone. There seems to be no indication that being Black implies riskier behaviour.

However, the National Chlamydia Screening Program and the Gonococcoal Resistance to Antimicrobials Surveillance Programme both found a substantially higher infection rate amongst Black participants than other ethnic groups. (Health Protection Agency, 2005). This study has indicated that a higher prevalence of risky behaviour is not likely to be the explanation, which has important implications for the design of interventions to reduce the infection rate. Considerations besides behaviour change are needed. For example, Laumann and Youm (1999) found that the higher rates of bacterial infections amongst African Americans could be explained by the patterns of sexual networks between different ethnic groups. “Safe” African Americans are more likely than White Americans to have had a “risky” partner in the past five years. Rates can also be affected by the prevalence of the disease in the population. With higher case rates, there is a higher probability that one individual in a Black couple is infected (and may not even know it).

There is a key group missing from the Government's proposals. This study has identified that previously married individuals have a high prevalence of risky behaviour, as did the initial analysis of NATSAL I (Johnson et al., 1994). With 167,116 divorces in 2004, large numbers of people enter into this group every year and potentially place themselves at risk of an STD (Office for National Statistics, 2005). However, little is known about why this group behaves as it does and further research is needed in inform the design of effective interventions to reduce risky behaviour.

Although number of partners in the last year may be a good indicator with which to identify at risk groups, it may seem a poor one on which to base a public health intervention. A health campaign that encouraged "avoid chlamydia and gonorrhoea: only have one sexual partner each year" would be laughable. Partnership formation and breakdown is largely divorced from disease risk. It is determined by the nature of each relationship and concepts such as love, trust and fidelity. To try to discourage partnership turnover would be pointless.

However, awareness of the importance of partnership turnover is useful because it provides a simple way for each person to assess their own risk. For instance, encouraging people who have had more than one partner to get tested for chlamydia and gonorrhoea could be an effective way to reduce disease prevalence. To help reduce incidence, it could target the 52% of people who have more than one partner but to not use condoms to change their behaviour, combining the message on partnership turnover with condom use. Through the media, we receive messages about our health every day and it can be too easy to ignore them. It is not difficult to understand why the Government would prefer to target certain groups, ensuring that the message is marketed to them in the most effective way possible. However, using a single, simple measure, it is possible for everyone to assess their own risk of infection and to take responsibility for their sexual health.

REFERENCES

Aral S (2004). "Sexual Risk Behaviour and Infection: Epidemiological Considerations". Sexually Transmitted Infections 80 (Suppl. ii): 8-12.

Austin H, Louv WC and Alexander J (1984). "A Case-Control Study of Spermicides and Gonorrhoea". JAMA 251:2822-2824.

Barlow D (1977). "The Condom and Gonorrhoea". Lancet 310: 811-813.

Bjekic M, Vlaginac H, Sipetic S and Marinkovic J (1997). "Risk Factors for Gonorrhoea: case-control study". Genitourinary Medicine 73: 518-521.

Burstein G, Zenilman J, Gaydos C, Diener-West M, Howell M, Braithwaite W and Quinn T (2001). "Predictors of repeat *Chlamydia trachomatis* infections diagnosed by DNA amplification testing among inner city females". Sexually Transmitted Infections 77: 26-32.

Carretero O and Opartil S (2000). "Essential Hypertension: Part 1 Definition and Etiology". Circulation 101: 329.

Copas A, Johnson A, and Wadsworth J (1997). "Assessing Participation Bias in Sexual Behaviour Surveys: implications for measuring HIV risk". AIDS 11: 783-790

D'Oro L, Parazzini F, Naldi L and Vecchia C (1994). "Barrier Methods of Contraception, Spermicides and Sexually Transmitted Diseases: a review". Genitourinary Medicine 70: 410-417.

Department of Health (2004). "Reid Announces £300m to Modernise Sexual Health Services" (26 November 2004).]http://www.dh.gov.uk/publicationsandstatistics/pressreleases/PressReleasesNotices/fs/en?CONTENT_ID=4096396&chk=p07Pcs. Downloaded 10 March 2006.

Egger M, Davey Smith G, and Phillips A (1997). "Meta-analysis: Principles and Procedures". British Medical Journal 315: 1533-1537.

Erens B, McManus S, Field J, Korovessis C, Johnson A, Fenton K, and Wellings K (2001). <u>National Survey of Sexual Attitudes and Lifestyles II: Technical Report</u>. National Centre for Social Research: London.

