Gini coefficients for measuring the distribution of sexually transmitted infections among individuals with different levels of sexual activity

Sandro Gsteiger [1] , **Nicola Low** [1] , **Pam Sonnenberg** [2] , **Catherine H Mercer** [2] , **Christian L Althaus** [Corresp. 1]

1   Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland

2   Institute for Global Health, University College London, London, United Kingdom

Corresponding Author: Christian L Althaus
Email address: christian.althaus@alumni.ethz.ch

Objectives. Gini coefficients have been used to describe the distribution of *Chlamydia trachomatis* (CT) infections among individuals with different levels of sexual activity. The objectives of this study were to investigate Gini coefficients for different sexually transmitted infections (STIs), and to determine how STI control interventions might affect the Gini coefficient over time.

Methods. We used population-based data for sexually experienced women from two British National Surveys of Sexual Attitudes and Lifestyles (Natsal-2: 1999-2001; Natsal-3: 2010-2012) to calculate Gini coefficients for CT, *Mycoplasma genitalium* (MG), and human papillomavirus (HPV) types 6, 11, 16 and 18. We applied bootstrap methods to assess uncertainty and to compare Gini coefficients for different STIs. We then used a mathematical model of STI transmission to study how control interventions affect Gini coefficients.

Results. Gini coefficients for CT and MG were 0.33 (95% confidence interval (CI): 0.18-0.49) and 0.16 (95% CI: 0.02-0.36), respectively. The relatively small coefficient for MG suggests a longer infectious duration compared with CT. The coefficients for HPV types 6, 11, 16 and 18 ranged from 0.15-0.38. During the decade between Natsal-2 and Natsal-3, the Gini coefficient for CT did not change. The transmission model shows that higher STI treatment rates are expected to reduce prevalence and increase the Gini coefficient of STIs. In contrast, increased condom use reduces STI prevalence but does not affect the Gini coefficient.

Conclusions. Gini coefficients for STIs can help us to understand the distribution of STIs in the population, according to level of sexual activity, and could be used to inform STI prevention and treatment strategies.

1 **Gini coefficients for measuring the distribution of sexually transmitted infections among**

2 **individuals with different levels of sexual activity**

4 Sandro Gsteiger[1], Nicola Low[1], Pam Sonnenberg[2], Catherine H. Mercer[2], Christian L. Althaus[1]

6 [1]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland

7 [2]Institute for Global Health, University College London, London, UK

9 Corresponding author: Christian L. Althaus, Institute of Social and Preventive Medicine (ISPM),

10 University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland; christian.althaus@alumni.ethz.ch.

11   **Abstract**

13   **Objectives.** Gini coefficients have been used to describe the distribution of *Chlamydia*

14   *trachomatis* (CT) infections among individuals with different levels of sexual activity. The

15   objectives of this study were to investigate Gini coefficients for different sexually transmitted

16   infections (STIs), and to determine how STI control interventions might affect the Gini

17   coefficient over time.

19   **Methods.** We used population-based data for sexually experienced women from two British

20   National Surveys of Sexual Attitudes and Lifestyles (Natsal-2: 1999-2001; Natsal-3: 2010-2012)

21   to calculate Gini coefficients for CT, *Mycoplasma genitalium* (MG), and human papillomavirus

22   (HPV) types 6, 11, 16 and 18. We applied bootstrap methods to assess uncertainty and to

23   compare Gini coefficients for different STIs. We then used a conceptual transmission model for

24   STIs to study how control interventions affect Gini coefficients.

25

26   **Results.** Gini coefficients for CT and MG were 0.33 (95% confidence interval (CI): 0.18-0.49)

27   and 0.16 (95% CI: 0.02-0.36), respectively. The relatively small coefficient for MG suggests a

28   longer infectious duration compared with CT. The coefficients for HPV types 6, 11, 16 and 18

29   ranged from 0.15-0.38. During the decade between Natsal-2 and Natsal-3, the Gini coefficient

30   for CT did not change. The transmission model shows that higher STI treatment rates are

31   expected to reduce prevalence and increase the Gini coefficient of STIs. In contrast, increased

32   condom use reduces STI prevalence but does not affect the Gini coefficient.

33

34   **Conclusions.** Gini coefficients for STIs can help us to understand the distribution of STIs in the

35   population, according to level of sexual activity, and could be used to inform STI prevention and

36   treatment strategies.

37 **Introduction**

38

39 Understanding how sexually transmitted infections (STIs) are distributed among individuals is

40 important both from a biological and from a public health perspective. Differences in STI

41 prevalence within a population, between groups with varying levels of sexual activity, can

42 provide information about biological and epidemiological characteristics of the infection. For

43 example, an STI with a long infectious duration, such as human papillomavirus (HPV), will tend

44 to be spread more evenly across a population than an STI with a short infectious duration, such

45 as *Neisseria gonorrhoeae* (NG). This observation can be explained by the fact that STIs with

46 short infectious durations require a higher rate of sexual partner change for sustained spread in

47 the population. NG is thus more concentrated in a small subgroup of individuals with high sexual

48 activity. Such ideas were initiated in the late 1970s and led to the concept of the 'core group' [1].

49 In 1990, Brunham & Plummer [2] inferred the size of core groups for various STIs from the

50 biological parameters that describe transmissibility and infectious duration, and discussed the

51 implications for selecting adequate STI control strategies.

