Journal of the Operational Research Society (2006) 57, 1400–1412
www.palgrave-journals.com/jors

Improving the cost-effectiveness of Chlamydia screening with targeted screening strategies

D Evenden[1], PR Harper[1]*, SC Brailsford[2] and V Harindra[3]

[1]School of Mathematics, University of Southampton, Southampton, UK; [2]School of Management, University of Southampton, Southampton, UK; and [3]St Mary's Hospital, Portsmouth, UK

Chlamydia is the most common sexually transmitted infection in the UK and constitutes a major public health problem. The UK Department of Health is phasing in a National Chlamydia Screening Programme (NCSP) but there is concern that blanket screening of the entire at risk population will simply add extra burden to the already overstretched health economy. This paper demonstrates that certain high-risk sub-groups within the general population are critical in the infection dynamics. Improved targeting of these high-risk populations achieves greater cost-effectiveness. Statistical risk-group clustering techniques have been used to identify indicators that are strong predictors in determining high-risk status while geomapping techniques visually display prevalence geographically across the region, thus identifying high prevalence postcode clusters and informing public health planners where to target intervention and screening strategies. A System Dynamics simulation model has been used to capture the infection dynamics and measure the cost-effectiveness of the intervention strategies. The model incorporates risk-group behaviour as identified by the above geomapping and statistical analysis components of the research. The combined use of computer simulation, statistical analysis and geomapping methodologies has provided a unique holistic view of the problem.

Journal of the Operational Research Society (2006) **57,** 1400–1412. doi:10.1057/palgrave.jors.2602134
Published online 28 December 2005

Keywords: healthcare; Chlamydia; system dynamics; geomapping; risk grouping

1. Introduction

Chlamydia trachomatis is the commonest sexually transmitted bacterial infection in the UK, with 89 431 diagnoses in Genito-Urinary Medicine (GUM) clinics in 2003 (Health Protection Agency, 2004), and has been widely reported in literature (Hicks *et al*, 1999; Department of Health, 2000; Hart *et al*, 2002; Honey *et al*, 2002) and recently in the media (eg, The Guardian, 2004). It constitutes a major public health concern. The majority of infections are asymptomatic, but can lead to serious long-term sequelae including pelvic inflammatory disease, tubal infertility and ectopic pregnancy. Screening programmes have been shown to be effective in Sweden and the United States (Herrmann *et al*, 1991; Addiss *et al*, 1993). However, the methodology employed may not be suitable in the UK and there are concerns that blanket screening of the whole population at risk will add extra burden to the overstretched health economy. Recently, the UK Department of Health provided funding to introduce national Chlamydia screening of people between the ages of 16–25 in 10 centres with the view to extend this programme to the rest of the country within the next few years as part of the National Chlamydia Screening Programme (NCSP) (Department of Health, 2005).

*Correspondence: *PR Harper, School of Mathematics, University of Southampton, Southampton, Hants S0171BJ, UK.*
E-mail: prh@maths.soton.ac.uk

Portsmouth was one of two UK Department of Health sites chosen for an opportunistic screening trial of Chlamydia, whereby 20 000 persons in the 16–24 age range were screened (Pimenta *et al*, 2003a, b). Portsmouth is an island city situated on the coast of Southern England and has a population of just under 200 000. Very high levels of population coverage were achieved in the opportunistic screening trial and this was regarded as an important factor in the success of future screening interventions. An infection prevalence of around 10% was observed, with an age peak noted at 18 years. Harindra *et al* (2002) provide more detailed insight of the methods and preliminary results from the trial.

This paper presents collaborating work with the University of Southampton and Consultants at the GUM Department, St Mary's Hospital, Portsmouth. The research was timely, given that the Portsmouth opportunistic trial had been completed and that it was felt that findings from this trial could help to inform the NCSP. The work was novel in crossing the boundaries of various disciplines, namely Operational Research, Statistics, Health Services Research and Geography. The methodologies adopted, as discussed in this paper, combined geomapping, statistical clustering methods, and System Dynamics (SD) modelling. The geomapping work, using the software MapInfo, allowed for the spreading patterns and infection clusters to be observed, and provided a critically important contribution to screening intervention planning. The analysis of socio-economic indicators, using regression models and tree-based classification trees, identified high-risk groups within the population for screening intervention targeting. The SD model, built using the software Vensim, captured the infection dynamics and cost-effectiveness of screening using strategies informed by the previous two components. Overall the multiple approach that was adopted, utilizing OR and statistical methods in combination with geomapping tools, facilitated a holistic view of the problem. Thus, the recommendations that emerged to help inform health policy were considered to be well founded. They were put forward to help form local policy to support the Chlamydia screening programme and to provide more general guidance and recommendations on a cost-effective screening methodology. Although we analyse and present data from the Portsmouth Chlamydia opportunistic screening trial, the methodology adopted here could be readily applied to other geographical areas, and indeed for other sexually transmitted infections or diseases.

Although there is an extensive literature on modelling infectious disease, there is relatively little published work on modelling Chlamydia infection. A review by Honey *et al* (2002) describes studies that show screening to be more cost-effective than just testing symptomatic women. The role of male partners, and the fact that men seem to be forgotten in the infection and treatment equation was recognized by Hart *et al* (2002), particularly where women were screened opportunistically. Previous papers to analyse the cost-effectiveness of screening have included Haddix *et al* (1995), Genç and Mardh (1996) and Buhaug *et al* (1989). The analyses reported in these papers ignore risk-groups within the population and the impact of screening on prevalence. Townshend and Turner (2000) developed a SD model that overcame many of these previous concerns, and was an excellent source of guidance for this work. Gove (1997) used a Discrete Event Simulation (DES) model to evaluate screening options for Chlamydia. The research presented here is novel in that we combine geomapping, risk groupings and computer simulation techniques, allowing for each component of the work to be informed by the other components. For example, within the SD model we have included different risk groups within the population based on sexual behaviour that were identified during the statistical clustering work. Furthermore, by considering the spatial prevalence of Chlamydia over the geographical region using geompaping techniques, we have been able to determining the relationship between socio-economic indicators and prevalence by postcode, which in turn informed the parameters for the SD model.

To summarize the key research objectives:

- Geographical mapping of Chlamydia prevalence in the Portsmouth region.
- Statistical determination of relationships between socio-economic indicators and prevalence by postcode.
- Identification of high-risk populations.
- Determination of those factors which help to plan and target screening and inform health education methods, in order to reduce the national incidence and prevalence of Chlamydia.

In order to meet the research objectives, the following methodologies were adopted and are discussed in subsequent sections of this paper:

- Preliminary analysis of the opportunistic screening trial data (Section 2).
- Geomapping analysis of the screening data (Section 3).
- Statistical analysis of prevalence and socio-economic indicators (Section 4).
- A comprehensive SD model of the Chlamydia infection process, to promote an understanding and justification of the targeted screening intervention (Section 5).

2. Analysis of the opportunistic screening data

Data were collected during the opportunistic screening trial held in the Portsmouth area from October 1999 to September 2000. A number of usable subsets of data were made available for analysis, including opportunistic screening data, GUM patients and partners tracing records, and referrals from positive results in the opportunistic screen.

