Variations in the epidemiology of primary, secondary and early latent syphilis, England and Wales: 1999 to 2008

Heather Jebbari, Ian Simms, Stefano Conti, Andre Marongiu, Gwenda Hughes, Helen Ward, Cassandra Powers, Daniel Rh Thomas, Barry Evans

Variations in the epidemiology of primary, secondary and early latent syphilis, England and Wales: 1999 to 2008

Heather Jebbari[1], Ian Simms[1], Stefano Conti[2], Andrea Marongiu[1], Gwenda Hughes[1], Helen Ward[1], Cassandra Powers[1], Daniel Rh Thomas[3], Barry Evans[1].

[1] MESH Dept., Health Protection Agency Centre for Infections, 61 Colindale Ave, London, NW9 5EQ

[2] Statistics, Modelling & Bioinformatics Unit, Health Protection Agency Centre for Infections, 61 Colindale Ave, London, NW9 5EQ

[3] NPHS Communicable Disease Surveillance Centre, Temple of Peace and Health, Cardiff CF10 3NW

* Corresponding author: **Dr Ian Simms,** (t) 020 8327 7571, (f) 020 8200 7868

ian.simms@hpa.org.uk

Key messages

- Over the last ten years the infectious syphilis epidemic has remained 'young', with a high proportion of primary and secondary cases.
- The increase in primary syphilis as a proportion of early syphilis indicates that infection is detected and managed at an earlier stage of infection.
- Sustained, intensive, targeted efforts to interrupt further transmission need to be maintained and intensified.
- Locally-based interventions that penetrate sexual networks identified through partner notification and surveillance initiatives will likely be the most effective method of controlling infection.

Key words:

Infectious syphilis; epidemiology; stage of infection; England & Wales

Word Count:

Abstract - 244

Main body of text – 2588

Abstract

Objectives

To investigate factors associated with variations in diagnoses of primary, secondary and early latent syphilis in England and Wales.

Methods

Data were derived from two sources: diagnoses made in genitourinary medicine (GUM) clinics reported on form KC60, and information collected through National Enhanced Syphilis Surveillance (NESS). Multinomial regression modelling was used for data analysis.

Results

Between 1999 and 2008, 12 021 NESS reports were received, 54% of KC60 reports. The dominant profile of the epidemic was one of White men who have sex with men aged 35 to 44, often co-infected with HIV, centred in larger cities. During this period, the proportion of primary cases increased over time, while the proportion of secondary cases fell. Primary cases exceeded secondary cases by 2004. The proportion of early latent cases remained relatively stable over time and tended to be lower than that of primary and secondary infection. Patients who attended because they had symptoms of infection, had been identified through partner notification, were HIV positive and were UK-born were more likely to present with primary or secondary infection than early latent infection. A higher proportion of early latent cases were seen among patients who were Asian, had contacted sexual partners through saunas, bars and the internet, had untraceable partners and had acquired infection in Manchester.

Conclusions

The continuing syphilis epidemic indicates that control has only been partially effective, with ongoing transmission being sustained. Intensive and targeted efforts delivered locally are required to interrupt further transmission.

Introduction

In England and Wales the resurgence of infectious syphilis started in 1997 with an outbreak amongst heterosexual men and women in Bristol. Since then diagnoses made at genitourinary medicine (GUM) clinics have risen from 285 in 1997 to 3265 cases in 2008.[1,2] This dramatic change in incidence has been focused on major urban areas, particularly London and Manchester, but outbreaks have been seen throughout England and Wales, and there are striking similarities within outbreaks seen in Europe, North America and Australia.[2] The epidemic has been influenced by developments in the HIV epidemic and behavioural change in men who have sex with men (MSM) and has been characterised by a rapid increase in diagnoses made in MSM, and a high proportion of HIV co-infections. An outbreak amongst heterosexuals, including young people, has developed alongside the larger MSM epidemic.

The natural history of syphilis suggests that an early epidemic dominated by primary and secondary infection would be expected to develop into an epidemic predominantly composed of early latent infections.[3,4] Each phase presents a different public health challenge. Here we investigate the factors that are associated with variations in the stage of diagnosed infections over the past ten years and consider the implications of these findings in terms of public health intervention.

Methods

Data sources

Two data sources were used: diagnoses made in GUM clinics reported on form KC60, and information collected through National Enhanced Syphilis Surveillance (NESS) which was also undertaken only within GUM clinics. Both methodologies have been previously described and discussed in detail.[5] In brief the KC60 was an aggregate mandatory return,

the number of cases is recorded by gender and male sexual orientation for primary, secondary and early latent syphilis[a]. Age group was also recorded but only for primary and secondary infection. NESS is a voluntary reporting system based on the collection of disaggregate (patient level) data including: gender, age, ethnic background, sexual orientation, stage of infection, HIV status, geographic area where the infection was likely to have been acquired and connections with social and sexual networks.

Definitions

Definitions of early latent syphilis vary between countries. The definitions used for primary, secondary and early latent infection within the UK are given in the guidelines published by the British Association for Sexual Health and HIV (BASHH www.bashh.org). The BASHH guidelines define early latent syphilis as "*Treponema pallidum* infection diagnosed on serological testing with no symptoms or signs within the first two years of infection: this is early latent syphilis and beyond that late latent syphilis". In the KC60 return, primary and secondary infections were recorded under a single combined category (A1, A2) and early latent (A3) was recorded separately. In contrast, under the NESS collection system primary (A1), secondary (A2) and early latent (A3) are recorded separately.

Data completeness & representativeness

The NESS dataset is a subset of the KC60 dataset. KC60 data are shown as a reference point for the NESS data, although only the patient level NESS dataset was used in the subsequent detailed analysis. Completeness of reporting varies over time and by Strategic Health Authority (SHA): reporting from the North West and North East SHAs was close to 100%. NESS began in 1999 in the North West, while Wales and most other English regions followed from 2003, the East of England and East Midlands regions submitting data for the first time in 2007. Coding accuracy varies between the two datasets.[6] For

example the patient management software used in some clinics automatically defaults male sexuality to 'heterosexual', leading to the incorrect assignment of some patients to this category. Regional variation in diagnoses reported through NESS was the area of most concern in terms of selection bias. To investigate this in more detail SHA of diagnosing GUM clinic was included in the statistical model.

Statistical analysis

Statistical analysis was only undertaken on NESS data. The relationship between stage of infection and selected explanatory variables was investigated by using multinomial (also known as multi-categorical logistic) regression modelling (details given in Annex).[7] The probability of presenting with syphilis at each stage of infection was estimated by year and for different patient characteristics from the model using maximum likelihood. Predicted probabilities were plotted to investigate infection dynamics over time. These figures are shown in the main text and Annex.

Results

Between 1999 and 2008, annual cases of infectious syphilis reported through the KC60 system rose from 415 to 3265, an increase in incidence from 0.79 per 100,000 (denominator = total population) 5.99/100,000. A cumulative total of 22 156 diagnoses were made over the period. For the same period, NESS reports were received for 12 021 cases, 54% of the KC60 total. The NESS and KC60 datasets were compared in terms of the outcome variable, stage of infection, and an overview of the proportions recorded in each dataset indicate that there was a good general level of qualitative agreement (Figure 1). This suggests that overall NESS gave a consistent representation of the distribution of diagnoses over time. The analysis was undertaken on those patients for whom complete information were available. After exclusion of cases with missing values 11 838 were used for the statistical analysis (Table 1). The majority (73%, 8656/11 838) of

cases were seen in MSM compared to 16% (1938/11 838) in heterosexual men and 10% in heterosexual women (1244/11 838). Most cases were seen in patients of White ethnic background (9491/11 594 or 82%), whereas Black and Asian groups accounted for 10% (1126/11 594) and 6% (656/11 594) respectively. Primary syphilis was reported in 43% (4698/10 997) of cases, secondary in 33% (3677/10 997) and early latent in 24% (2622/10 997). Of the 233 patients identified through antenatal screening 97% were women and 141 (61%) were diagnosed with early latent syphilis. MSM reported a median of 2 (inter-quartile range 1 to 5) sexual partners in the 3 months prior to diagnosis, whereas the median number reported by heterosexual men and women was 1 (IQR 1 to 2). Just over a third of syphilis patients (3087/7870 or 39%) thought that they had been infected through oral sex, 92% of whom (2847/3087) were MSM.