52

53 The Gini coefficient can be used to quantify the degree of concentration of an STI in a

54 population. Originally introduced for describing inequalities in income distributions [3], the Gini

55 coefficient provides a general tool to measure the distribution or imbalance of a disease outcome

56 in relation to an exposure variable or risk factor [4], such as the geographic location or sexual

57 behavior. A Gini coefficient of zero denotes perfect equality where an infection is equally

58 distributed across a population. For infections that are concentrated in specific subpopulations,

59 the Gini coefficient can increase up to a maximal value of one. The Lorenz curve is a visual

60 representation of the cumulative distribution of a disease when ordered according to the risk

61 factor [5]. The diagonal line on a Lorenz curve plot denotes perfect equality, e.g., every

62 subpopulation has the same prevalence of an STI.

63

64 Several groups have used Gini coefficients and Lorenz curves to describe how *Chlamydia*

65 *trachomatis* (CT), NG, syphilis or herpes are distributed across different geographical regions in

66 Canada [6], the UK [7] and the US [8–10]. These findings have helped to summarize inequalities

67 in STI distributions, assess the suitability of geographically targeted interventions, and provide

68 insights into the epidemic phases of the STIs over time. Chen et al. [11] were the first to apply

69 the concept of Gini coefficients in mathematical transmission models. They proposed a

70 metapopulation modeling framework that better captures the sociogeographic epidemiology of

71 NG and compared the resulting Gini coefficients with empirical estimates.

72

73 The way in which modifiable factors, such as sexual activity and STI control interventions affect

74 the Gini coefficient has been less-well studied. Previously, we described the distribution of CT

75 infections among individuals with different sexual activity using Lorenz curves and Gini

76 coefficients to calibrate dynamic transmission models [12]. We used data from the second British

77 Survey of Sexual Attitudes and Lifestyles (Natsal-2, 1999-2001) [13], which included CT test

78 results (samples were later tested for HPV [14]). The most recent survey (Natsal-3, 2010-2012)

79 also provides information on HPV and *Mycoplasma genitalium* (MG) positivity and offers a

80 unique opportunity to study relationships between sexual behavior and STI prevalence [15–17].

81

82 This study had two objectives. First, we wanted to estimate and compare the Gini coefficients for

83 different STIs and over time using data from two Natsal surveys. Second, we used a

84 mathematical transmission model to obtain general insights into how Gini coefficients are

85 expected to change as a result of STI control interventions.

86

87 **Methods**

88

89 *Data*

90 Natsal-3 is a population-based probability sample survey of sexual attitudes and lifestyles

91 conducted in Britain (England, Scotland, and Wales) and carried out from 2010 to 2012 [18,19].

92 The sample consists of 15162 women and men aged 16-74 years. A subsample of participants

93 aged 16-44 years who reported at least one sexual partner over their lifetime were asked to

94 provide urine samples, resulting in laboratory confirmed STI test results from 2665 women and

95 1885 men [15]. Urine was tested for the presence of CT, MG, type-specific HPV, NG, and HIV

96 antibodies. To compare Gini coefficients for CT over time, we also used data from the Natsal-2

97 survey, which was carried out in 1999-2001 [20]. This survey includes 11,161 women and men

98 aged 16-44 years. Urine samples for ligase chain reaction (LCR) testing for CT were available

99   for a subset of 2055 and 1474 sexually active women and men aged 18-44 years [13]. Unless

100   otherwise stated, we used the subpopulations that provided urine samples for our analysis from

101   both surveys. Sexual behavior and urine sample data were individually weighted to adjust for

102   unequal selection probabilities and to correct for the age, gender and regional profiles in the

103   survey sample. For simplicity, we did not include same-sex contacts for the sexual behavior

104   variables because only 2.2% of the population reported a new same-sex partner during time

105   period covered by Natsal-3. The full datasets of both surveys are available from the UK Data

106   Archive at the University of Essex (http://data-archive.ac.uk).

107

108   *Statistical analysis*

109   We used Lorenz curves to plot the cumulative proportion of specific STI infections (y_i) as a

110   function of the cumulative proportion of the population (x_i), after population sub-groups i (

111   $i = 1, 2, ..., n$) have been ranked according to their level of sexual activity. The Gini coefficient is

112   defined as the area between the line of equality and the Lorenz curve over the total area below

113   the line of equality:

114
$$G = 1 - \sum_{i=1}^{n} (x_i - x_{i-1})(y_i + y_{i-1}),$$

115   where $x_0 = y_0 = 0$ and $x_n = y_n = 1$.