Fenton K, Korovessis C, Johnson A, McCadden A, McManus S, Wellings K, Mercer C, Carder C, Copas A, Nanchahal K, Macdowall W, Ridgway G, Field J and Erens B (2001a). "Sexual Behaviour in Britain: Reported Sexually Transmitted Infections and Prevalent Genital *Chlamydia Trachomatis* Iinfection". Lancet 358: 1851–1854.

Fenton K, Johnson A, McManus S and Erens B (2001b). "Measuring Sexual Behaviour: Methodological Challenges in Survey Research". Sexually Transmitted Infections 77: 84-92.

Fenton K, Mercer C, McManus S, Erens B, Wellings K, Macdowall W, Byron C, Copas A, Nanchahal K, Field J and Johnson A (2005). "Ethnic Variations in Sexual Behaviour in Great Britain and Risk of Sexually Transmitted Infections: a Probability Survey". Lancet 365: 1246-1255.

Fortenberry J, Brizendine E, Katz B, Wools K, Blythe M and Orr D (1999). "Subsequent Sexually Transmitted Infections Among Adolescent Women in with Genital Infection Due to *Chlamydia trachomatis*, *Neisseria gonorrhoeae* or *Trichomonas vaginalis*". Sexually Transmitted Diseases 26: 26-32.

Gershman K and Barrow J (1996). "A Tale of Two Sexually Transmitted Diseases: Prevalences and Predictions of Chlamydia and Gonorrhoea in Women Attending Colorado Family Planning Clinics". Sexually Transmitted Diseases 23: 481-487.

Gunn R, Fitzgerald S and Aral S (2000). "Sexually Transmitted Disease Clinic Clients at Risk for Subsequent Gonorrhoea and Chlamydia Infections". Sexually Transmitted Diseases 27: 343-348.

Hart G (1992). "Factors Associated with Genital Chlamydial and Gonoccocal Infection in Females". Genitourinary Medicine 68 :217-220.

Hart G (1993). "Factors Associated with Genital Chlamydial and Gonoccocal Infection in Males". Genitourinary Medicine 69: 393-396.

Health Protection Agency (2005). Mapping the Issues: HIV and other sexually transmitted diseases in the United Kingdom 2005. Health Protection Agency Centre for Infections: London.

Hijazi L, Throw C, and Winter AJ (2002). "Factors Affecting Co-Infection with Genital Chlamydia and Genital Gonorrhoea in an Urban Genitourinary Medicine Clinic". Sexually Transmitted Infections 78: 386-390.

Hughes G, Catchpole M, Rogers P, Brady A, Kinghorn G, Mercey D, and Thin N (2000a). "Comparison of Risk Factors for Four Sexually Transmitted Infections: results from a study of attenders at three genitourinary medicine clinics in England". Sexually Transmitted Infections 76: 262-267.

Hughes G, Andrews N, Catchpole M, Goldman M, Forsyth-Benson D, Bond M, and Myers A (2000b). "Investigation of the Increased Incidence of Gonorrhoea Diagnosed in Genitourinary Medicine Clinics in England, 1994-1996". Sexually Transmitted Infections 76:18-24.

Jaccard J, Macdonald R, Wan C, Guilamo-Ramos V, Dittus P and Quinlan S (2004). "Recalling Sexual Partners: the Accuracy of Self-reports". Journal of Health Psychology 9: 699-712.

Johnson A, Wadsworth J, Wellings K, Field J and Bradshaw S (1994). Sexual Attitudes and Lifestyles. Blackwell Scientific Publications: Oxford.

Jonsson M, Karlsson R, Rylander E, Gustavsson A, and Wadell G (1995) "The Associations Between Risk Behaviour and Reported History of Sexually Transmitted Diseases Among Young Women: a Population-based Study". International Journal of STD & AIDS 8: 501-505.

Lacey C, Merrick D, Bensley D and Fairley I (1997). "Analysis of the Sociodemography of Gonorrhoea in Leeds". British Medical Journal 314: 1715.

Latino M, Bello L, Lanza A, Leotta E, Tersiev P, Intinis G, Spagnolo E, Smirne C and Grio R (2002). "*Chlamydia trachomatis* infection among sexually active young women in Italy". Sexually Transmitted Infections 78.

Laumann E and Youm Y (1999). "Racial/Ethnic Group Differences in the Prevalence of Sexually Transmitted Diseases in the United States: a Network Explanation". Sexually Transmitted Diseases 26: 250-261.