Results of the multinomial regression model are summarised in Tables 2 and 3. Each estimated regression coefficient provides a measure of increased (if positive) or decreased (if negative) likelihood of presenting with either primary or secondary syphilis, relative to early latent. For example, an HIV positive individual would be expected to have a 1.62=exp{0.48} higher odds ratio of presenting primary syphilis infection (as opposed to early latent) compared to an HIV negative patient.

Patients who attended because they had symptoms of infection were significantly more likely to present with primary infection relative to early latent infection ($p<0.001$) (Table 2). Patients brought in through partner notification ($p=0.051$), who were HIV positive ($p=0.052$) and who were UK-born ($p=0.064$) were also more likely to present with primary infection, although this was of borderline significance. In contrast, patients who were Asian ($p=0.039$), or those who had contacted sexual partners through saunas ($p=0.001$), bars ($p<0.01$) and the internet ($p=0.01$), had untraceable partners ($p=0.029$) and had acquired

infection in Manchester (p=0.043) were significantly less likely to present with primary infection relative to early latent infection.

The pattern was similar when secondary and early latent infection were compared (Table 3). Patients who attended because they had symptoms of infection (p<0.001), had been identified through partner notification (p=0.022), were HIV positive (p<0.001) and were UK-born (p=0.045) were significantly more likely to present with secondary infection relative to early latent infection (Table 3). Patients who had contacted sexual partners through saunas (p=0.005), bars (p=0.002) and the internet (p=0.005) were significantly less likely to present with secondary infection relative to early latent infection.

There was no significant association between stage of infection at presentation and the health region reporting the data, suggesting that inconsistent regional reporting is unlikely to have influenced the main findings (Tables 2 and 3).

In general, the proportion of diagnosed cases of primary infection increased over time, whereas the proportion of secondary cases fell (Figure 2 and Annex Figures 1 to 5). For all the variables studied, the proportion of primary cases had exceeded that of secondary cases by 2004. The proportion of early latent cases remained relatively stable over time and was lower than that of primary and secondary infection. MSM were more likely to present with primary infection whereas heterosexual men were more likely to present with secondary infection (Figure 2). Women were most likely to present with primary infection but a relatively high and increasing proportion presented with early latent infection (Figure 2). The relative proportion of the stages of infection varied substantially depending on the reasons for attending clinical care, a higher proportion of early latent cases being seen in those attending for routine clinical care (Annex Figure 4). The proportion of

primary cases seen in patients who acquired their infection in Brighton exceeded that of secondary cases at least three years before this was observed for those infected in London, Manchester and other parts of the UK (Annex Figure 5).

Discussion

Over the last ten years the epidemic of infectious syphilis has remained 'young', with a high proportion of primary and secondary cases. The decline in diagnoses of secondary infection and the increase in diagnoses of primary infection indicate that infection is being detected and managed at an earlier stage of infection through increased disease awareness, better case ascertainment and increased access to GUM services. However, in the early 2000s waiting times to access GUM services increased substantially, caused by large increases in attendance at already overstretched clinical services and this is likely to have contributed to an increase in the duration of infectiousness.[2] The problem was recognized by the Department of Health (England) and the proportion of people contacting GUM services seen within 48 hours now approaches 100%.

The consistent level of early latent cases, although low, indicates that primary and secondary infections are going undiagnosed but it is not possible to predict the number of undetected cases present in the population. Undiagnosed cases have a number of public health implications. Some infections will be resolved through indirect antimicrobial treatment but of those that progress to tertiary syphilis, around a third are likely to develop clinical manifestations of late syphilis which may be seen in clinical settings such as neurology and cardiology over future decades. The high proportion of syphilis diagnoses in MSM that are co-infected with HIV reflects the close relationship between the epidemics. Syphilis infection is known to facilitate HIV transmission, and consequently may be contributing to increased HIV incidence. This highlights recommendations that MSM

should have an annual sexual health screen, including testing for HIV where not already diagnosed, supported by improved laboratory turnaround times for the diagnosis of syphilis.[8,9] MSM also need to be aware that syphilis can be transmitted through oral sex. There is also an increased risk of congenital syphilis in reproductive age women. Congenital syphilis can be prevented through first trimester screening supported by treatment and partner notification. This control method is cost-effective but highly dependent on well structured healthcare pathways. In 2007 95% of pregnant women were screened for syphilis in England, although uptake varied from 82% to 98% between regions.[10] This level of screening is slightly higher than that for HIV screening (93% nationally, range: 85% to 97%) despite there being no national target for syphilis screening in pregnancy. Over the period studied, this important public health intervention identified 22% of infections seen amongst women. The cases of congenital syphilis that have emerged reflect failures of both the antenatal screening and adult syphilis intervention strategies.[11,12]

The dominant profile of the epidemic is one of White, MSM aged 25 to 34, many of whom are co-infected with HIV and over a third believed that they had acquired infection through oral sex. This risk profile has been observed consistently since the re-emergence of infectious syphilis over a decade ago. [1] The associations between primary and secondary infection and social networks within saunas, bars and the internet reflects risk taking amongst MSM accessing both traditional 'sexual marketplaces' and internet chat rooms. The easy acquisition of sexual partners via the internet has joined previously isolated networks, and reduced the time taken for epidemics to evolve.

Infections amongst heterosexuals have also increased with the East Midlands region seeing a high proportion (54%) of heterosexual cases. The profile of these cases is more

diverse than those seen in MSM. In the early stages of the epidemic, heterosexual outbreaks were generally isolated and linked, for example, to travel abroad.[1] More recently heterosexual infections have been acquired through local sexual networks within the UK, sex work, and in some cases amongst students and young people aged less than 20.[13,14] In England and Wales, GUM clinics offer free, open access services that are widely advertised but some of those at risk of infection find accessing services difficult. For example, outbreak investigations showed that some young people diagnosed with syphilis were marginalised in society in terms of socio-economic circumstance.[15] They were not registered with health services and did not attend services when they experienced issues with their health. Increased risk of infection amongst patients of Asian ethnicity may reflect infection acquired abroad, but it may also be caused by barriers to accessing sexual health services that deter patients from seeking clinical advice.[16] These observations highlight the importance that all young people should be able to access comprehensive sexual health services for testing, treatment and management.

The main potential limitation of the study is the use of voluntary NESS dataset. The NESS dataset includes around 60% of the diagnoses recorded in the KC60 dataset and consequently it may not be representative of the overall pattern of diagnoses. High regional variation is a feature of the epidemic, but analysis of the NESS dataset showed that the presence of region as a factor in the analysis did not have a significant effect on the model, indicating that the enhanced data-set was not geographically biased.

Over the past decade diagnoses of infectious syphilis have increased in Western Europe and the US but few countries collect detailed information by stage of infection. In the US in 2008 primary and secondary infection accounted for 52% of diagnoses of early syphilis whereas early latent accounted for 48%.[17] In contrast, 24% of diagnoses of early syphilis made in England and Wales were early latent. However, comparisons are difficult to

make given variations in the quality of surveillance data, sexual behaviour, access to clinical services and treatment, and the different case definitions used for early latent syphilis. For example, the definition used in the US includes latent cases seen within the previous 12 months, a shorter period than the 24 months used in the UK definition.

The re-establishment of syphilis as an endemic infection reflects a failure of control strategies. Partner notification, which is central to the detection of the infection, has been of limited use to control efforts due to the high proportion of anonymous partners seen in MSM. Other forms of control include a combination of interventions such as the promotion of early diagnosis and treatment through measures, such as facilitating access to sexual health services, antenatal screening, diagnosis and treatment and promoting behavioural changes such as increased condom use and reducing the number of sexual partners. As the syphilis epidemic continues to develop, sustained, intensive and targeted efforts to interrupt further transmission need to be maintained and intensified. Locally-based interventions that penetrate sexual networks identified through local partner notification and surveillance initiatives will likely be the most effective method of controlling infection.

Word count = 2588

Acknowledgements

We would like to thank GUM clinicians for reporting cases of infectious syphilis through the enhanced surveillance systems operated by the HPA Regional Epidemiology Units and HPA Centre for Infections, and Dr Marion Lyons for providing surveillance data for Wales.

Competing interests

None

Copyright licence

Footnote

a The KC60 statistical return was replaced by the Genitourinary Medicine Clinic Activity Dataset as at 1st April 2009.