116

117   We derived Lorenz curves and estimated Gini coefficients for CT, MG, and HPV types 6, 11, 16

118   and 18. We focused on these four HPV genotypes because they are present in the widely used

119   quadrivalent vaccine and are frequently considered in dynamic transmission models [21]. We did

120   not include NG and HIV in our analysis because of the small numbers of positive tests. We used

121   the number of new opposite-sex partners in the last year as the exposure variable summarizing

122   sexual activity because of its strong association with STI prevalence [12], and frequent use to

123   parameterize dynamic transmission models [22]. Owing to the larger sample size for female

124   respondents and a potential bias resulting from lower test sensitivity to detect HPV infections in

125   male urine [15], we focused our analysis on women and provide a separate analysis for men as

126   Supplemental Information. We constructed bootstrap confidence intervals (CIs) for the Gini

127   coefficients and point-wise bootstrap confidence bands for the Lorenz curves by sampling with

128   replacement [8–10,23]. We calculated the 2.5th and 97.5th percentiles from 1000 bootstrapped

129   Gini coefficients and Lorenz curves.

130

131   *Transmission model*

132   We adapted a previously described mathematical model of STI transmission [24] to investigate

133   how changes in infectious duration and transmissibility affect the prevalence and the Gini

134   coefficient of an STI in a simulated population. We stratified the population according to sexual

135   activity [1,25], i.e., we assumed n different sexual activity classes with 0, 1, 2, …, n-1 new

136   opposite-sex partners per year. For simplicity, we assumed that sexual activity and the natural

137   history and transmission of the infection are the same in women and men. The SIS (susceptible-

138   infected-susceptible) transmission dynamics can be described by the following ordinary

139   differential equation (ODE):

140
$$\frac{dy_i}{dt} = \mu \sum_{j=1}^{n} x_j y_j + \beta c_i (1 - y_i) \sum_{j=1}^{n} \rho_{ij} y_j - \gamma y_i - \mu y_i,$$

141   where y_i is the proportion of individuals in sexual activity class i who are infected, and x_i is the

142   proportion of all individuals who belong to sexual activity class i. The first and last term of the

143   equation describe how individuals can change their sexual activity class at rate μ and be

144   redistributed to either the same or another sexual activity class proportional to the size of sexual

145   activity classes [24,26]. The middle terms describe the process of transmission and the clearance

146   of infection at rate γ. Susceptible individuals $1 - y_i$ have an average of c_i new opposite-sex

147   partners per year. β is the per partnership transmission probability and y_j is the probability that a

148   partner in sexual activity class j is infected. ρ_{ij} represents the elements of the sexual mixing

149   matrix [27]

150
$$\rho_{ij} = \epsilon \delta_{ij} + (1 - \epsilon) \frac{c_j x_j}{\sum_{l=1}^{n} c_l x_l},$$

151   where δ_{ij} denotes the Kronecker delta (it is equal to 1 if $i = j$ and to 0 otherwise).

152

153   The proportion of individuals in sexual activity classes x_i with $c_i \in 0, 1, 2, ..., n - 1$ new opposite-

154   sex partners per year were based on Natsal-2. All data for 18-44 year old women and men were

155   pooled and weighted. Since the transmission model is primarily used for illustrative purposes, we

156   did not include changes in sexual behavior, between Natsal-2 and Natsal-3, which were small

157   [19]. We set $\mu = 1$ per year as described previously [24,26], and set $\epsilon = 0$, i.e., we assumed

158   random proportional mixing between different sexual activity classes [27]. We then chose a

159   particular combination of the infectious duration ($1/\gamma$) and the per partnership transmission

160   probability (β) and ran the model into endemic equilibrium. We recorded the simulated STI

161   prevalence and computed the Gini coefficient as described above, i.e., by calculating the

162   cumulative proportion of infections as a function of the cumulative proportion of the population,

163   ranked by sexual activity i. We repeated this process for various combinations of the infectious

164   duration and the per partnership transmission probability, which allowed us to map between the

165   (unobservable) model parameters and the simulated (observable) summary measures prevalence

166   and Gini coefficient.

167

168   We further used the transmission model to investigate how two hypothetical CT control

169   interventions affect infection prevalence and the Gini coefficient. First, we calibrated the

170   infectious duration and the per partnership transmission probability such that the simulated CT

171   prevalence and Gini coefficient correspond to the estimates from Natsal-2 ($1/\gamma = 1.75$ years, $\beta = $

172   19%). Second, we simulated the expected changes in prevalence and the Gini coefficient as a

173   result of a) an increase in CT screening coverage aiming to detect asymptomatic CT cases, and

174   b) the effects of an educational campaign that leads to a change in sexual behavior and/or an

175   increase in condom use. We assumed that the first intervention would reduce the overall

176   infectious duration by 10%. For the second intervention, we assumed a 10% reduction in the

177   product of the per partnership transmission probability (β) and the number of opposite-sex

178   partners (c_i). Note that these hypothetical interventions do not necessarily predict the quantitative

179   effects of real-world interventions; instead, they provide a qualitative picture of how prevalence

180   and Gini coefficients are expected to change.

181

182   Data analyses and model simulations were performed in the R software environment for

183   statistical computing [28]. All code files are available on GitHub

184   (https://github.com/calthaus/gini).

185

186

187

188 **Results**

189

190 *Lorenz curves and Gini coefficients*

191 The Lorenz curves for MG and HPV types 6, 11 and 16 are closer to the diagonal line than the

192 Lorenz curves for CT and HPV 18 (Fig. 1A). This indicates that CT and HPV-18 in women are

193 more strongly associated with the number of new opposite-sex partners in the last year than MG

194 and the other type-specific HPV. The Gini coefficients mirror this observation and are higher for

195 CT (0.33) and HPV 18 (0.38) than for MG and the different HPV types (≤ 0.22) (Table 1).

196 However, the bootstrapped confidence intervals for the Lorenz curves are wide (figure 1B),

197 resulting in considerable uncertainty in the estimated Gini coefficients. The Lorenz curves for

198 CT for the two survey periods of Natsal-2 (1999-2001) and Natsal-3 (2010-2012) are similar

199 (figure 1C) with a Gini coefficient in Natsal-2 of 0.30 (95% CI: 0.12-0.50) and in Natsal-3 of

200 0.33 (95% CI: 0.18-0.49).