Over 25 000 Chlamydia tests were carried out during this trial, though these included repeat tests (eg to check that the infection had been cured) and visits by the same person to different medical facilities (and thus getting a new patient number). These raw data were recorded in an Excel file consisting of 25 553 records. Even though this particular data contained no test results, it was useful in that it contained records of patient sex, ethnic group, and postcode, all referenced by a patient index P-number. These data also included the partner tracing details where this was possible. Although these data had patient name fields, these were deleted to ensure patient anonymity. Harindra *et al* (2002), and Pimenta *et al* (2003a, b), provide more insight into the design of the opportunistic screening trial in Portsmouth, particularly in the context of screening for co-infections and presumptive treatment. These initial findings suggested that the screening trial, and the heavy reliance on patient cooperation, was well received in this instance. The trial itself was seen to be highly effective.

The results data for the opportunistic trial consisted of 17 342 patient records. This was provided as an Excel file and consisted of patient number, the test date, date of birth, testing locations, laboratory specimen number, and test result. To find patient postcode, the patient number (eg 2234) was reformatted to the fixed format patient index P-number (eg P02 234) using Excel text manipulation functions. This could then be cross-referenced to the 25 553 records of the raw data to extract postcode and patient details. When records were sorted by postcode in the Portsmouth area, there were just over 11 100 records pertaining to a valid postcode district (eg PO1), and just under this number pertaining to a valid postcode sector (eg PO1 1). Figure 1 shows how the various datasets were extracted and manipulated to obtain the final data set to be analysed.

The screened percentage in each postcode district is shown in Figure 2. In order to calculate these figures, we needed first to obtain census population statistics for 30 202 women in the 16- to 24-year-age group for the PO postcode areas, which were extracted from the UK Office of National Statistics (ONS) census data (http://www.statistics.gov.uk). It was then necessary to convert the Census data from electoral ward to postcode district in order to permit comparison with the number of tests by postcode district. As shown in Figure 2, the values achieved were large (typically 30–40%) and this provided confidence that further calculations based on these data gave results that were representative of the overall population.

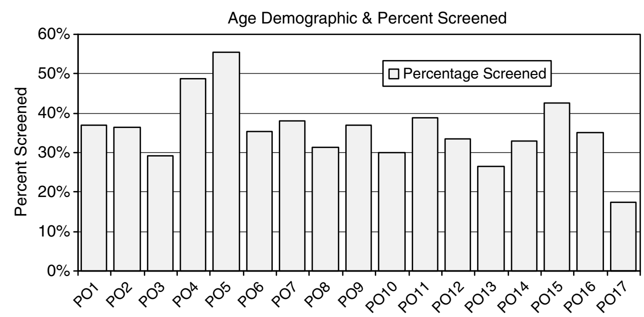

Figure 2 Screening penetration: percentage of 16- to 24-year-old population screened by postcode district.

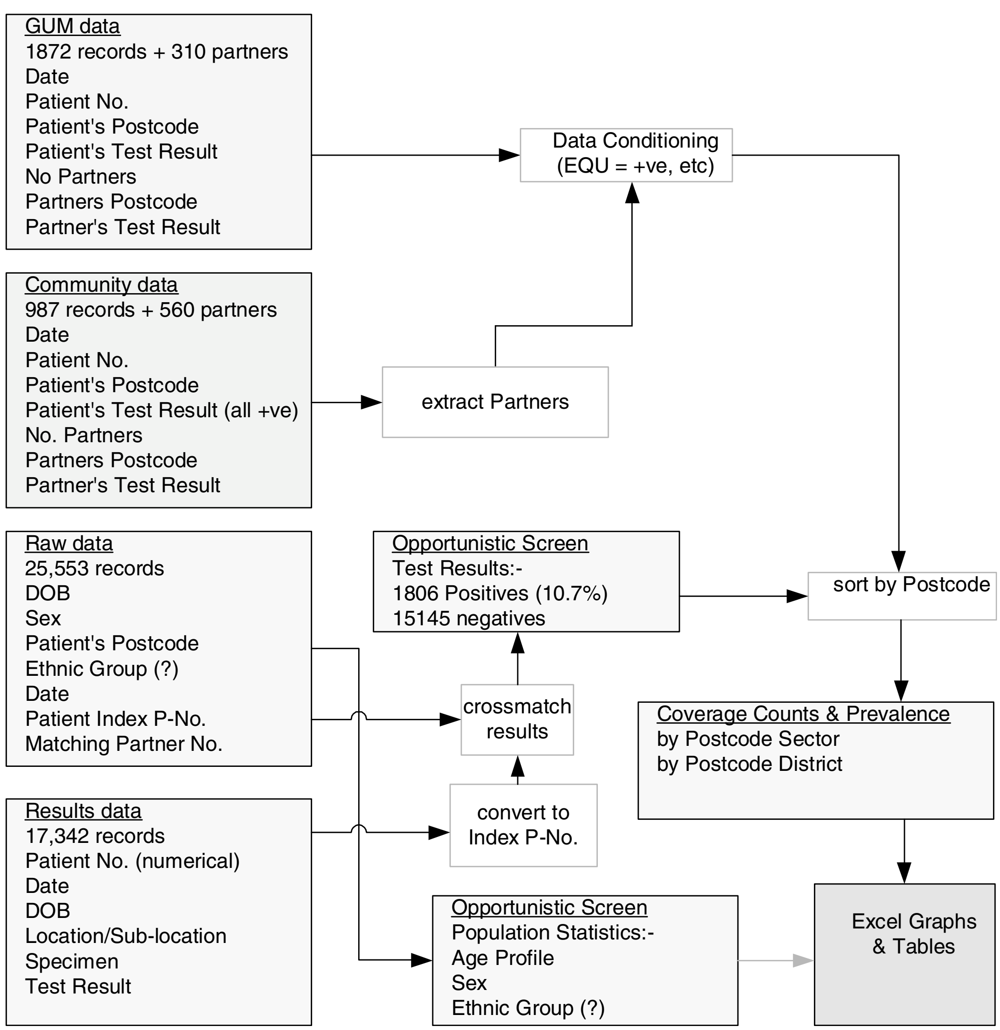

Figure 1 Screening data extraction and manipulation.

Table 1 Prevalence by patient type

Patient group	Prevalence (%)	95% CI (%)	Sample
GUM patients	17.8	1.8	1632
Partners of all GUM patients	23.0	5.1	254
Partners of positive GUM patients	51.9	7.1	77
Partners of community referrals	45.9	4.4	488
Opportunistic Screen—PO only	9.07	0.53	11 140
Opportunistic Screen—raw data	10.41	0.45	17 342

Infection prevalence was calculated by patient type and is shown in Table 1. When the number of positive results or the sample size was small the exact method based on the binomial distribution was used. Equations (1) and (2) provided an exact solution, where N_D was the number of positives in a group of N tests, with prevalence p, and where k was the summing variable.

For the lower CI limit (1−α)100%:

$$\mathrm{Binom}(k;\, N,\, p) = 0.975 = \sum_{k=0}^{N_D-1} \binom{N}{k} \times p_L^k \times (1-p_L)^{N-k} = 1 - \alpha/2 \tag{1}$$

For the upper CI limit (1−α)100%:

$$\mathrm{Binom}(k;\, N-1,\, p) = 0.025 = \sum_{k=0}^{N_D} \binom{N}{k} \times p_U^k \times (1-p_U)^{N-k} = \alpha/2 \tag{2}$$

This shows that prevalence among GUM patients was significantly higher than the opportunistic screen prevalence, suggesting that GUM users constitute a higher risk group. The prevalence among partners of GUM patients and partners of community-screened patients was consistent.