References

1 Health Protection Agency. STI Table 1 Country 2008. Health Protection Agency,2009.http://www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1215589015024.

2 Simms I, Fenton KA, Ashton M, *et al.*. The re-emergence of Syphilis in the United Kingdom: the new epidemic phases. *Sex Transm Dis* 2005:**32**:220-6.

3 Grassly NC, Fraser C, Garnett GP. Host immunity and synchronised epidemics of syphilis across the United States. *Nature* 2005;**433**:417-421.

4 Simms I. Syphilis. In: *International Encyclopedia of Public Health*. Ed. HK Heggenhougen. Elsevier Publications. 2008. ISBN 978-0-12-373960-5

5 The UK Collaborative Group for HIV and STI surveillance. *Testing times.* HIV and Sexually Transmitted Infections in the United Kngdom: 2007. London: Health Protection Agency, Centre for Infections. November 2007.

6 Menon-Johansson AS, Young Y, Maguire H, *et al.*. Surveillance of an infectious syphilis outbreak in south London, UK: a comparison of reporting systems. *Int J STD AIDS.* 2009;**20**:196-9.

7 Fahrmeir L, Tutz G. Multivariate Statistical Modelling Based on Generalized Linear

Models. Springer Verlag, 2001 (2nd edition).

8 Health Protection Agency. Serological diagnosis of syphilis. National Standard Method VS0P44 Issue 1. 2007. www.hpa-standardmethods.org.uk/pdf_sops.asp

9 Amin AK, Manuel RJ, Ison CA, *et al.*. Audit of laboratory diagnostic methods for syphilis in England and Wales. *Sex Transm Infect* 2009;**85**:88-91.

10 Health Protection Agency & UK National Screening Committee. Infectious diseases in pregnancy screening programme: report on 2005-2007 data. London: Health Protection Agency, March 2009.

11 CDSC. Outbreak of heterosexually acquired syphilis in Cambridgeshire. CDR Weekly Health Protections Agency website. November 10, 2000. Available at: http://www.hpa.org.uk/cdr/archives/2000/cdr4500.pdf. Accessed 22 June 2009.

12 Simms I, Ward H. Congenital syphilis in the United Kingdom. *Sex Transm Infect* 2006;**82**:1.

13 Singh S, Bell G, Talbot M. The characteristics of a recent syphilis outbreak in Sheffield, UK, and an evaluation of contact t*racing as a method of control. Sex Transm Infect 2007;***83:***193-9.

14 Lomax N, Wheeler H, Anaraki H, *et al.*. Management of a syphilis outbreak in street sex workers in east London. *Sex Transm Inf 2006; **82:**437-8.*

15 Welfare W, Lacey H, Lighton L, *et al.*. An outbreak of syphilis among young heterosexuals in an English Town. *In press Int J STD AIDS.*

16 Dhar J, Griffiths CA, Cassell JA, Sutcliffe L, Brook GM. How and why do South Asians attend GUM clinics? Evid*ence from contrasting GUM clinic*s across England? *Sex Transm Infect* 2010;**86**:366-370.

17 Centers fo*r Disease Contro*l *and Preventio*n. *Sexually Transmitted Disease Surveilla**nc**e, 2008. Atlanta, GA: U.S. Department of Heal**th** and Human Services; November 2009.*

Table & figures

T*able 1 Absolute (relative) frequencies of syphilis diagnoses reported through NESS by gender and male sexual orientation: 1999 to 2008*

*Parameter**		*Male heterosexual n (%)*	*Female heterosexual n (%)*	Men having sex with men n (%)	Total n (%)
Total		1938 (16.37)	1244 (10.51)	8656 (73.12)	11 838 (100.00)
Stage of infection	*Primary*	916 (47.27)	409 (32.88)	3373 (38.97)	4698 (39.68)
	Secondary	414 (21.36)	246 (19.77)	3017 (34.85)	3677 (31.07)
	Early latent	419 (21.62)	417 (33.952)	1786 (20.63)	2622 (22.15)
	Not available	189 (9.75	172 (13.83)	480 (5.55)	841 (7.10)
Whether born in the UK	*UK born*	707 (36.48)	469 (37.70)	1736 (20.06)	2912 (24.59)
	Non-UK born	1065 (54.95)	682 (54.82)	6143 (70.97)	7890 (66.65)
	Not available	166 (8.57)	93 (7.48)	777 (8.98)	1036 (8.75)
Ethnicity	*White*	1056 (54.49)	720 (57.88)	7715 (89.13)	9491 (80.17)
	Black	499 (25.75)	341 (27.41)	286 (3.30)	1126 (9.52)
	Asian	268 (13.83)	111 (8.92)	277 (3.20)	656 (5.54)
	Other ethnicity	61 (3.15)	45 (3.62)	215 (2.48)	321 (2.72)
	Not available	54 (2.79)	27 (2.17)	163 (1.88)	244 (2.07)
Reason for attending clinical services	*Routine screen*	361 (18.63)	314 (25.24)	2275 (26.28)	2950 (24.92)
	Symptoms	1058 (54.59)	259 (20.82)	4283 (49.48)	5600 (47.31)
	Campaign	14 (0.72)	3 (0.24)	21 (0.24)	38 (0.33)
	Contact tracing	200 (10.32)	191 (15.35)	722 (8.34)	1113 (9.40)
	Antenatal screening	7 (0.36)	276 (22.19)	1 (0.01)	284 (2.40)
	Other	163 (8.41)	136 (10.93)	641 (7.41)	940 (7.94)
	Not available	135 (6.97)	65 (5.23)	713 (8.24)	(7.71)
HIV status	*HIV positive*	110 (5.68)	54 (4.34)	3071 (35.48)	3235 (27.33)
	HIV negative	1350 (69.66)	923 (74.20)	4393 (50.75)	6666 (56.31)
	Not available	478 (24.66)	267 (21.46)	1192 (13.77)	1937(16.37)
Infection acquired through oral sex	*Yes*	192 (9.91)	48 (3.86)	2847 (32.89)	3087 (26.08)
	No	1095 (56.50)	735 (59.08)	2953 (34.12)	4783 (40.41)
	Not available	651 (33.59)	461 (37.06)	2856 (32.99)	3968 (33.52)
Location	*London*	551 (28.43)	297 (23.87)	2903 (33.54)	3751 (31.68)
	Brighton	20 (1.03)	14 (1.13)	437 (5.05)	471 (3.98)
	Manchester	66 (3.41)	37 (2.97)	1018 (11.76)	1121 (9.47)
	Elsewhere in UK	586 (30.24)	420 (33.76)	1828 (21.12)	2834 (23.94)
	Overseas	320 (16.51)	208 (16.72)	682 (7.88)	1210 (10.22)
	Not available	395 (20.38)	268 (21.54)	1788 (20.66)	2451 (20.70)
Commercial sex worker	*Yes*	11 (0.57)	111 (8.92)	92 (1.06)	214 (1.81)
	No	1534 (79.15)	873 (70.18)	6947 (80.26)	9354 (79.01)
	Not available	393 (20.28)	260 (20.90)	1617 (18.68)	2270 (19.18)
Social network	*Sauna*	26 (1.34)	8 (0.64)	1064 (12.29)	1098 (9.28)
	Bar	180 (9.29)	50 (4.02)	1265 (14.61)	1495 (12.63)
	Internet	9 (0.46)	5 (0.40)	535 (6.18)	549 (4.64)
	Not available	1723 (88.91	1181 (94.94)	5792 (66.91)	8696 (73.46)
Age	*≤15*	2 (0.10)	10 (0.80)	2 (0.02)	14 (0.12)
	16 to 19	51 (2.63)	128 (10.29)	170 (1.96)	349 (2.95)
	20 to 24	220 (11.35)	283 (22.75)	920 (10.63)	1423 (12.02)
	25 to 34	611 (31.53)	466 (37.46)	2710 (31.31)	3787 (31.99)
	35 to 44	542 (27.97)	209 (16.80)	3022 (34.91)	3773 (31.88)
	45+	459 (23.68)	114 (9.16)	1562 (18.05)	2135 (18.03)
	Not available	53 (2.73)	34 (2.73)	270 (3.12)	357 (3.02)
Strategic Health Authority	*East Midlands*	208 (10.73)	165 (13.26)	197 (2.28)	570 (4.81)
	East of England	21 (1.08)	14 (1.13)	43 (0.50)	78 (0.66)
	London	557 (28.74)	345 (27.73)	2888 (33.36)	3790 (32.02)
	North East	111 (5.73)	58 (4.66)	483 (5.58)	652 (5.51)
	North West	295 (15.22)	211 (16.96)	2315 (26.74)	2821 (23.83)
	South Central	65 (3.35)	34 (2.73)	198 (2.29)	297 (2.51)
	South East Coast	39 (2.01)	30 (2.41)	516 (5.96)	585 (4.94)
	South West	89 (4.59)	71 (5.71)	442 (5.11)	602 (5.08)
	West Midlands	144 (7.43)	121 (9.73)	256 (2.96)	521 (4.40)