201

202 [INSERT FIG. 1 HERE.]

203

204 [INSERT TABLE 1 HERE.]

205

206 We used the transmission model to perform a mapping between model parameters and the

207 simulated STI prevalence and Gini coefficient. Within the parameter ranges that are

208 representative for the considered STIs, we found a close to linear relationship between the Gini

209 coefficient, infection prevalence, infectious duration and per partnership transmissibility (Fig.

210 2A, dashed grid). STIs with a short infectious duration (e.g., 1 year) require frequent sexual

211 partner changes (i.e., a high number of opposite-sex partners) and are therefore characterized by

212 a high Gini coefficient. Longer infectious durations increase prevalence, facilitate STI

213 transmission between individuals with a low number of opposite-sex partners, and consequently

214 decrease the Gini coefficient. Interestingly, different values for the per partnership transmission

215 probability only influence prevalence but do not affect the Gini coefficient. These insights from

216 the transmission model potentially allow the inference of biological parameters for the different

217   STIs in Natsal-3 (Fig. 2A, colored dots). Although the confidence intervals and the associated

218   uncertainty are large, CT and HPV 18 seem to be consistent with an infectious duration between

219   1 and 2 years. MG and the other HPV types are consistent with longer infectious durations.

220

221   [INSERT FIG. 2 HERE.]

222

223   *STI control interventions*

224   We used the transmission model to examine how two hypothetical control interventions for CT

225   affect the Gini coefficient and infection prevalence. First, we assumed that the increase in CT

226   screening between the two survey periods of Natsal-2 (1999-2001) and Natsal-3 (2010-2012)

227   through the National Chlamydia Screening Programme (NCSP) in England [29] has resulted in a

228   reduction in the overall infectious duration. The transmission model predicts that this would

229   result in an increase of the Gini coefficient with a concurrent drop in prevalence (Fig. 2B). In

230   contrast, the point estimates of the Gini coefficient and prevalence both show a slight, albeit no

231   statistically significant increase between Natsal-2 and Natsal-3. The expected changes in

232   prevalence and Gini coefficient might be relatively small, however, and may remain within the

233   95% CIs of the point estimates. Second, we simulated the effects of a decrease in the number of

234   opposite-sex partners and/or an increase in condom. Such a behavior change would be expected

235   to reduce the prevalence of CT without affecting the Gini coefficient.

236

237   **Discussion**

238

239   Building upon earlier work [12], we constructed Lorenz curves and estimated Gini coefficients in

240   women to investigate how different STIs are distributed according to sexual activity. Gini

241   coefficients for CT and HPV 18 appear to be higher than for MG and HPV 6, 11 and 16. We

242   found no evidence that the Gini coefficient for CT changed between the two survey periods of

243   Natsal-2 and Natsal-3. Using a mathematical model of STI transmission, we found that a CT

244   screening intervention should reduce prevalence and increase the Gini coefficient, whilst condom

245   use reduced prevalence but did not affect the Gini coefficient.

246

247   A main strength of this study was the availability of Natsal-2 and Natsal-3, two very large data

248   sets that measure both STI positivity and self-reported sexual behavior in probability samples of

249   the British general population. These comprehensive data sets allow comparison between

250   different STIs and over time. Calculating infection prevalence and Gini coefficients is

251   straightforward if suitable data are available. In contrast, obtaining biological parameters that

252   determine the transmission dynamics, such as the infectious duration or the transmission

253   probability, are notoriously difficult to obtain [30,31]. In this study, we used a mathematical

254   model of STI transmission with a detailed description of different sexual activity classes [24] to

255   explore the relationship between Gini coefficient, infection prevalence, infectious duration and

256   transmissibility.

257

258   There are several limitations to this study. First, despite the large overall sizes of the two Natsal

259   surveys, the relatively low prevalence of STIs in the general population [13–16] resulted in

260   relatively large uncertainties in the Lorenz curves and the corresponding Gini coefficients,

261   particularly for males (see Supplemental Information). Owing to the small sample sizes, we

262   pooled data over all age groups and restricted our analyses to the whole survey population.