Of 17 342 records with patient sex and a test result available 12 653 were cross-matched. Of these 12 454 (98.43%) were female of which 8.97% tested positive. Only 199 (1.57%) records were male of which 16 (8.04%) tested positive. These results are shown in Figure 3, along with 95% confidence intervals.

Since the confidence intervals overlap it suggested that prevalence among male and female groups was not significantly different. This was confirmed using a χ^2 test for homogeneity where $\chi^2 = 0.2$, df = 1, and P-value = 0.65, suggesting homogeneity between male and female results (Yates' correction to this method provides the same conclusions). For this reason, and since the number of men in the sample was small, no further distinction was made between male and female records.

Figure 4, based on 11 140 patient records partitioned by postcode district, shows the prevalence of infection, with confidence intervals calculated for each district. These confidence intervals varied in width because the partitioned sample size within each district was different. In particular PO17 has a very large confidence interval due to the small sample.

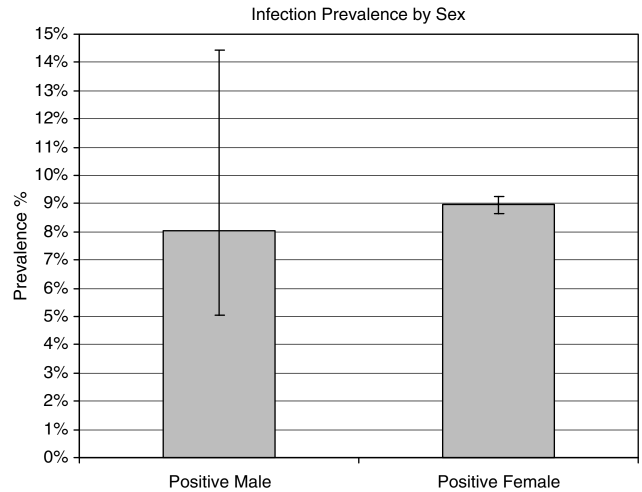

Figure 3 Prevalence by sex.

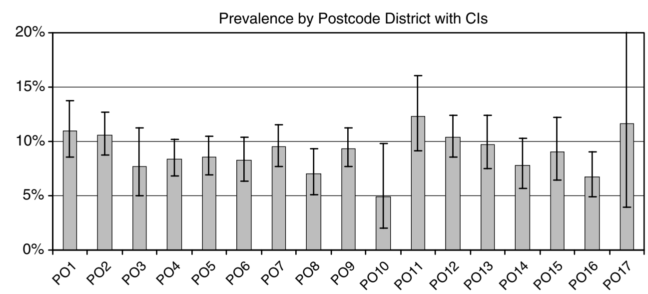

Figure 4 Opportunistic screen infection prevalence by postcode district.

The age profile of patients tested, their test results, and prevalence within each age group were investigated, using 16 411 records for persons aged between 11 and 40 years. Figure 5 shows the prevalence calculated from the above data with confidence intervals for each estimate. A peak in infection prevalence is shown at the age of 18 years.

GUM patients and community referral patients were asked to provide the postcode of the most recent partners. Where these data were available it was possible to assess the pattern of relationships across the postcode structure. The number of relationships which occurred in the same postcode were counted, and also within the same postcode sector and district. Using a VBA program it was possible to search for the number of partnerships in adjacent districts. This largest proportion of partnerships was found to be within the immediate geographic area of the patient, with a significant proportion in the same postcode (which includes the same address). Figure 6 shows partnership location for

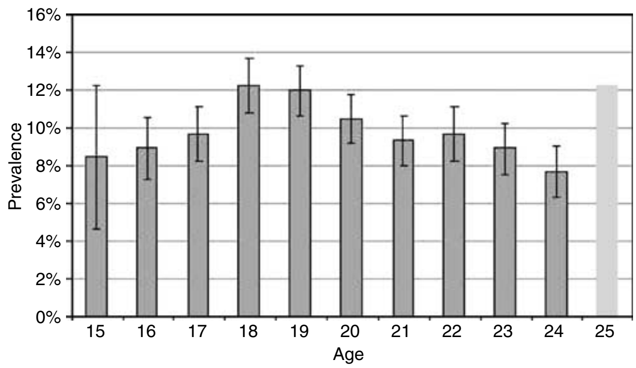

Figure 5 Age and prevalence distribution.

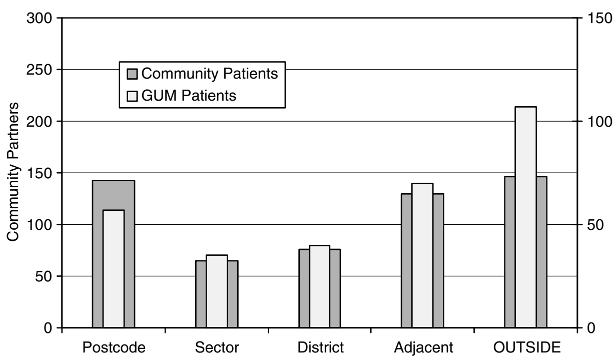

Figure 6 Partner locations by patient type.

both community referrals and GUM patients. Partnerships within the same postcode sector, district or adjacent district are proportionally similar for both patient types, and these satisfied a homogeneity test ($P = 0.995$). This suggested that the behaviour of the GUM patients and community patients was similar in the distribution of partnerships in the sectors, districts and adjacent districts. GUM patients have proportionally fewer partners in the same postcode and proportionally more partners outside these regions. An independence test on these two categories shows that this difference is significant ($P = 0.002$). Therefore, we conclude that GUM patients behave differently compared to the rest of the community, whose partners tended to be closer to home.

The GUM data and community referrals data sets allowed partnership change to be assessed. From the GUM and community data 1872 patients and 987 patients, respectively, indicated the number of partners in the last 3 months, shown in Table 2. One GUM patient was recorded with 12 partners and another with 20 partners.

To summarize key findings from the preliminary data analysis:

- The 17 553 screening trial results constituted a huge sample of data. The conclusions drawn from the analysis can be regarded as representative of the Portsmouth region.
- Screening rates were high. Typically 30–40% of the 16- to 24-year-age group was included. Over the 1-year screening trial a monthly average of nearly of 1500 patients were processed.
- Measured prevalence rates were high. For all Chlamydia tests 10.41% were positive and 12.37% were positive or equivocal (where a positive result could not be confirmed).
- Prevalence among males was not statistically different to that among females. Consequently, male partners should not be ignored in a screening programme of female patients. In addition, infection rate among partners was high in both the GUM and community referrals data sets. Partner tracing and treatment is important in the screening strategy to mitigate the risk of re-infection.
- At the postcode district level (eg PO1) prevalence varies between 4.9 and 12.3%. At the postcode sector level (eg PO1 1) prevalence varies between 3.3 and 18.1%. This analysis forms the basis for the geomapping work (Section 3).
- The age profile showed a clear prevalence peak at the age of 18 years.
- GUM patients have been shown to exhibit a different behaviour pattern with a greater proportion of partners outside of their immediate vicinity. It was felt that increased mobility does have significant effects on the risk dynamics, and this may be related to higher levels of disposable income.
- The partnership change frequency was almost identical between GUM patients and community referrals.