	Yorkshire & The Humber	22 (1.14)	8 (0.64)	67 (0.77)	97 (0.83)
	Not available	387 (19.97)	187 (15.03)	1251 (14.45)	1825 (15.42)
Year	*1999*	5 (0.26)	10 (0.80)	32 (0.37)	47 (0.39)
	2000	7 (0.36)	5 (0.40)	74 (0.85)	86 (0.73)
	2001	62 (3.20)	44 (3.54)	355 (4.10)	461 (3.89)
	2002	147 (7.59)	79 (6.35)	643 (7.43)	869 (7.34)
	2003	242 (12.49)	144 (11.58)	808 (9.33)	1194 (10.09)
	2004	278 (14.34)	197 (15.84)	1091 (12.60)	1566 (13.23)
	2005	258 (13.31)	167 (13.42)	1170 (13.52)	1595 (13.47)
	2006	372 (15.20)	226 (18.17)	1397 (16.14)	1995 (16.85)
	2007	303 (15.63)	188 (15.11)	1568 (18.11)	2059 (17.40)
	2008	231 (11.92)	153 (12.30)	1309 (15.12)	1693 (14.30)
	Not available	33 (1.70)	31 (2.49)	20 (2.41)	273 (2.31)

* Partner notification is not shown because it is a discrete variable.

Table 2 Comparison of primary with early latent infection: estimates and confidence intervals for multinomial regression modelling*

Parameter		Estimate	Standard Error (95% CI)	*p*-value
Intercept		-12.36	142.88 (-292.66 to 267.93)	0.466
Strategic Health Authority	*East Midlands (baseline*)*			
	East of England	12.49	306.01 (-587.83 to 612.82)	0.484
	London	0.04	1.13 (-2.18 to 2.26)	0.486
	North West	0.65	1.04 (-1.38 to 2.68)	0.266
	South Central	10.99	215.44 (-411.65 to 466.64)	0.480
	South East Coast	-0.20	1.27 (-2.69 to 2.29)	0.437
	South West	0.62	1.12 (-1.58 to 2.81)	0.291
	West Midlands	0.26	1.18 (-2.05 to 2.56)	0.413
	Yorkshire & The Humber	10.68	290.09 (-558.41 to 579.76)	0.485
Gender	*Male (baseline*)*			
	Female	0.85	1.29 (-1.68 to 3.37)	0.255
Whether born in the UK	*UK-born (baseline*)*			
	Non UK-born	0.56	0.37 (-0.16 to 1.29)	0.064
Ethnicity	*White*			
	Black	-0.06	0.94 (-1.92 to 1.79)	0.473
	Asian	-1.20	0.68 (-2.53 to 0.14)	0.039
	Other ethnicity	0.02	0.92 (-1.79 to 1.82)	0.493
Sexual orientation	*Heterosexual (baseline*)*			
	Men who have sex with men	0.31	0.66 (-0.97 to 1.60)	0.317
Reason for attending clinical services	*Routine STI screen (baseline*)*			
	Symptoms	2.34	0.31 (1.74 to 2.94)	<0.001
	Campaign	0.26	1.56 (-2.79 to 3.31)	0.433
	Partner notification	0.77	0.47 (-0.15 to 1.70)	0.051
	Antenatal screening	-14.78	717.42 (-1422.19 to 1392.64)	0.492
	Other	-0.32	0.42 (-1.14 to 0.50)	0.224
HIV serostatus	*HIV negative (baseline*)*			
	HIV positive	0.48	0.30 (-0.10 to 1.06)	0.052
Infection acquired through oral sex	*No (baseline*)*			
	Yes	0.12	0.29 (-0.46 to 0.69)	0.346
Partner notification†	*Traceable*	-0.01	0.04 (-0.08 to 0.06)	0.399
	Untraceable	-0.02	0.01 (-0.03 to 0.00)	0.029
Location	*London (baseline*)*			
	Brighton	-0.76	0.75 (-2.23 to 0.70)	0.153
	Manchester	-0.97	0.56 (-2.08 to 0.14)	0.043
	Elsewhere UK	0.36	0.56 (-0.74 to 1.46)	0.259
	Overseas	-0.55	0.50 (-1.52 to 0.42)	0.134
Commercial sex worker	*No (baseline*)*			
	Yes	0.32	0.76 (-1.17 to 1.80)	0.337
Social network	*None (baseline*)*			
	Sauna	-0.95	0.31 (-1.56 to -0.34)	0.001
	Bar	-1.02	0.30 (-1.61 to -0.44)	<0.001
	Internet	-1.01	0.31 (-1.61 to -0.41)	0.001
Age	*16 to 19 (baseline*)*			
	20 to 24	-1.24	1.23 (-3.66 to 1.17)	0.157
	25 to 34	-1.15	1.21 (-3.52 to 1.23)	0.172
	35 to 44	-0.92	1.23 (-3.32 to 1.49)	0.227
	45+	-1.02	1.23 (-3.44 to 1.40)	0.204
Time	*Year*	0.01	0.07 (-0.13 to 0.15)	0.462

***** baseline individual: white heterosexual, UK-born HIV negative male, aged 16 to 19, who attended GUM services in the East Midlands for a routine STI screen, likely to have acquired syphilis in London but unlikely to have acquired infection through oral sex, not a commercial sex worker and with no stated social/sexual networks.

† no baseline is shown for partner notification because it is a discrete variable.

Interpretation: each estimated regression coefficient provides a measure of increased (if positive) or decreased (if negative) likelihood of presenting with primary, relative to early latent infection. For example, an HIV positive individual would be expected to have a

1.62=exp{0.48} higher odds ratio of presenting primary syphilis infection (as opposed to early latent) compared to an HIV negative patient.

Table 3 Comparison of secondary with early latent infection: estimates and confidence intervals for multinomial regression modelling*

Parameter		Estimate	Standard Error (95% CI)	*p*-value
Intercept		175.39	142.23 (-103.63 to 454.41)	0.109
Strategic Health Authority	*East Midlands (baseline*)*			
	East of England	13.80	306.01 (-586.52 to 614.13)	0.482
	London	1.42	1.29 (-1.11 to 3.94)	0.136
	North West	1.32	1.20 (-1.04 to 3.68)	0.136
	South Central	11.52	215.44 (-411.13 to 434.16)	0.479
	South East Coast	0.37	1.42 (-2.41 to 3.15)	0.397
	South West	1.53	1.27 (-097 to 4.03)	0.115
	West Midlands	0.37	1.37 (-2.32 to 3.06)	0.393
	Yorkshire & The Humber	11.19	290.09 (-557.90 to 580.28)	0.485
Gender	*Male (baseline*)*			
	Female	-0.74	1.60 (-3.88 to 2.39)	0.321
Whether born in the UK	*UK-born (baseline*)*			
	Non UK-born	0.61	0.36 (-0.10 to 1.32)	0.045
Ethnicity	*White (baseline)*			
	Black	0.60	0.94 (-1.24 to 2.44)	0.261
	Asian	-0.77	0.66 (-2.06 to 0.52)	0.12
	Other ethnicity	-0.40	0.90 (-2.17 to 1.38)	0.33
Sexual orientation	*Heterosexual (baseline*)*			
	Men who have sex with men	0.37	0.67 (-0.95 to 1.69)	0.292
Reason for attending clinical services	*Routine STI screen (baseline*)*			
	Symptoms	2.32	0.31 (1.71 to 2.92)	<0.001
	Campaign	-11.76	342.18 (-683.04 to 659.52)	0.486
	Partner notification	0.96	0.48 (0.02 to 1.90)	0.022
	Antenatal screening	-12.26	717.42 (-1419.67 to 1395.16)	0.493
	Other	0.33	0.39 (-0.43 to 1.10)	0.197
HIV serostatus	*HIV negative (baseline*)*			
	HIV positive	0.97	0.29 (0.40 to 1.54)	<0.001
Infection acquired through oral sex	*No (baseline*)*			
	Yes	-0.16	0.29 (-0.73 to 0.41)	0.29
Partner notification	*Traceable*	0.00	0.03 (-0.05 to 0.06)	0.438
	Untraceable	0.00	0.01 (-0.01 to 0.01)	0.441
Location	*London (baseline*)*			
	Brighton	-0.46	0.74 (-1.91 to -0.99)	0.268
	Manchester	-0.18	0.57 (-1.31 to 0.94)	0.374
	Elsewhere UK	0.85	0.57 (-0.27 to 1.96)	0.068
	Overseas	-0.37	0.48 (-1.32 to 0.58)	0.223
Commercial sex worker	*No (baseline*)*			
	Yes	-1.43	0.95 (-3.29 to 0.44)	0.067
Social network	*None (baseline*)*			
	Sauna	-0.78	0.30 (-1.38 to -0.19)	0.005
	Bar	-0.86	0.29 (-1.43 to -0.28)_	0.002
	Internet	-0.76	1.30 (-1.34 to -0.18)	0.005
Age	*16 to 19 (baseline*)*			
	20 to 24	-1.30	0.27 (-3.86 to 1.25)	0.158
	25 to 34	-0.72	1.27 (-3.22 to 1.77)	0.285
	35 to 44	-0.76	1.29 (-3.29 to 1.77)	0.278
	45+	-0.89	1.30 (-3.44 to 1.65)	0.245
Time	*Year*	-0.09	0.07 (-0.23 to 0.05)	0.108