263   Investigating sex- or age-specific differences in Gini coefficients would certainly be interesting

264   but is currently not feasible. Second, the limited sample size did not allow us to calculate Gini

265   coefficients for NG and HIV, which have a very low prevalence, and arguably high Gini

266   coefficient [15]. Third, the comparison of prevalence and Gini coefficients of CT between

267   Natsal-2 and Natsal-3 should also be treated with caution. The two surveys used different

268   laboratory tests and were not powered to detect changes in CT prevalence [15]. Fourth, we

269   defined the exposure variable as the number of new opposite-sex partners in the last year. Using

270   other exposure variables for our analysis would obviously affect the Lorenz curves and the Gini

271   coefficients, but would not necessarily be applicable for the mapping between model parameters

272   and the summary measures. Fifth, mathematical modeling necessarily involves several

273   assumptions and simplifications. As in the data analysis, we did not stratify the population

274   according to age, assumed the sexual behavior in women and men to be the same, and considered

275   the general population of those reporting sex with opposite-sex partners as a whole. Changes in

276   sexual behavior between Natsal-2 and Natsal-3 were minimal [19], so we did not take these into

277   account. We assumed fully proportional mixing which typically results in the best description of

278   infection prevalence in different sexual activity classes in models with a constant per partnership

279   transmission probability [22,24,27]. We also did not consider sex-specific differences in the

280   infectious duration and the transmissibility of the various STIs, which might limit the application

281   of Gini coefficients for the inference of these parameters. Further, we assumed that individuals

282   who clear an infection can become reinfected, although CT and type-specific HPV infections

283   might confer temporal immunity [32,33]. Together, these modeling assumptions highlight that

284   the insights from our conceptual transmission model are primarily qualitative and that

285   quantitative results should be treated with caution.

286

287   The calculated Gini coefficients and prevalence of CT and HPV 18 in women in Natsal-3

288   suggest an infectious duration of 1-2 years, which is in good agreement with previous estimates

289   [30,34–36]. Our mapping indicates that the infectious durations for the other high-risk HPV type

290   16 and low-risk HPV types 6 and 11 could be longer than 2 years. This interpretation would

291   contrast with previous studies that estimate similar infectious durations for HPV-16 and HPV-18

292   [34,35], and shorter infectious durations of less than a year for HPV 6 and 11 [34]. The

293   discrepancy could be a result of ignoring the effects of temporal immunity to reinfection. The

294   infectious duration for the other bacterial STI, MG, seems to be longer than for CT. There is

295   considerable uncertainty regarding the infectious duration of MG. One analysis, which used data

296   from a study of female students in London (UK), estimated the mean infectious duration at 15

297   months [37], which is in the same range as CT. The per partnership transmission probability is a

298   highly model-dependent parameter and depends on the type of sexual partnerships that are

299   considered. It is maybe not surprising that our mapping, which suggests relatively low per

300   partnership transmission probabilities of 10% to 25%, is not consistent with estimates from other

301   modeling studies [31,33].

302

303   Our findings allowed us to interpret differences in Gini coefficients and changes over time for

304   different STIs, illustrating that Gini coefficients can serve beyond their original role as simple

305   statistical measures of exposure-outcome associations. We found that changes in the

306   transmission probability only influence infection prevalence. This means that while decreasing

307   the transmission probability (e.g., through increased condom use) decreases the overall burden of

308   an STI, it does not affect how an STI is distributed among individuals with different sexual

309   activity. Hence, the target groups for future control interventions should remain the same. In

310   contrast, we showed that changes in the infectious duration (e.g., through increased testing and

311   treatment uptake) affects both the prevalence and Gini coefficient of an STI. A stronger

312   concentration of the infections among individuals with increased sexual activity would require a

313   change in the target groups for control interventions. A prerequisite for this newly proposed use

314   of Gini coefficients for STIs will be the availability of large data sets with a sufficiently high

315   infection prevalence. These could either be population-based surveys or smaller cohorts that

316   focus on individuals with an increased risk of acquiring STIs, such as men who have sex with

317   men (MSM).

318

319   In summary, our study illustrates that the Gini coefficient for measuring the distribution of an

320   STI among individuals with different sexual activity represents a simple proxy measure, which

321   combines epidemiological and behavioral data. Estimating Gini coefficients for the general

322   population or particular sub-populations, in combination with mathematical modeling, has the

323   potential to make inference of biological parameters that determine STI transmission and to

324   assess the impact of control measures.

325   **Acknowledgements** Natsal-3 is a collaboration between University College London (UCL),

326   London School of Hygiene and Tropical Medicine (LSHTM), National Centre for Social

327   Research (NatCen), Public Health England (PHE), and the University of Manchester. Natsal-2

328   was a collaboration between UCL, LSHTM, NatCen, and the Health Protection Agency (now

329   PHE). We thank the study participants, the team of interviewers from NatCen Social Research,

330   and operations and computing staff from NatCen Social Research. Natsal-3 was approved by the

331   Oxfordshire Research Ethics Committee A (Ref: 10/H0604/27); Natsal-2 was approved by the

332   University College Hospital and North Thames Multi-Centre Research Ethics Committee and all

333   the Local Research Ethics Committees in Britain.

334

References

1   Hethcote Herbert W, Yorke JA. *Gonorrhea transmission dynamics and control*. Springer-Verlag, Berlin 1984.

2   Brunham RC, Plummer FA. A general model of sexually transmitted disease epidemiology and its implications for control. *Med Clin North Am* 1990;**74**:1339–52.

3   Gini CW. Variability and Mutability, contribution to the study of statistical distributions and relations. *Studi Economico-Giuridici della R Universita de Cagliari* 1912.

4   Lee WC. Characterizing exposure-disease association in human populations using the Lorenz curve and Gini index. *Stat Med* 1997;**16**:729–39.

5   Lorenz MO. Methods of measuring the concentration of wealth. *Journal of the American Statistical Association* 1905;**9**:209–19.