Table 2 Partnership frequencies (number of partners over a 3-month period)

Number of partners	*GUM patients*	*Community referrals*
0	111	49
1	1423	824
2	294	97
3	33	13
4+	12	5

3. Geomapping analysis

As part of the preliminary data analysis, as described above, results by postcode district were reported and presented in tables and bar chart format to Consultants at Portsmouth. This approach had two main disadvantages. The first was that postcodes were found to be fairly anonymous descriptors unless one happens to be very familiar with the area. The other problem was that there were many tens of postcode sectors. At this greater level of detail it was difficult to produce meaningful, easily interpreted bar charts. This provided the motivation for geomapping.

Geomapping is the art of representing data or measurements superimposed on a map of the area to which it relates. It is a powerful technique since the data values are placed in location context. The maps themselves may be conventional street maps, ordnance survey maps, or other representations, which may even include terrain elevation with 3D perspective views. For the purpose of this research project, it was sufficient to use two approaches; ordnance survey maps, which place the data in an effective and familiar context; and postcode polygon maps, to clearly display the relationships between the data. The MapInfo software tool (www.mapinfo.com) was used to manage, manipulate, and display the mapping data. Plotting prevalence data in MapInfo was, however, was nontrivial, and required various data sources and conversion programs in order to obtain sufficiently detailed plots. Support for this work was provided by the Geodata Institute (within the School of Geography) at the University of Southampton. However, despite the initial efforts, once accomplished geomapping analysis permitted:

- A clear indication of the geographical prevalence and numbers of patients in the 16- to 24-year-age group across the opportunistic screening region.
- Clustering of areas of high numbers of positive test results and high levels of test prevalence.
- A presentation of the infection clusters with their relative magnitudes, allowing for future screening and intervention strategies to be focused in the important infection hotspots and to facilitate intervention planning.
- Analysis and geomapping of the confidence intervals and prevalence uncertainties, providing necessary checks and balances to ensure that any intervention policy is robust.

Numerous maps were produced for the study. Only a small sample is presented here to illustrate the use of geompaping. Spatial prevalence was plotted at both postcode district and sector levels. Figure 7 shows the geographical extent of the opportunistic trial based on the 11 140 patients tested in the Portsmouth area. The postcode sector polygons have been omitted for clarity, although the bars are located at the centres of each postcode sector. The numbers are also omitted but the numbers of tests carried out in each postcode sector is indicted by the relative height of each bar. The largest count (501) occurred in PO4 0 near the Havelock area of Portsmouth, with other large counts in the populated areas around Fareham, Gosport, Portsmouth and Havant.

Figure 8 plots test results at the district level. The left-hand bar (darkest) shows the total number of tests; the middle bar shows the number of positives; the right-hand bar (lightest) shows other results (equivocal, etc). Note that to show these results alongside each other a non-linear (logarithmic) scale was used.

Similar plots were obtained for prevalence, with and without equivocal results, by postcode district. Furthermore, the same maps were produced at the more detailed postcode sector level. For example, Figure 9 shows prevalence by sector. Here, prevalence is indicated by the disc diameter (larger the diameter, higher the prevalence). Actual values are not shown on the map for clarity, but relative visual comparison can be made here. The size of the square at the

Figure 7 Trial coverage by postcode sector.

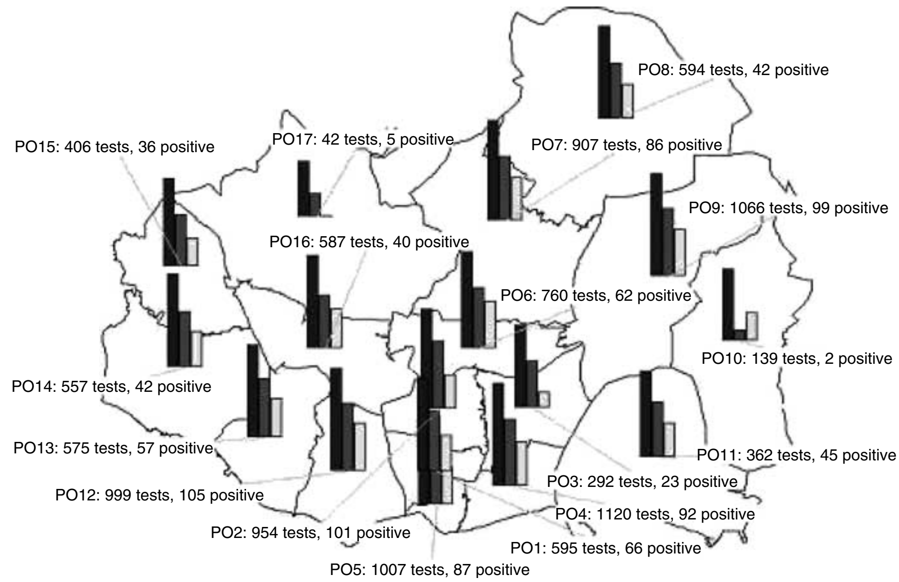

Figure 8 Total and positive test results by postcode district.

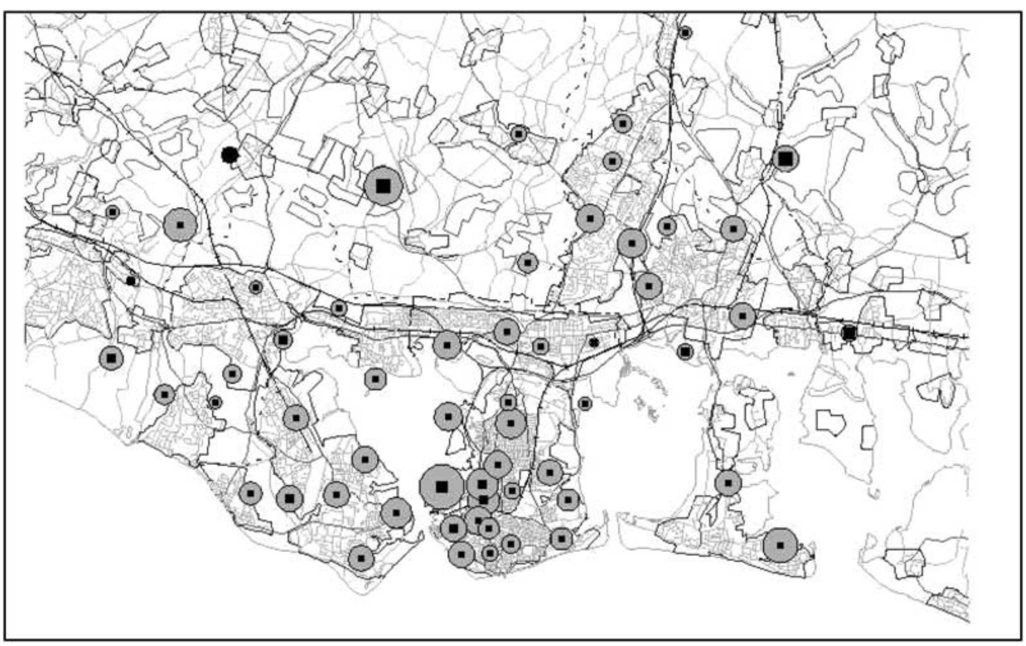

Figure 9 Prevalence by postcode sector.

centre of each disk indicates confidence interval (larger the square, larger the CI).