Low N, Sterne J and Barlow D. "Inequalities in Rate of Gonorrhoea and Chlamydia Between Black Ethnic Groups in South East London: Cross-sectional Study". Sexually Transmitted Infections 77: 15-20.

Lyss S, Kamb M, Peterman T, Moran J, Newman D, Bolan G, Douglas J, Iatesta M, Mallotte C, Zenilamn J, Ehret J, Gaydoa C and Newhall W (2003). "*Chlamydia trachomatis* among Patients Infected with and Treated for *Neisseria gonorrhoeae* in Sexually Transmitted Disease Clinics in the United States". Annals of Internal Medicine 139:178-186.

Magidson, J., and Vermunt J.K. (2003) "A Nontechnical Introduction to Latent Class Models". http://www.statisticalinnovations.com/articles/lcmodels2.pdf. Downloaded 27 May 2006.

Magidson, J. and Vermunt, J.K.(2004a). "Latent class analysis". In Lewis-Beck M, Bryman A, and Liao TF (eds.), The Sage Encyclopedia of Social Sciences Research Methods. Sage Publications: California, pp. 549-553.

Magidson, J and Vermunt, J.K.. (2004b). "Latent Class Models". In Kaplan D (ed.), The Sage Handbook of Quantitative Methodology for the Social Sciences. Sage Publications: California, pp. 549-553.

McCutcheon A (1987). Latent Class Analysis. Sage Publications: London.

Mehta S, Erbelding E, Zenilman J and Rompalo A (2003). “Gonorrhoea Reinfection in Heterosexual STD Clinic Attendees: Longitudinal Analysis of Risks for First Reinfection”. Sexually Transmitted Infections 79: 124-128.

Mertz K, Finelli L, Levine W, Mognoni R, Berman S, Fishbein M, Garnett G and Louis M (2000). “Gonorrhoea in Male Adolescents and Young Adults in Newark, New Jersey”. Sexually Transmitted Diseases 27:201-207.

Niccolai L, Rowhani-Rahbar A, Jenkins H, Green S and Dunne DW (2005). “Condom Effectiveness for Prevention of *Chlamydia trachomatis* Infection”. Sexually Transmitted Infections 81: 323-325.

Nyland K (2005). “Latent Class Analysis in Mplus Version 3”. http://www.ats.ucla.edu/STAT/mplus/seminars/lca/default.htm. Downloaded 3 December 2005.

Office for National Statistics (2001). “One Number Census Quality Assurance Information: Quality Assurance Themes”. http://www.statistics.gov.uk/census2001/pdfs/1991_underenumeration.pdf. Downloaded 28 July 2006.

Office for National Statistics (2005). “Divorces fell slightly in 2004 in England and Wales”. http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=14124. Downloaded 21 July 2006.

Pemberton J, McCann JS, Mahony JDH MacKenzie G, Dougan H and Hay I (1972). “Socio-Medical Characteristics of Patients Attending a V.D. Clinic and the Circumstances of Infection”. British Journal of Venereal Disease 48: 391-396.

Radcliffe K, Ahmad S, Gilleran G and Ross J (2001). “Demographic and Behavioural Profile of Adults Infected with Chlamydia: a Case-control Study”. Sexually Transmitted Infections 77: 265-270.

Ramstedt K, Forssman L, Gieskecke J and Granath F (1992). “Risk Factors for *Chlamydia Trachomatis* Infection in 6810 Young Women Attending Family Planning Clinics”. International Journal of STD & AIDS 3: 117-122.

Rietmeijer C, Van Bemmelen R, Judson F and Douglas J (2002). “Incidence and Repeat Infection Rates of *Chlamydia Trachomatis* Among Male and Female Patients in an STD Clinic”. Sexually Transmitted Diseases 29: 65-72.

Rosenberg MJ, Davidson AJ, Chen JH, Judson FM and Douglas JM (1992). “Barrier Contraceptives and Sexually Transmitted Diseases in Women: a Comparison of Female-Dependent Methods and Condoms”. American Journal of Public Health 82: 669-674.

Skrondal A and Rabe-Hesketh S (2004). Generalised Latent Variable Modelling. Chapman and Hall: London.

Storr C, Hongling Z, Kung-Yee L, and Anthony J (2004). "Empirically Derived Latent Classes of Tobacco Dependence Syndromes Observed in Recent-Onset Tobacco Smokers: Epidemiological Evidence from a National Probability Sample Survey". Nicotine and Tobacco Research 6: 533-545.

Uebersax JS (2001). "Latent Class Analysis – Frequently Asked Questions". http://ourworld.compuserve.com/homepages/jsuebersax/local.htm. Downloaded 17 January 2006.