* baseline individual: white heterosexual, UK-born HIV negative male, aged 16 to 19, who attended GUM services in the East Midlands for a routine STI screen, likely to have acquired syphilis in London but unlikely to have acquired infection through oral

sex, not a commercial sex worker and with no stated social/sexual networks.

† no baseline is shown for partner notification because it is a discrete variable.

Interpretation: each estimated regression coefficient provides a measure of increased (if positive) or decreased (if negative) likelihood of presenting with secondary, relative to early latent infection. For example, an HIV positive individual would be expected to have a 1.62=exp{0.48} higher odds ratio of presenting primary syphilis infection (as opposed to early latent) compared to an HIV negative patient.

Annex

Methods

Multinomial regression modelling is a extension to standard logistic regression allowing for three or more levels (say *J)* to be used for the categorical response variable.[1] A multinomial model was fitted with the *R* package VGAM to obtain maximum likelihood estimates.[2] In its most basic form, a multinomial model pairs each response category *j=1,...,J-1* with an arbitrary baseline level (*J*). Accordingly the *J-1* relative risks RR_{1J}, RR_{2J},..., $RR_{(J-1)J}$, induced by the ensuing set of pairwise contrasts, are modelled to be log-linearly related to a set of selected predictors. Here primary, secondary and early latent syphilis are considered (*J*=3), and so the model simultaneously estimates underlying log-odds ratios through the use of iterative re-weighted least squares.

Early latent infection was used as the baseline and the log-odds ratios were simultaneously modelled for each non-baseline stage of infection (primary and secondary) relative to the baseline. Parameter estimates from each regression equation were interpreted as in standard logistic regression.

Various multinomial models were fitted to the NESS data-set, each including a different set of explanatory variables and network of interactions. A stepwise search strategy was carried out to identify within the set of available predictors (and interactions) those significantly influencing the likelihood of syphilis infection at its various stages. Candidate multinomial models were derived from the full model with no interaction terms through both forward selection and backward elimination, whereby a predictor/interaction at a time is respectively included or excluded from the pool of explanatory variables. A number of 'core' predictors identified by the authors were also consistently retained within the analysis. Sparsely populated categories were merged to ensure robust

estimation of the underlying infection rates. The model offering the best compromise between goodness of fit and model complexity, as measured by the AIC statistic, was ultimately used for drawing predictions over time.[3]

Interaction terms were excluded because they did not increase the predictive power of the model. Alternative models encoding more structured time trends were also considered. Although a quadratic pattern was tested, it produced a model that had no appreciable difference in terms of overall accuracy-complexity balance, as indicated by corresponding AIC statistics (1382.045 from a linear time trend, 1383.154 from quadratic time trend).

Maximum likelihood estimates of regression coefficients were computed with corresponding standard errors, 95% confidence intervals (CI) and *p*-values. Some non-significant variables were retained within the model to preserve its robustness and facilitate the interpretation of its predictions. Baseline variables were arbitrarily taken as a white heterosexual, UK-born HIV negative male, who attended GUM services for routine STI screen within the East Midlands NHS SHA, was likely to have acquired syphilis in London although unlikely through oral sex, was not a commercial sex worker and had no stated social/sexual networks.

References

1 Fahrmeir L, Tutz G. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer Verlag, 2001 (2nd edition).

2 R Development Core Team (2010): "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria; url: http://www.R-project.org.

3 Spiegelhalter DJ, Best NG, Carlin BP, *et al.*. Bayesian Measures of Model Complexity and Fit. *J R Stat Soc Ser B Stat Methodol* 2005;**64**(4):583-639.

Figures

Figure 1

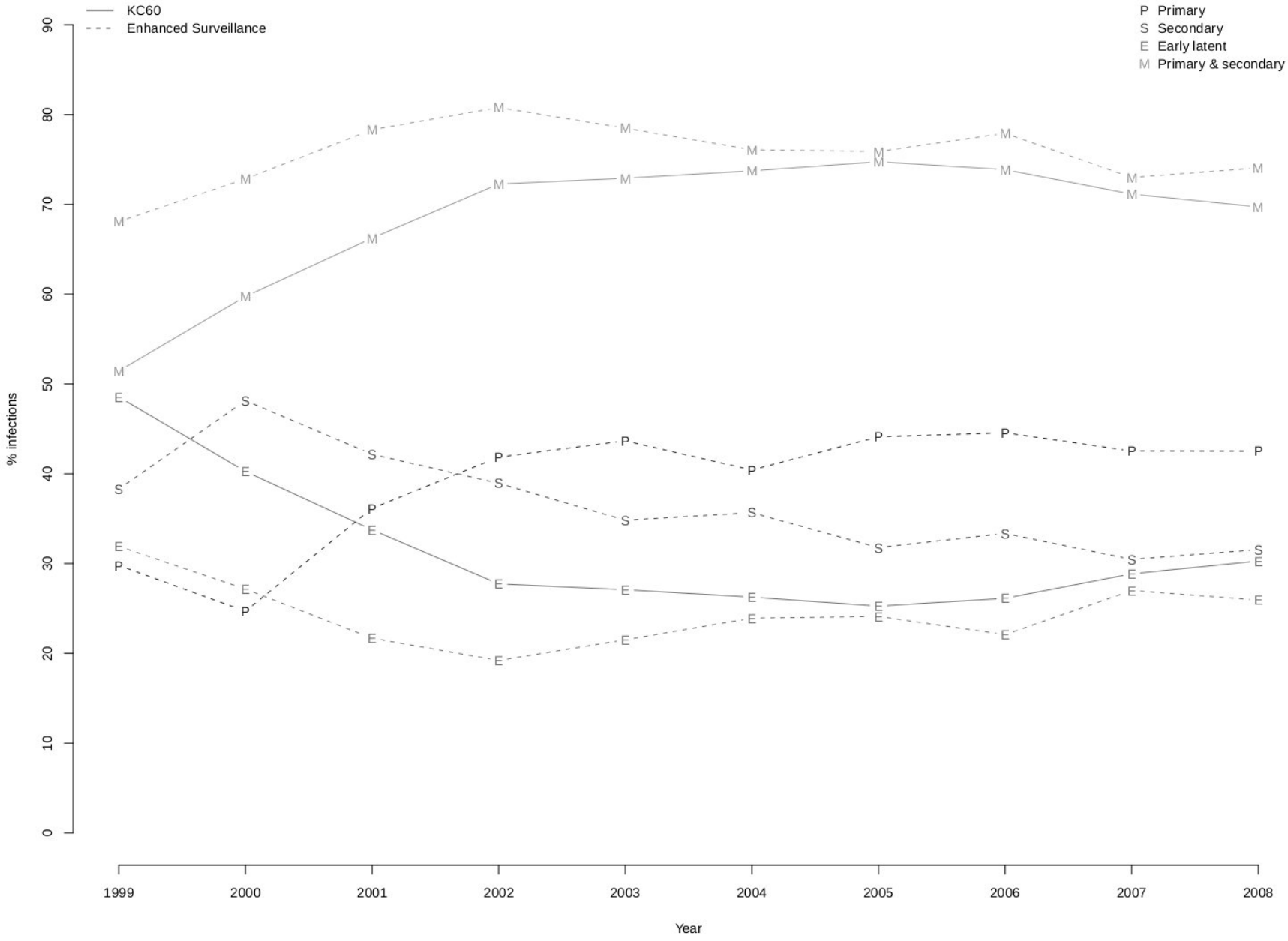

KC60
Enhanced Surveillance
P Primary
S Secondary
E Early latent
M Primary & secondary
% infections
0
10
20
30
40
50
60
70
80
90
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Year