To target intervention strategies, we split the postcode sectors into groups to reflect the distribution of prevalence values, as shown in Figure 10. The horizontal axis shows increasing levels of prevalence. In conjunction with analysis of the actual values testing positive and equivocal, we decided to group the sectors into four categories of prevalence ranges. This process allowed the top nine sectors to be identified as a distinct group, followed by a second group of seven sectors. Both of these have above average prevalence levels. The remaining sectors were split into two larger groups with middling and low prevalence ranges. Table 3 shows these groupings.

The key tasks and findings of the geomapping work may be summarized as:

- The extent of the screening trial coverage has been plotted provided a clear indication of the numbers of patients in the 16- to 24-year-age group across the opportunistic screening trial region.

- Test result data and prevalence levels have been plotted. This shows the clustering of areas of high numbers of positive test results and high levels of test prevalence.
- Presentation of the infection clusters with their relative magnitudes allows future screening and intervention strategies to be focussed in the important infection 'hotspots'.
- By prioritising prevalence values, four categories of infection prevalence ranges have been identified. These have been plotted as maps to facilitate intervention planning.

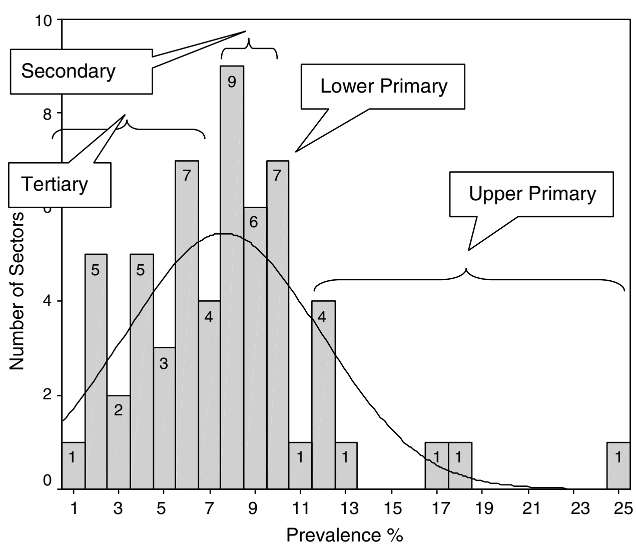

Figure 10 Distribution of prevalence.

4. Socio-economic indicators and prediction

Building on the geomapping work and identification of high prevalence sectors, this section describes analysis to ascertain whether of socio-economic variables could act as indicators of infection prevalence. Statistical analysis was carried out using SPSS (www.spss.com) and Minitab (www.minitab.com) and consisted of multivariate regression analysis, and tree-based regression analyses, CART and CHAID (Breiman *et al*, 1984). Multivariate regression analysis allowed for the primary determinants of prevalence to be identified. Since the collinear variables were removed in the regression process, only the important variables remained that best explained the variation in prevalence. CART and CHAID analyses were then used for clustering of prevalence groups, in order to identify higher risk predictors.

A number of types of socio-economic indicator variable were investigated. The primary source was the Office of National Statistics (ONS). ONS provided census data in a predefined format and content. Other census data were available from Manchester University's Casweb facility via MIMAS (www.mimas.ac.uk). Casweb provided a much higher degree of detail and flexibility, available in large and complex datasets that allowed user definable file format and content.

We used seven indices provided by the ONS to capture aspects of deprivation, plus an overall index. Various factors are used to calculate an index for each electoral ward. In addition, the wards are ordered according to its index to provide a rank position for that ward. Rank position 1 was the most deprived, and 8414 the least deprived. Correlation analysis was conducted on both index and rank. Deprivation indices are determined in the following domains:

- Income deprivation.
- Employment deprivation.

Table 3 Allocation of sectors to target groups

Category	*Target sector*		*Category*	*Target sector*	
Upper primary	PO9 6	PO12 1	Lower Primary	PO13 0	PO7 7
Top 9	PO1 3	PO1 4	Second 7	PO6 4	PO6 3
Prevalence	PO17 6	PO1 1	Prevalence	PO12 4	PO9 3
Average 17.3%	PO11 9	PO11 0	Average 13.51%	PO7 8	
SD 4.0%	PO15 6		SD 0.2%		
Secondary	PO2 0	PO13 9	Tertiary	PO7 5	PO15 7
Middle 20	PO5 4	PO4 8	Lower 23	PO10 8	PO9 1
Prevalence	PO9 2	PO3 6	Prevalence	PO4 9	PO6 2
Average 11.62%	PO2 8	PO9 5	Average 7.87%	PO8 8	PO5 2
SD 0.7%	PO12 3	PO13 8	SD 1.7%	PO9 4	PO16 7
	PO2 7	PO5 3		PO4 0	PO17 5
	PO12 2	PO16 0		PO7 6	PO14 2
	PO14 4	PO6 1		PO8 9	PO3 5
	PO16 9	PO14 1		PO1 5	PO8 0
	PO1 2	PO5 1		PO16 8	PO10 7
				PO2 9	PO15 5
				PO14 3	

- Health deprivation and disability.
- Education, skills and training deprivation.
- Housing deprivation.
- Geographical access to services deprivation.
- Child poverty (a subset of income deprivation).

Correlation with infection prevalence, as measured at the electoral ward level, gave some very significant correlations. There were potentially important differences of emphasis in the GUM data. This is shown in Table 4, where significant correlations at the 99% level are indicated by double asterisks, and at the 95% level are indicated by single asterisks. Lower P-values indicate a greater correlation than higher P-values. With one exception, all deprivation indices showed that worse deprivation was associated with higher prevalence. The exception was Geographical Access, where worse deprivation was associated with reduced prevalence, and this was the only socio-economic indicator in both opportunistic screen and GUM datasets to be correlated in this way. Positive GUM patients appeared to be from less deprived wards, since Access and Education deprivation were the only variables to be strongly correlated with GUM patient's prevalence levels.

Other Casweb indictors were also tested for correlation with prevalence. These included deprivation, age, ethnic origin, car ownership, social class, income support and jobseekers allowance, and vital statistics (birth and death rates). All indicators, except vital statistics, were found to be statistically significantly correlated (at the 95% level) with prevalence. Results were as might be expected, for example higher prevalence was associated with higher income support levels, lower car ownership and lower-skilled professions.

Multiple regression was carried out to reduce the number of previously observed correlated variables to a minimum and useful (practical) subset that best-described infection prevalence. It was found that most of the variables were collinear, that is, they generally exhibited similar behaviour, and thus duplicated the true underlying mechanisms. Regression analysis was used to identify an independent and non-collinear set of variables. Kolmogorov–Smirnov and Shapiro–Wilk's tests were used to make informed judgment on necessary transformations, which included squaring prevalence to satisfy normality assumptions.