6   Elliott LJ, Blanchard JF, Beaudoin CM, Green CG, Nowicki DL, Matusko P, Moses S. Geographical variations in the epidemiology of bacterial sexually transmitted infections in Manitoba, Canada. *Sex Transm Infect* 2002;**78 Suppl 1**:i139–44.

7   Monteiro EF, Lacey CJN, Merrick D. The interrelation of demographic and geospatial risk factors between four common sexually transmitted diseases. *Sex Transm Infect* 2005;**81**:41–6.

8   Kerani RP, Handcock MS, Handsfield HH, Holmes KK. Comparative geographic concentrations of 4 sexually transmitted infections. *Am J Public Health* 2005;**95**:324–30.

9   Chesson HW, Sternberg M, Leichliter JS, Aral SO. Changes in the state-level distribution of primary and secondary syphilis in the USA, 1985-2007. *Sex Transm Infect* 2010;**86 Suppl 3**:iii58–62.

10 Chesson HW, Sternberg M, Leichliter JS, Aral SO. The distribution of chlamydia, gonorrhoea and syphilis cases across states and counties in the USA, 2007. *Sex Transm Infect* 2010;**86 Suppl 3**:iii52–57.

11 Chen MI, Ghani AC, Edmunds WJ. A metapopulation modeling framework for gonorrhoea and other sexually transmitted infections in heterosexual populations. *J R Soc Interface* 2009;**6**:775–91.

12 Althaus CL, Turner KME, Schmid BV, Heijne JCM, Kretzschmar M, Low N. Transmission of Chlamydia trachomatis through sexual partnerships: a comparison between three

366   individual-based models and empirical data. *J R Soc Interface* 2012;**9**:136–46.

367 13 Fenton KA, Korovessis C, Johnson AM, McCadden A, McManus S, Wellings K, Mercer

368   CH, Carder C, Copas AJ, Nanchahal K, Macdowall W, Ridgway G, Field J, Erens B. Sexual

369   behaviour in Britain: reported sexually transmitted infections and prevalent genital

370   Chlamydia trachomatis infection. *Lancet* 2001;**358**:1851–4. doi:10.1016/S0140-

371   6736(01)06886-6

372 14 Johnson AM, Mercer CH, Beddows S, de Silva N, Desai S, Howell-Jones R, Carder C,

373   Sonnenberg P, Fenton KA, Lowndes C, Soldan K. Epidemiology of, and behavioural risk

374   factors for, sexually transmitted human papillomavirus infection in men and women in

375   Britain. *Sex Transm Infect* 2012;**88**:212–7. doi:10.1136/sextrans-2011-050306

376 15 Sonnenberg P, Clifton S, Beddows S, Field, N, Soldan K, Tanton C, Mercer CH, da Silva

377   FC, Alexander S, Copas AJ, Phelps A, Erens B, Prah P, Macdowall W, Wellings K, Ison

378   CA, Johnson AM. Prevalence, risk factors, and uptake of interventions for sexually

379   transmitted infections in Britain: findings from the National Surveys of Sexual Attitudes and

380   Lifestyles (Natsal). *Lancet* 2013;**382**:1795–806. doi:10.1016/S0140-6736(13)61947-9

381 16 Sonnenberg P, Ison CA, Clifton S, Field N, Tanton C, Soldan K, Beddows S, Alexander S,

382   Khanom R, Saunders P, Copas AJ, Wellings K, Mercer CH, Johnson AM. Epidemiology of

383   Mycoplasma genitaliumin British men and women aged 16–44 years: evidence from the

384   third National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *International Journal of*

385   *Epidemiology* 2015;**44**:1982–94. doi:10.1093/ije/dyv194

386 17 Tanton C, Mesher D, Beddows S, Soldan K, Clifton S, Panwar K, Field N, Mercer CH,

387   Johnson AM, Sonnenberg P. Human papillomavirus (HPV) in young women in Britain:

388   Population-based evidence of the effectiveness of the bivalent immunisation programme and

389   burden of quadrivalent and 9-valent vaccine types. *Papillomavirus Res* 2017;**3**:36–41.

390   doi:10.1016/j.pvr.2017.01.001

391 18 Erens B, Phelps A, Clifton S, Mercer CH, Tanton C, Hussey D, Sonnenberg P, Macdowall

392   W, Field N, Datta J, Mitchell K, Copas AJ, Wellings K, Johnson AM. Methodology of the

393   third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Sexually*

394   *Transmitted Infections* 2013;sextrans–2013–051359. doi:10.1136/sextrans-2013-051359

395 19 Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, Macdowall W, Lewis R,

396   Field N, Datta J, Copas AJ, Phelps A, Wellings K, Johnson AM. Changes in sexual attitudes

397   and lifestyles in Britain through the life course and over time: findings from the National

398   Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet* 2013;**382**:1781–94.

399   doi:10.1016/S0140-6736(13)62035-8

400   20 Johnson AM, Mercer CH, Erens B, Copas AJ, McManus S, Wellings K, Fenton KA,

401   Korovessis C, Macdowall W, Nanchahal K, Purdon S, Field J. Sexual behaviour in Britain:

402   partnerships, practices, and HIV risk behaviours. *Lancet* 2001;**358**:1835–42.