Upchurch D, Brady W, Reichart C, and Hook E (1990). "Behavioural Contributions to Acquisition of Gonorrhoea in Patients Attending an Inner City Sexually Transmitted Disease Clinic". Journal of Infectious Diseases 161: 938-941.

Van Duynhoven Y, Nagelkerke N, and Van de Laar M (1999). "Reliability of Self-Reported Sexual Histories: Test-Retest and Interpartner Comparison in a Sexually Transmitted Diseases Clinic". Sexually Transmitted Diseases 26: 33-42.

Vermunt JK (1997). LEM: a general program for the analysis of categorical data. Department of Methodology and Statistics: Tilburg University.

Vuylsteke B, Vandenbruaene M, Vandenbalcke P, Van Dyck E and Laga M (1999). "*Chlamydia Trachomatis* Prevalence and Sexual Behaviour among Female Adolescents in Belgium". Sexually Transmitted Infections 75: 152-155.

Weinstock H, Bolan G, Kohn R, Balladares C, Back A and Oliva G (1991). "*Chlamydia Trachomatis* Infection in Women: a Need for Universal Screening in High Prevalence Populations?". American Journal of Epidemiology 135: 41-46.

Whittington L, Kent C, Kissinger P, Oh K, Fortenberry J, Hillis S, Litchfield B, Bolan G, St Louis M, Farley T and Handsfield H (2001). "Determinants of Persistent and Recurrent *Chlamydia Trachomatis* Infection in Young Women". Sexually Transmitted Diseases 28: 117-123.

Winter AJ, Sriskandabalan P, Wade AAH, Cummins C and Barker P (2000). "Sociodemography of Genital *Chlamydia Trachomatis* in Coventry, UK 1992-1996". Sexually Transmitted Infections 76: 103-109.

World Health Organisation (2000). Obesity: Preventing and Managing the Global Epidemic. World Health Organisation: Geneva.

Zenilman J, Hook E, and Shepherd M (1994). "Alcohol and Other Substance Use in STD Clinic Patients: Relationships with STDs and Prevalent HIV Infection". Sexually Transmitted Diseases 21:220-225.

Zenilman J, Weisman C, Rompalo A, Ellish N, Upchurch D, Hook E and Celentano D (1995). "Condom Use to Prevent Incident STDs: the Validity of Self-reported Condom Use". Sexually Transmitted Diseases 22: 15-21.

Appendix

Table A.1. Definition of "multiple partners"

Studies	Definition
• ***Chlamydia***	
Fenton (2001a)	2-4 partners in the last year
Gershman (1996)	More than 1 partner in the last 90 days
Hart (1992)	More than 1 partner
Hart (1993)	More than 1 partner
Hughes (2000)	3+ partners in the last year
Jonsson (1995)	2-3 lifetime partners
Latino (2002)	More than 1 partner in the last 6 months
Radcliffe (2001)	2+ partners in the last year
Vuylsteke (1999)	2+ lifetime partners
• ***Gonorrhoea***	
Bjekic (1997)	3+ partners in the last year
Hart (1992)	More than 1 partner
Hart (1993)	More than 1 partner
Hughes (2000)	3+ partners in the last year
Upchurch (1990)	2+ partners in last month

Table A.2. Definition of "short term relationship"

Studies	Definition
• ***Chlamydia***	
Fenton (2001)	1+ new partner in the last 12 months
Gershman (1996)	1+ new partner in the last 90 days
Hart (1992)	1+ partner, but no steady partner, in last 3 months
Hart (1993)	1+ partner, but no steady partner, in last 3 months
Ramstedt (1992)	1+ new partner in last 12 months
Weinstock (1991)	1+ new partner in last 3 months
• ***Gonorrhoea***	
Bjekic (1997)	1+ new partner in the last month
Hart (1992)	1+ partner, but no steady partner, in last 3 months
Hart (1993)	1+ partner, but no steady partner, in last 3 months
Mertz (2000)	Casual partner during preceding month
Upchurch (1990)	1+ new partner in the last month

Table A.3. Definition of "alcohol consumption"

Studies	Definition
• ***Chlamydia***	
Radcliffe (2001)	More than 5 units of alcohol per week
Vuylsteke (1999)	Drinking at the weekend and several times during the week
Zenilman (1994)	Drank more than 2 times in the last week
• ***Gonorrhoea***	
Bjekic (1997)	Frequent alcohol consumption
Zenilman (1994)	Drank more than 2 times in the last week