Selecting a range of candidate variables based on the correlation analysis gave a useful regression equation to describe or predict prevalence. The key variables found using Minitab's best subsets function were education deprivation rank, ratio of 20–24 years old to ward population, child poverty rank, and number in the ward of age 16–17 years old. The R^2 value was 0.650 which indicated that 65% of the variation was described by these variables. The ANOVA result showed that the regression was statistically significant, since P-value <0.05. The coefficient table showed that all the variables were significant, with the possible exception of the 16–17 years age group. This could have been excluded with a small reduction in R^2 value to 0.621. The regression equation has been plotted in Figure 11 below, and seems to explain the prevalence variation reasonably well.

CART and CHAID analysis was undertaken using AnswerTree in SPSS. These methods utilize splitting rules and variance reduction techniques. A trade-off is made between misclassification cost and tree complexity to prune the tree structure to its optimal and simplest structure. Only the CART tree is presented here (Figure 12) as CHAID gave almost identical results. Reassuringly, CART and CHAID suggested the same key variables as identified in the regression analysis. CART analysis found education deprivation as the primary determinant, whereby low deprivation rank (ie worse deprivation) provides a group of higher mean prevalence. In the left-hand branch, this was in turn split by the variable describing the population ratio in the 20–24 years age group. Wards with a proportion of 20–24 years old greater than 8.63% of the total population had the highest prevalence. In the right-hand branch, wards where the

Table 4 Indices of deprivation at ward level

Deprivation index	*Significance* Trial data P-value	GU data P-value
Multiple rank position	0.000**	0.141
	0.000**	0.023*
Income rank position	0.001**	0.136
	0.000**	0.028*
Employment rank position	0.002**	0.189
	0.001**	0.038*
Health rank position	0.001**	0.017*
	0.001**	0.026*
Education rank position	0.000**	0.002**
	0.000**	0.000**
Housing rank position	0.004**	0.057
	0.001**	0.042*
Access to services rank position	0.000**	0.005**
	0.000**	0.001**
Child poverty rank position	0.001**	0.263
	0.001**	0.111

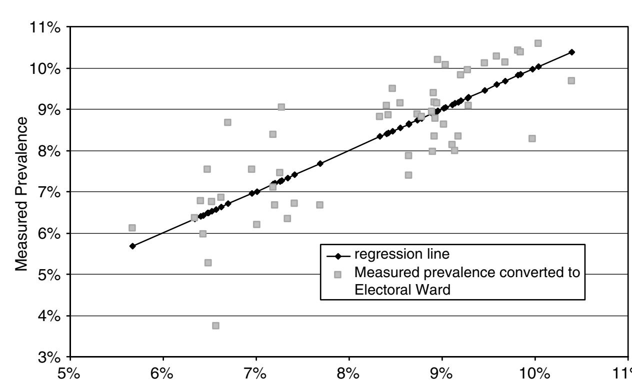

Figure 11 Opportunistic screen prevalence regression plot.

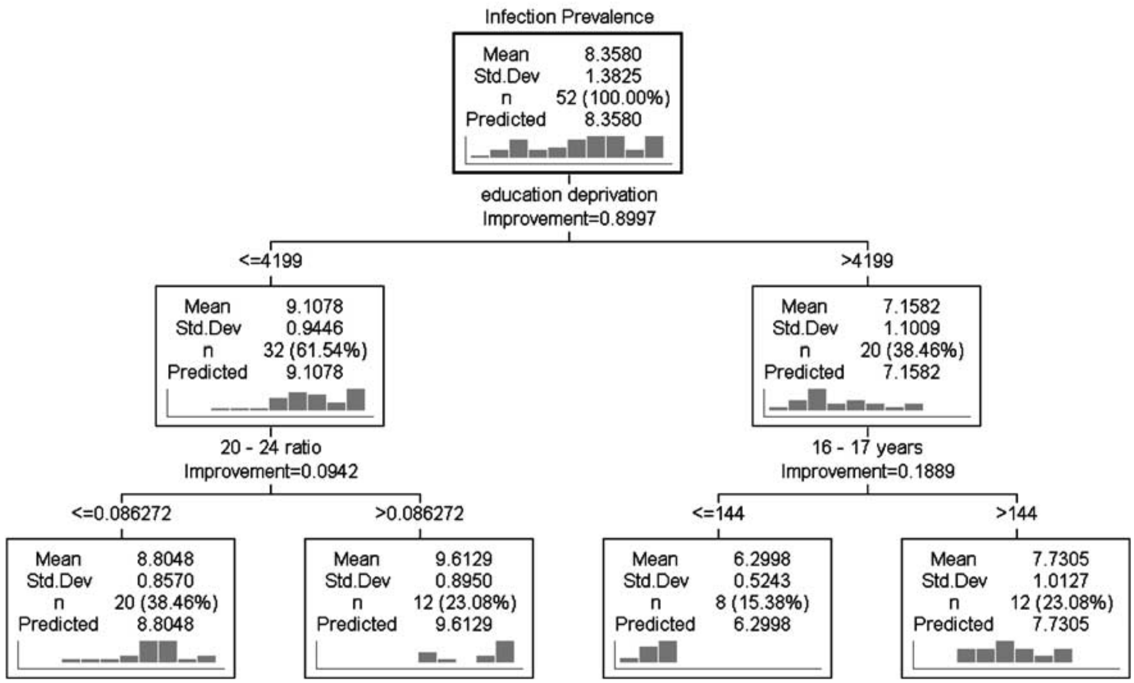

Figure 12 Opportunistic screen prevalence: results from CART.

number of 16–17 years old was less than 144 have the lowest overall prevalence. Education deprivation became a continual theme in exploratory analyses, with some binary trees branching more than once using this variable. This was perhaps explained by CHAID, which allows multiple splits from each node, and showed opportunistic screen prevalence split into three clear groups, based primarily on education deprivation.

In this section, three main techniques have been employed to gain a multiple-perspective insight into the relationships between socio-economic indicators and infection prevalence. Key findings are summarized as:

- Indices of deprivation were found to be very important, with significant correlations with both screening trial data and GUM data.
- A worse level of education deprivation was associated with higher prevalence. The converse was true for access to services deprivation.
- Other variables were investigated with a view to finding a better set of indicators or surrogates for deprivation, but the regression analysis confirmed that infection prevalence at the ward level was best explained by deprivation indices and age.
- CART analysis provided a simple sorting rule set, easier to interpret than a regression equation, and provided a graphical explanation for the non-statistician. This confirmed the importance of education deprivation. Education deprivation was also identified as the primary determinant of prevalence using CHAID.

5. Simulation modelling of infection dynamics and costs-effectiveness

To model the dynamics of infection recovery and sequelae, and in order to quantify cost-effectiveness of various screening strategies, a SD model was developed using the Vensim software (www.vensim.com). SD is ideally suited to modelling infections and large population movements. It was particularly relevant in modelling Chlamydia, in that the repeat re-infection mechanism was captured, along with the increased risk of sequelae given repeat infection. These time-dependent effects cannot easily be captured decision analysis models.

Adjustable parameters were included for screening and treatment options, and the model allowed interactions to be assessed and cost-effectiveness to be estimated. More detailed information on the model structure, parameters and results have already been reported in Evenden *et al* (2005). In this paper, we present limited results in order to demonstrate how previous components of the research (geomapping and risk grouping) are combined with the SD model.