Figure 2

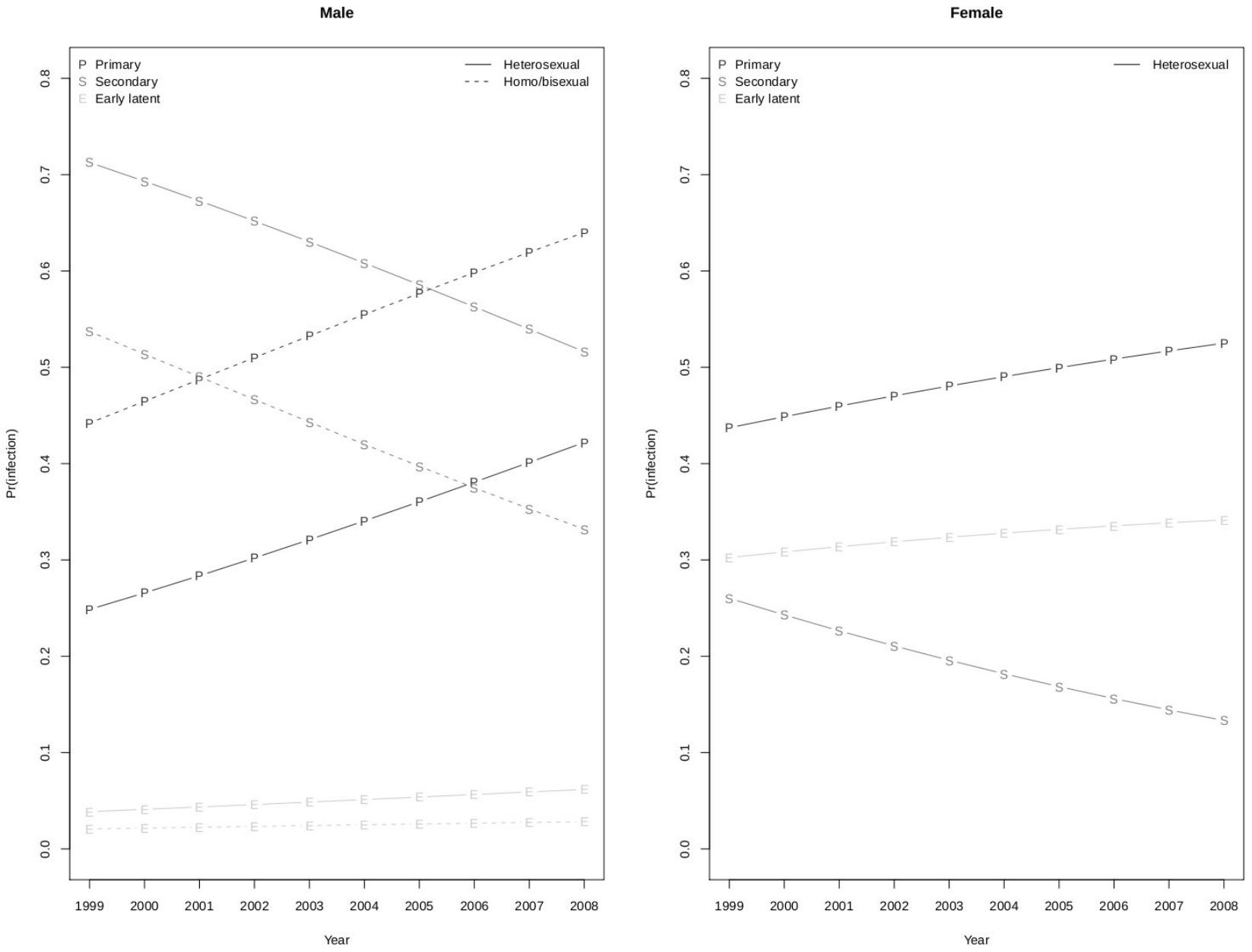

Male
Female
P Primary
S Secondary
E Early latent
Heterosexual
Homo/bisexual
Pr(infection)
Year
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

Figure A1

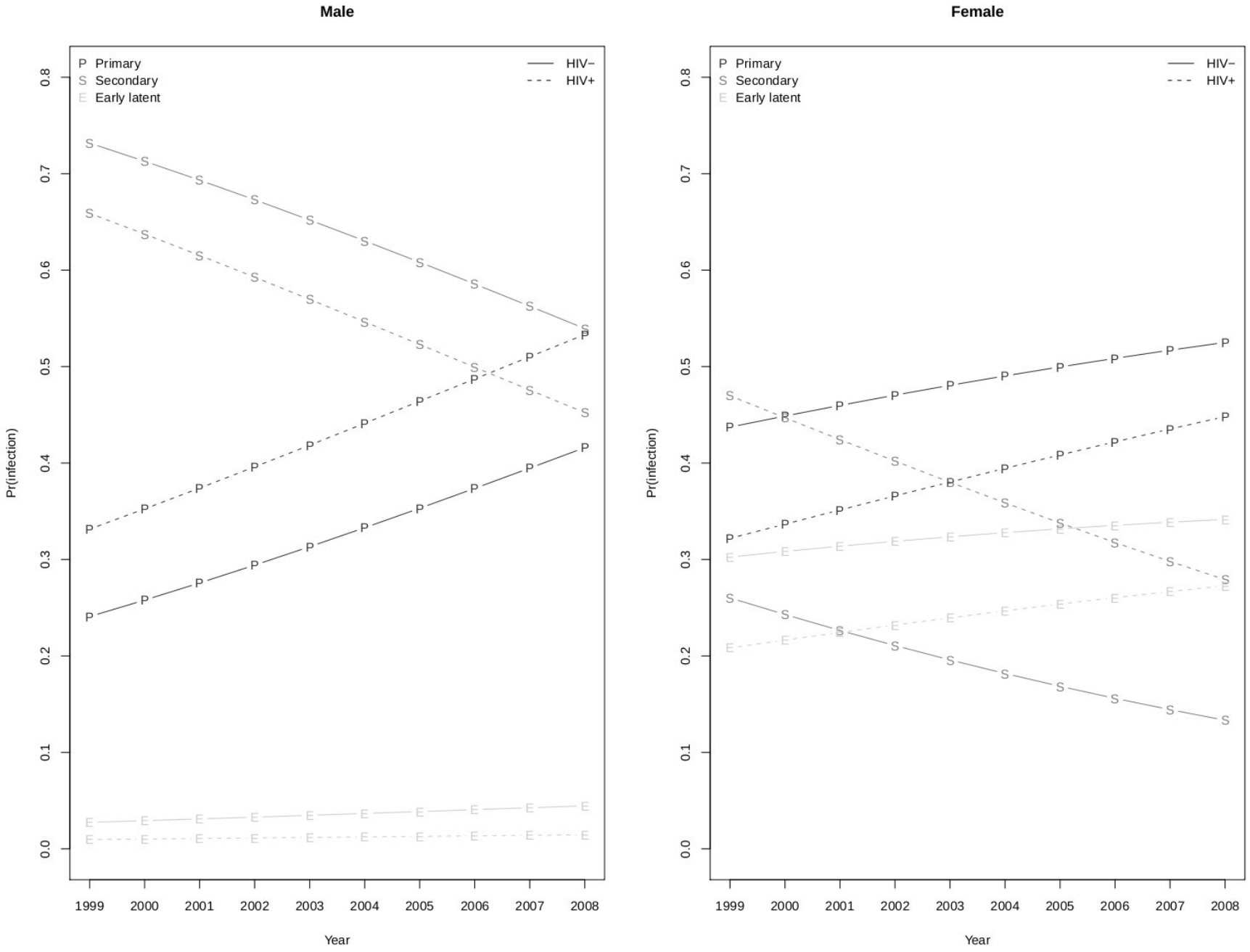

Male
Female
P Primary
S Secondary
E Early latent
HIV−
HIV+
Pr(infection)
Year
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