403   doi:10.1016/S0140-6736(01)06883-0

404   21 Brisson M, Bénard É, Drolet M, Bogaards JA, Baussano I, Vänskä S, Jit M, Boily MC,

405   Smith MA, Berkhof J, Canfell K, Chesson HW, Burger EA, Choi YH, Freiesleben De Blasio

406   B, De Vlas SJ, Guzzetta G, Hontelez JAC, Horn J, Jepsen MR, Kim JJ, Lazzarato F,

407   Matthijsse SM, Mikolajczyk R, Pavelyev A, Pillsbury M, Shafer LA, Tully SP, Turner HC,

408   Usher C, Walsh C. Population-level impact, herd immunity, and elimination after human

409   papillomavirus vaccination: A systematic review and meta-analysis of predictions from

410   transmission-dynamic models. *The Lancet Public Health* 2016;**1**:e8–e17.

411   22 Althaus CL, Heijne JC, Herzog SA, Roellin A, Low N. Individual and population level

412   effects of partner notification for Chlamydia trachomatis. *PLoS One* 2012;**7**:e51438.

413   23 Davison AC, Hinkley DV. *Bootstrap Methods and their Application*. 1 edition. Cambridge;

414   New York, NY, USA:: Cambridge University Press 1997.

415   24 Althaus CL, Choisy M, Alizon S. Number of sex acts matters for heterosexual transmission

416   and control of *Chlamydia trachomatis*. *PeerJ PrePrints* 2015;**3**:e1164.

417   doi:10.7287/peerj.preprints.940v1

418   25 Garnett GP, Anderson RM. Factors controlling the spread of HIV in heterosexual

419   communities in developing countries: patterns of mixing between different age and sexual

420   activity classes. *Philos Trans R Soc Lond B Biol Sci* 1993;**342**:137–59.

421   26 Fingerhuth SM, Bonhoeffer S, Low N, Althaus CL. Antibiotic-Resistant Neisseria

422   gonorrhoeae Spread Faster with More Treatment, Not More Sexual Partners. *PLoS Pathog*

423   2016;**12**:e1005611. doi:10.1371/journal.ppat.1005611

424   27 Garnett GP, Mertz KJ, Finelli L, Levine WC, St Louis ME. The transmission dynamics of

425   gonorrhoea: modeling the reported behaviour of infected patients from Newark, New Jersey.

426   *Philos Trans R Soc Lond B Biol Sci* 1999;**354**:787–97.

427   28 R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria::

428   R Foundation for Statistical Computing 2016.

429   29   Chandra NL, Soldan K, Dangerfield C, Sile B, Duffell S, Talebi A, Choi YH, Hughes G,

430   Woodhall SC. Filling in the gaps: estimating numbers of chlamydia tests and diagnoses by

431   age group and sex before and during the implementation of the English National Screening

432   Programme, 2000 to 2012. *Euro Surveill* 2017;**22**. doi:10.2807/1560-

433   7917.ES.2017.22.5.30453

434   30   Althaus CL, Heijne JC, Roellin A, Althaus CL. Transmission dynamics of Chlamydia

435   trachomatis affect the impact of screening programmes. *Epidemics* 2010;**2**:123–31.

436   31   Althaus CL, Heijne JCM, Low N. Towards More Robust Estimates of the Transmissibility of

437   Chlamydia trachomatis. *Sex Transm Dis* 2012;**39**:402–4.

438   32   Brunham RC, Rey-Ladino J. Immunology of chlamydia infection: Implications for a

439   chlamydia trachomatis vaccine. *Nat Rev Immunol* 2005;**5**:149–61. doi:10.1038/nri1551

440   33   Bogaards JA, Xiridou M, Coupé VMH, Meijer CJLM, Wallinga J, Berkhof J. Model-based

441   estimation of viral transmissibility and infection-induced resistance from the age-dependent

442   prevalence of infection for 14 high-risk types of human papillomavirus. *Am J Epidemiol*

443   2010;**171**:817–25. doi:10.1093/aje/kwp466

444   34   Insinga RP, Dasbach EJ, Elbasha, Liaw KL, Barr E. Incidence and duration of cervical

445   human papillomavirus 6, 11, 16, and 18 infections in young women: An evaluation from

446   multiple analytic perspectives. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:709–15.

447   doi:10.1158/1055-9965.EPI-06-0846

448   35   Johnson HC, Elfström KM, Edmunds WJ. Inference of type-specific HPV transmissibility,

449   progression and clearance rates: a mathematical modeling approach. *PLoS One*

450   2012;**7**:e49614. doi:10.1371/journal.pone.0049614

451   36   Price MJ, Ades AE, Angelis DD, Welton NJ, Macleod J, Soldan K, Turner K, Simms I,

452   Horner PJ. Mixture-of-exponentials models to explain heterogeneity in studies of the

453   duration of Chlamydia trachomatis infection. *Stat Med* 2013;**32**:1547–60.

454   doi:10.1002/sim.5603

455   37   Smieszek T, White PJ. Apparently-Different Clearance Rates from Cohort Studies of

456   Mycoplasma genitalium Are Consistent after Accounting for Incidence of Infection,

457   Recurrent Infection, and Study Design. *PLoS One* 2016;**11**:e0149087.

458   doi:10.1371/journal.pone.0149087

Estimated Gini coefficients for different sexually transmitted infections in women.