With the agreement of Consultants at St Mary's Hospital, we developed a simplified, efficient, and user-friendly model with the practical needs of policy makers in mind. Evenden *et al* (2005) describes why an SD approach was adopted together with the novel aspects of our model compared to the Townshend and Turner (2000) SD model, and other models in the literature, and is not repeated here. The basis of the chosen model is presented in Figure 13 and shows the causal loop diagram (CLD) in conjunction with the corresponding simplified stock and flow diagram (SFD).

Two risk groupings were used to capture a higher and lower level of sexual activity and infection. Parameters, such as frequency of partners and size of population, were informed by the geomapping and statistical analyses. Experimentation with the model showed that one of the most important factors was infection from the high-risk infected group into the low-risk susceptible group. The fraction of Chlamydia infections resulting in sequelae was set to 20%. This value was based on a range of sources investigated (Genç and Mardh, 1996; Magid *et al*, 1996; Howell *et al*, 1998; Townshend and Turner, 2000; van Valkengoed *et al*, 2001; Gift *et al*, 2002; Skaza and Eržen, 2002; Yeh *et al*, 2003), whence a number of other key parameters including treatment and screening costs were extracted and confirmed by expert opinion from Consultants in GUM at St Mary's Hospital. A probability tree was built to represent the various sequelae possibilities, including pelvic inflammatory disease (PID), infertility, ectopic pregnancy and chronic pelvic pain. Table 5 shows a selection of the base-case parameters.

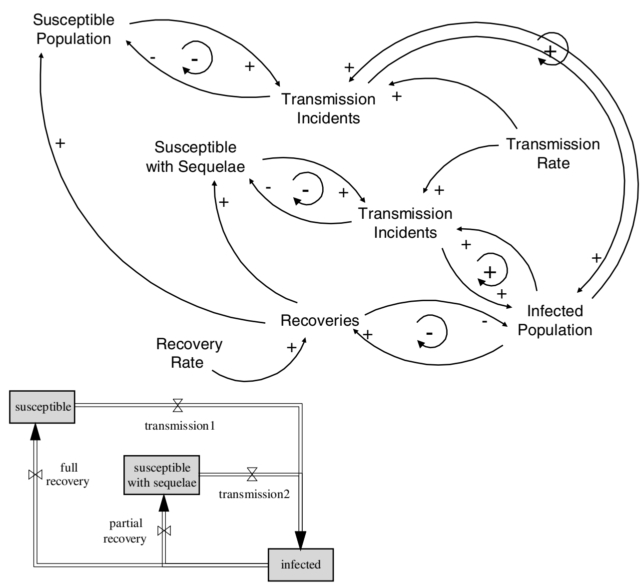

Figure 13 SD model structure.

In addition to these base-case parameters, infection prevalence was modelled at three levels of infection prevalence (5, 8 and 10%). In order to validate the model, parameters were set to those from other published studies in the literature, such as Howell *et al* (1998), and the results compared to those published. Statistical tests indicated that results from our model, for simple base-line scenarios, were not significantly different. Various sensitivity analyses were also performed, for example by adjusting the values of screening rates for both low- and high-risk groups. Sensitivity analysis showed that rate and cost of sequelae, low-risk partnership rate, and mixing rate were the dominant variables. This was as anticipated, suggesting that the model was not exhibiting any unexpected behaviour.

To illustrate cost-effectiveness results, we present Figure 14, which shows overall costs as a function of the

Table 5 Base-case modelling parameters

Parameter	*Value*
Population	10 000
Percentage of population initially in high-risk group	2.5%
Percentage of population initially in low-risk group	97.5%
Risk group mixing probability (within-group)	90%
Risk group mixing probability (between-group)	10%
High- and low-risk group sequelae fraction	20%
Low-risk group partnership frequency	1 every 5 years
High-risk group partnership frequency	1 every 2 months
High- to low-risk group mixing ratio	10%
Infection mean recovery time	30 months
Modelling period	24 months
Screening cost	£10
Treatment cost	£10
Sequelae cost	£2000

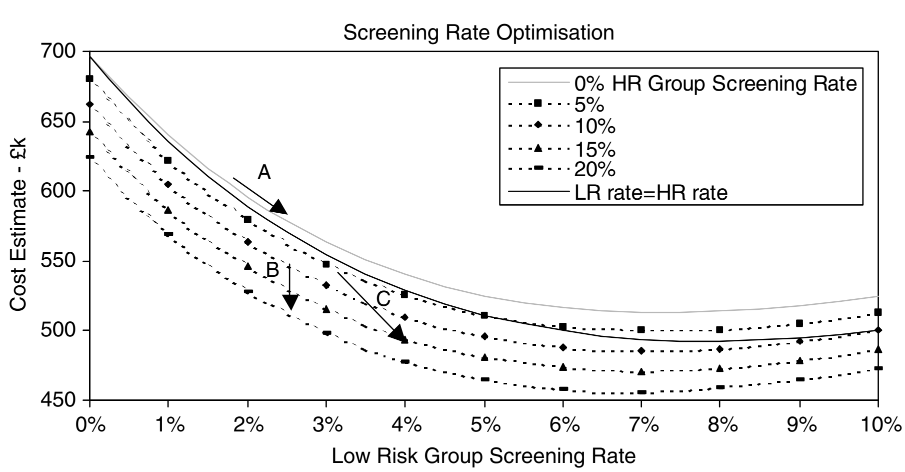

Figure 14 Cost savings for infection prevalence at 10%.

screening rate of the low-risk group. This is shown for six cases of screening of the high-risk group indicated in the legend. An idealized intervention is described using the labelled arrows. Various other scenarios were also simulated and are discussed in Evenden *et al* (2005).

The intersection of the topmost curve with the vertical axis shows the cost of not screening. Arrow A along the solid line shows the direction and overall cost reduction by increasing screening within the general population (ie both high- and low-risk groups). Arrow B shows the benefits of further reducing costs by increased targeted screening of the high-risk group, where cost savings can be accrued more rapidly.

Clearly this is an ideal situation since it would not necessarily be possible to target the high-risk group so precisely without also screening more from the low-risk group. For this reason most practical interventions will probably consist of a mix with a predominance of high-risk persons, as indicated by arrow C. Here we make reference to the geomapping work and statistical analysis, which will inform staff at St Mary's Hospital, Portsmouth, on the location of these high-risk groups and their likely socio-economic characteristics. This combined geomapping and modelling approach is a key benefit of this research over other published studies.

The key findings of the SD work were:

- SD modelling can provide key insights to allow the infection dynamics to be better understood. Key parameters can be varied and the effect on the dependent variables can be observed.
- Since prior infection does not convey immunity, re-infection is one of the main characteristics, which makes its prevalence such a long-term problem. It is a behavioural as well as a medical problem.
- The role of the high-risk groups in the infection dynamic is very important, as it provides a key source of new infection into the low-risk groups. Within the high-risk group itself the prevalence stabilises at a high level, with the majority infected.
- Screening provides immediate cost benefits—costs may be reduced by up to a half at high levels of infection prevalence. At reasonably achievable level, of screening, say 1–2% of the overall population, cost savings are worthwhile.
- To achieve optimal cost savings, a larger proportion of the high-risk groups need to be screened. It will be easier to screen a large proportion of the smaller high-risk groups than it will be to screen a smaller proportion of the larger low risk groups.
- For every high-risk person screened per month, around £1,500 can be saved.
- For every low-risk person screened per month, around £200 can be saved.
- These high-risk groups have been identified in geographic location terms and socio-economic terms (Sections 3 and 4) to enable this critical strategy to be planned, implemented and to succeed.