Figure A2

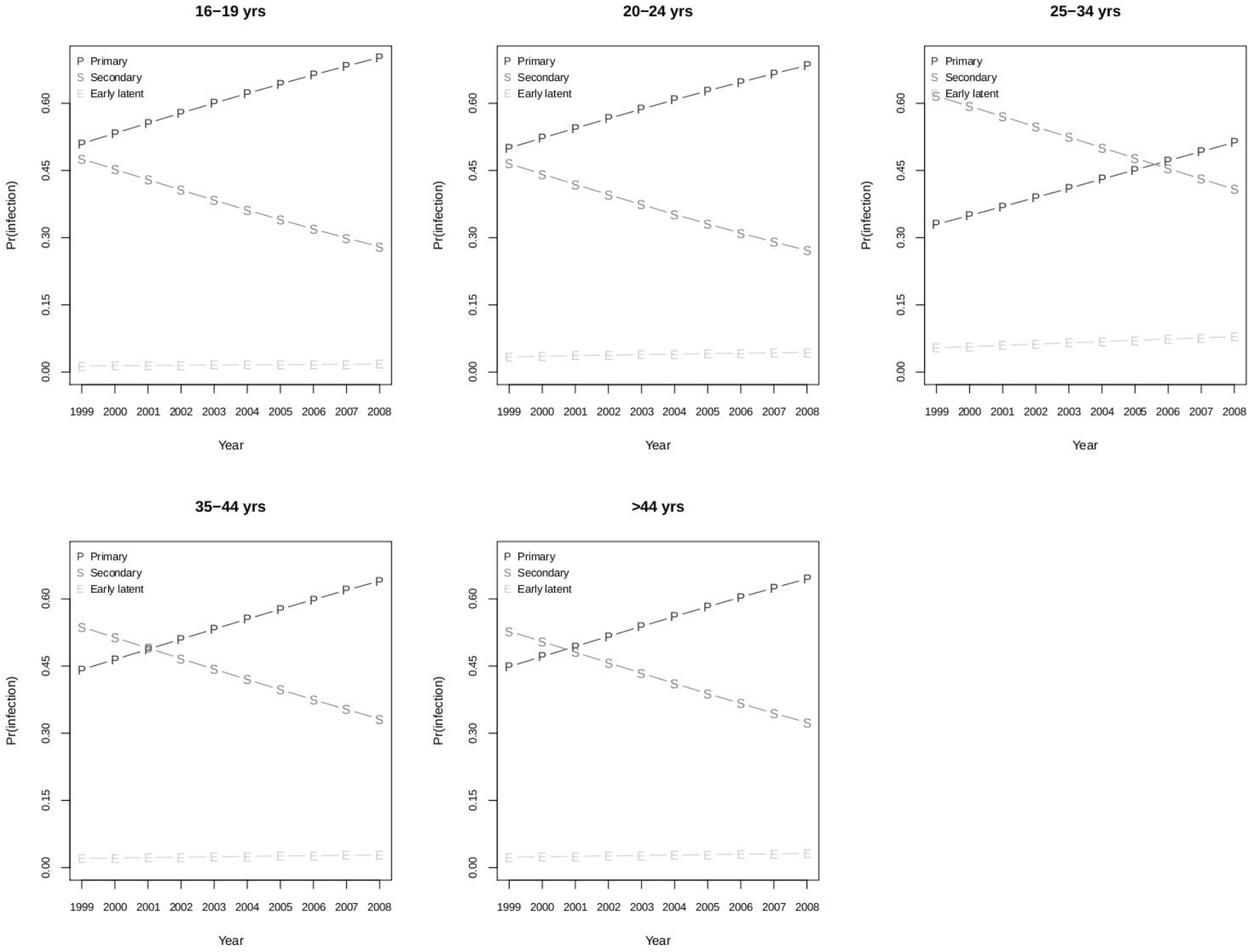

16−19 yrs
20−24 yrs
25−34 yrs
35−44 yrs
>44 yrs
P Primary
S Secondary
E Early latent
Pr(infection)
0.00
0.15
0.30
0.45
0.60
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
Year

Figure A3

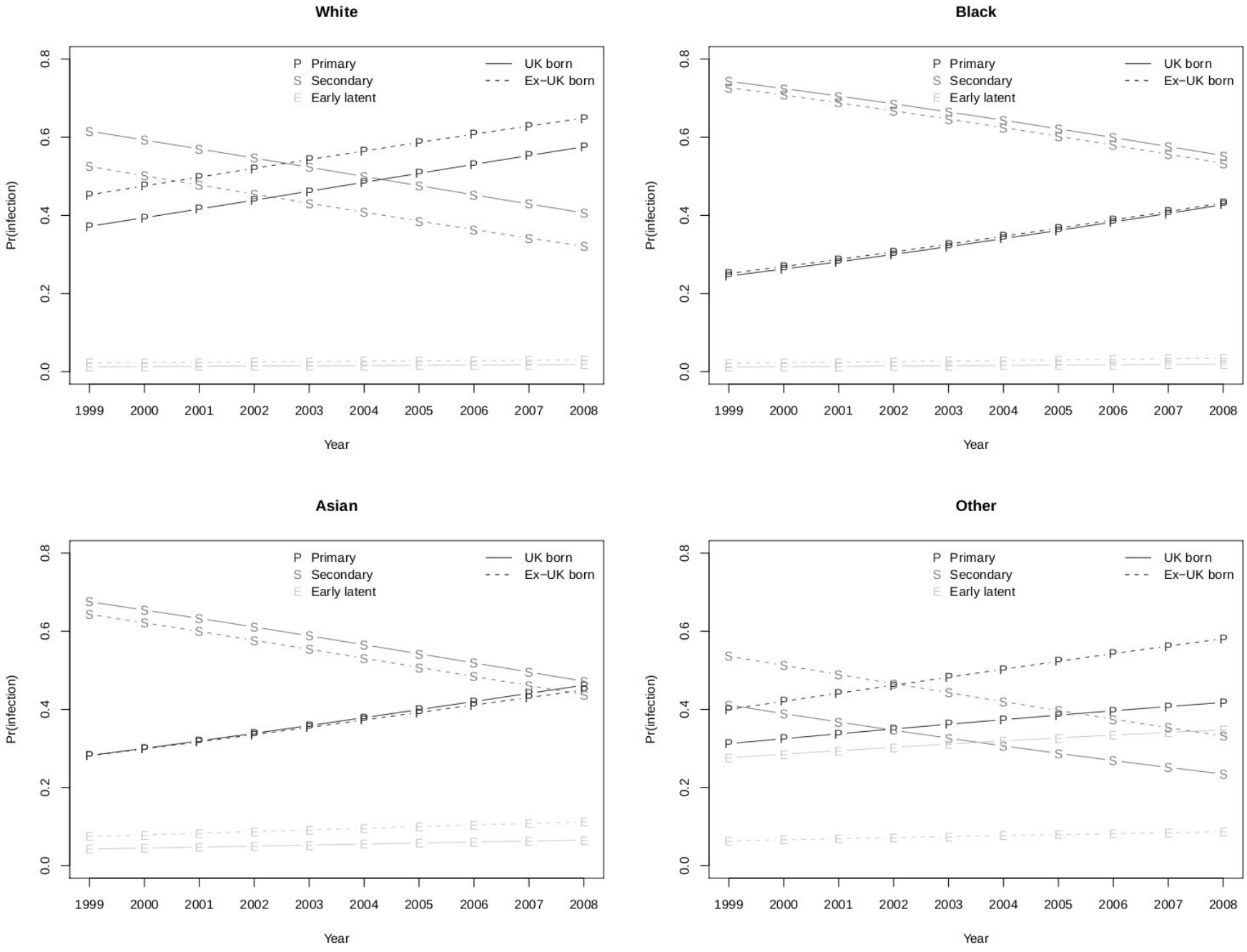

White
Black
Asian
Other
P Primary
S Secondary
E Early latent
UK born
Ex−UK born
Pr(infection)
Year
0.0
0.2
0.4
0.6
0.8
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