1   **Table 1: Estimated Gini coefficients for different sexually transmitted infections in women.**

Infection	Gini coefficient	95% confidence interval (CI)
Chlamydia trachomatis	0.33	0.18 - 0.49
Mycoplasma genitalium	0.16	0.02 - 0.36
HPV 6	0.22	0.02 - 0.43
HPV 11	0.17	0.04 - 0.31
HPV 16	0.15	0.04 - 0.27
HPV 18	0.38	0.10 - 0.64

Figure 1

Lorenz curves representing the cumulative proportion of STI infections in women as a function of the cumulative proportion of the population, after population sub-groups have been ranked by the number of new opposite-sex partners in the last year.

(A) Lorenz curves for different STIs. Data: Natsal-3. (B) Uncertainty around Lorenz curve for CT. The blue areas represent point-wise 50% (dark blue) and 95% (light blue) confidence bands. Data: Natsal-3. (C) Comparison of Lorenz curves for CT between Natsal-2 (dashed line) and Natsal-3 (solid line). In all graphs, the diagonal line (black dotted line) denotes perfect equality, i.e., an equal dispersion of the infection across population sub-groups.
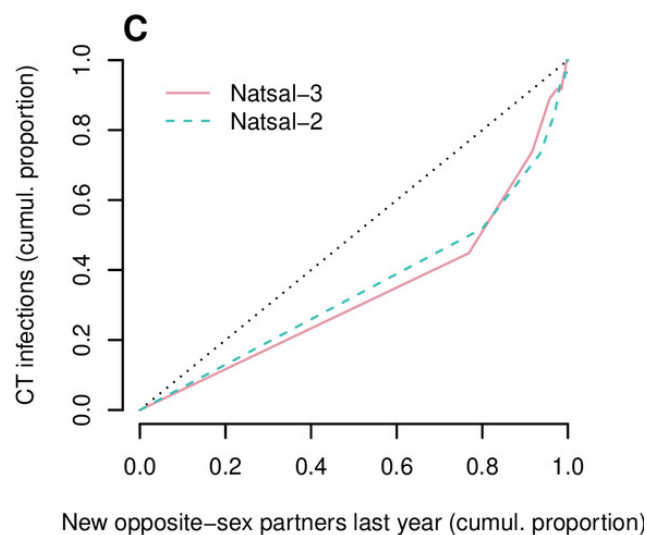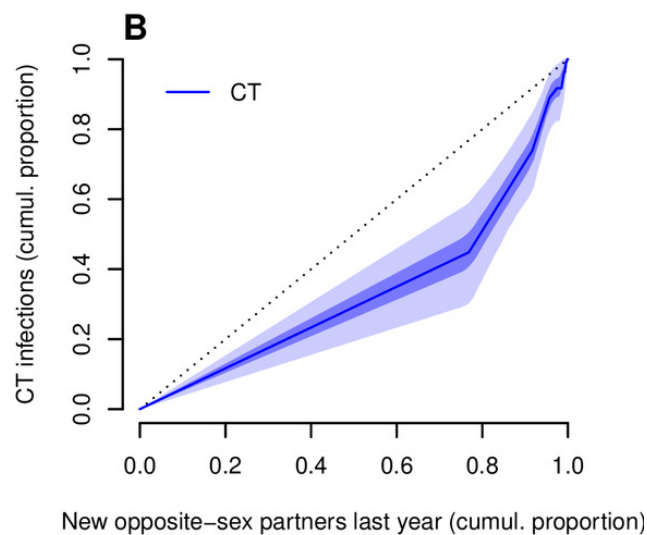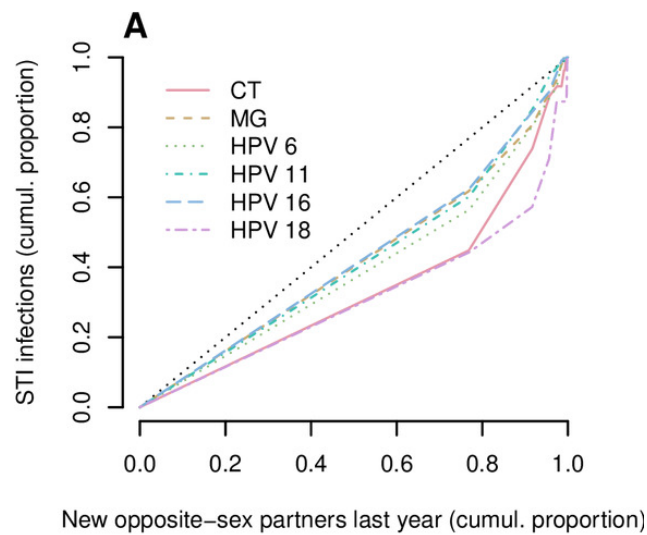
Figure 2

Relationship between Gini coefficient, STI prevalence, infectious duration and transmissibility.

(A) Gini coefficients and STI prevalence for women in Natsal-3 (colored dots). Modelled values for different combinations of the infectious duration and the per partnership transmission probability are projected on the graph (dashed grid). (B) Expected impact of control measures on Gini coefficients and prevalence of female CT between Natsal-2 and Natsal-3. The black arrows denote a 10% reduction in the per partnership transmission probability (horizontal arrows) or the infectious duration (diagonal arrows).