6. Conclusions and policy implications

In this study, we have conducted a detailed analysis of 17 553 screening trial results from the UK Department of Health Chlamydia opportunistic trial in Portsmouth. We developed a system dynamics model, with parameter values informed by the analysis of the trial data, which has shown that a high-risk sub-group of the general population, despite being relatively small in size but with a high number of sexual partnerships per case, is critical in the infection dynamics of Chlamydia. Such a group has the largest proportion of sequelae, and provides a major source of infection into the low-risk population sub-group. While the benefits of a universal screening approach have been calculated, it is clear that greater benefits accrue from a higher level of screening of the high-risk group. Thus, blanket screening of the entire at risk population, as proposed in the Department of Health's national screening programme, might be seen simply as to add extra burden to the overstretched health economy, whereas improved targeting of high-risk populations has been shown here to achieve greater cost-effectiveness and to control the recent alarming rise in cases of Chlamydia.

A modelling framework combining computer simulation, geomapping and risk-group clustering techniques, has facilitated a holistic view of the problem. Thus, it has been possible to find the indicators that determine high-risk and high-prevalence within the Portsmouth population, as well as geographically displaying their location across the region by postcode district and sector. Age and indices of deprivation were found to be good predictors, especially education deprivation and proportion of young people within the resident postcode sector. CART, CHAID and multiple regression approaches all gave consistently similar results.

Staff in the GUM department, St Mary's Hospital, are now able to evaluate their screening intervention planning and to re-organize their services in order to target the high-risk groups. For example, they now plan to utilize resources more effectively by visiting schools and public places within the identified primary target postcode sectors in order to increase awareness of Chamydia infection. It will be possible, using the same modelling framework, to analyse other trial data when this comes available from other geographical regions of the Department of Health screening programme (NCSP). This will enable us to see if the same indicators of high-risk behaviour in Portsmouth are consistent with other regions. Clearly this level of information, coupled with geomapping, will assist regional planning of the screening programme. Furthermore, although we present how the methodology has been used to consider the spatial prevalence and screening interventions for Chlamydia, this approach could be readily applied to other sexually transmitted infections or infectious diseases, such as HIV/AIDS surveillance and intervention planning.

References

Addiss DG *et al* (1993). Decreased prevalence of *Chlamydia trachomatis* infection associated with a selective screening programme in family planning clinics in Wisconsin. *Sex Transm Dis* **20**: 28–34.

Breiman L *et al* (1984). *Classification and Regression Trees*. Chapman & Hall: London.

Buhaug H, Skjeldestad FE, Backe B and Dalen A (1989). Cost effectiveness of testing for Chlamydial infections in asymptomatic women. *Med Care* **27**: 833–841.

Department of Health (2000). *Summary Report: A Pilot Study of Opportunistic Screening for Genital Chlamydia trachomatis Infection in England*. The Sexual Health and Substance Misuse Team, Department of Health: London.

Department of Health (2005). *National Chlamydia Screening Programme (NCSP)*. www.dh.gov.uk.

Evenden D, Harper PR, Brailsford SC and Harindra V (2005). System dynamics modelling of Chlamydia infection for screening intervention planning and cost benefit estimation. *IMA J Man Math* **16**: 265–279.

Genç M and Mardh PA (1996). A Cost-effectiveness analysis of screening and treatment for *Chlamydia trachomatis* infection in asymptomatic women. *Ann Intern Med* **124**: 1–7.

Gove DJ (1997). *Simulation modelling of infectious diseases*. PhD theses, University of Southampton, Southampton, UK.

Gift TL, Walsh C, Haddix A and Irwin KL (2002). A cost-effectiveness evaluation of testing and treatment of *Chlamydia trachomatis* infection among asymptomatic women infected with Neisseria gonorrhoeae. *Sex Trans Dis* **29**: 542–551.

The Guardian (2004). *Just say no to avoid sex infections, young told*. 25th November, 2004. http://www.guardian.co.uk/uk_news/story/0,,1358905,00.html.

Haddix AC, Hillis SD and Kassler WJ (1995). The cost effectiveness of azithromycin for *Chlamydia trachomatis* infections in women. *Sex Trans Dis* **22**: 274.

Harindra V, Tobin JM and Underhill G (2002). Opportunistic chlamydia screening: should positive patients be screened for co-infections? *Int J STD AIDS* **13**: 821–825.

Hart GJ, Duncan B and Fenton KA (2002). Chlamydia screening and sexual health, Editorial. *Sex Transm Infect* **78**: 396–397.

Health Protection Agency (2004). *HIV and other Sexually Transmitted Infections in the United Kingdom in 2003. Annual Report, November 2004*. http://www.hpa.org.uk/infections/topics_az/hiv_and_sti/sti-chlamydia/chlamydia.htm.

Herrmann B, Johnsson A and Mardh PA (1991). A retrospective study of efforts to diagnose infections by *Chlamydia trachomatis* in a Swedish county. *Sex Transm Dis* **18**: 233–237.

Hicks NR *et al* (1999). Chlamydia infection in general practice. *BMJ* **318**: 790–792.

Honey E *et al* (2002). Cost effectiveness of screening for *Chlamydia trachomatis*: a review of published studies. *Sex Transm Infect* **78**: 406–412.

Howell MR, Quinn TC and Gaydos CA (1998). Screening for *Chlamydia trachomatis* in asymptomatic women attending family planning clinics. *Ann Intern Med* **128**: 277–284.

Magid D, Douglas JM and Schwartz JS (1996). Doxycycline compared with azithromycin for treating women with genital *Chlamydia trachomatis* infections: an incremental cost-effectiveness analysis. *Ann Intern Med* **124**: 389–399.

Pimenta JM *et al* (2003a). Opportunistic screening for genital *Chlamydial infection*. I: Acceptability of urine testing in primary and secondary healthcare settings. *Sex Transm Dis* **79**: 16–21.

Pimenta JM *et al* (2003b). Opportunistic screening for genital Chlamydial infection. II: Prevalence among healthcare attenders, outcome, and evaluation of positive cases. *Sex Transm Dis* **79**: 22–27.

Skaza A and Eržen I (2002). Cost-effectiveness of screening for *Chlamydia trachomatis* in adolescent females in Slovenia. *Dermatovenerologica* **11**. www.mf.uni-lj.si/acta-apa/

Townshend JRP and Turner HS (2000). Analysing the effectiveness of chlamydia screening. *J Oper Res Soc* **51**: 812–824.

van Valkengoed IGM *et al* (2001). Cost effectiveness analysis of a population based screening program for asymptomatic *Chlamydia trachomatis* infections in women by means of home obtained urine specimens. *Sex Transm Inf* **77**: 276–282.

Yeh JM, Hook EW and Goldie SJ (2003). A refined estimate of the average lifetime cost of pelvic inflammatory disease. *Sex Transm Dis* **30**: 369–378.

Received May 2005;
accepted October 2005