Figure A4

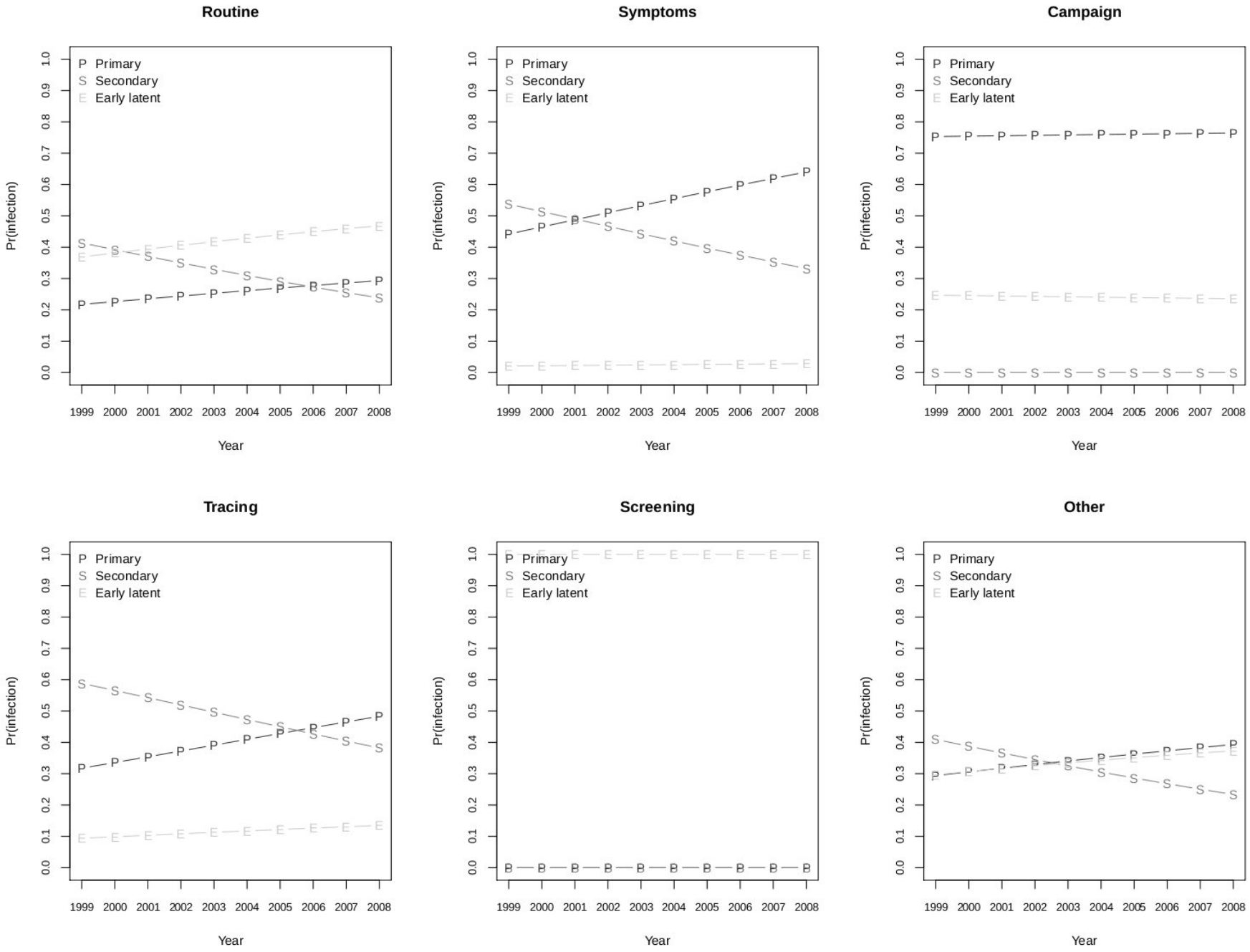
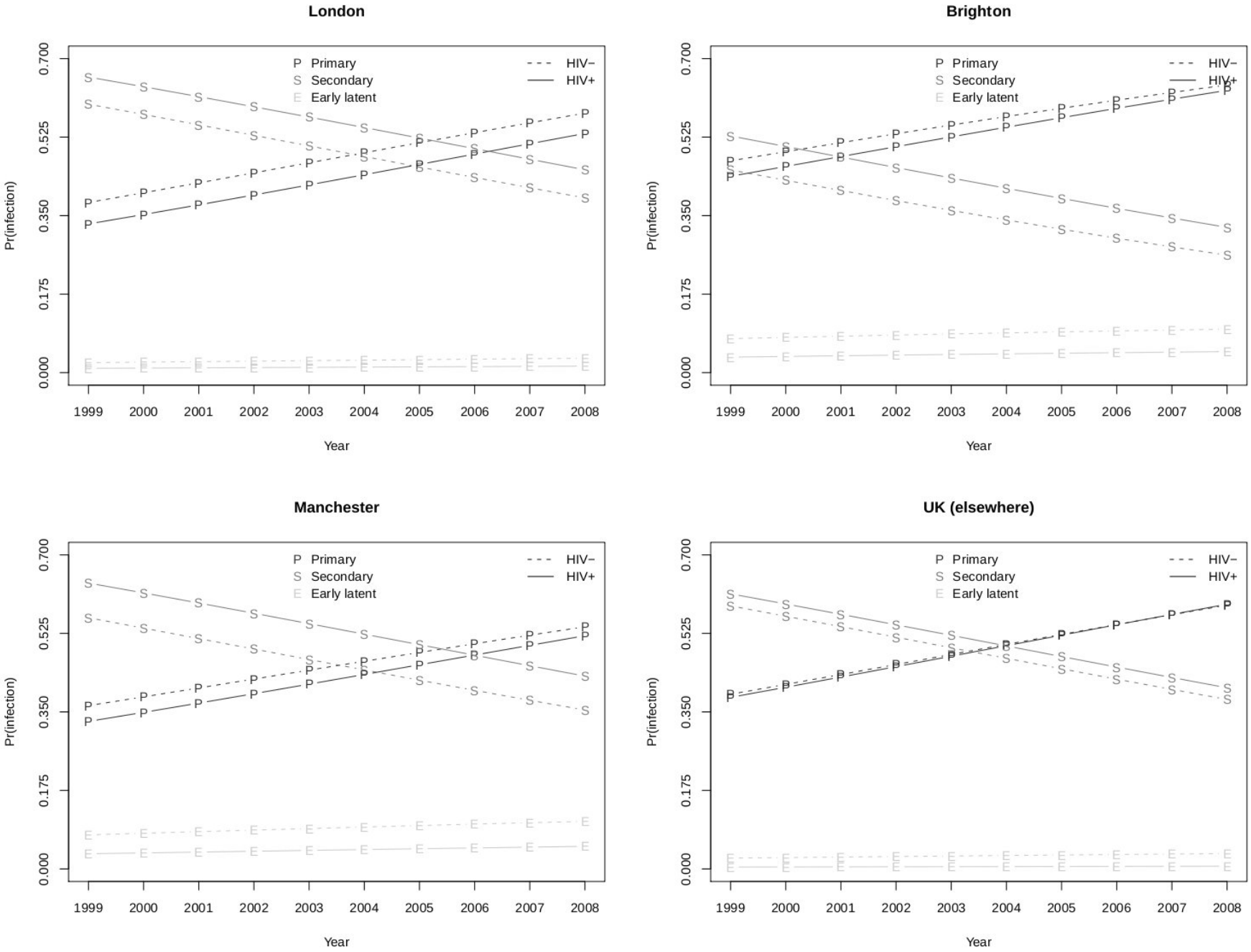

Routine
Symptoms
Campaign
Tracing
Screening
Other
P Primary
S Secondary
E Early latent
Pr(infection)
Year
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

Figure A5


London
Brighton
Manchester
UK (elsewhere)
P Primary
S Secondary
E Early latent
HIV−
HIV+
Pr(infection)
0.000
0.175
0.350
0.525
0.700
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Year