Proportion of Tubal Factor Infertility due to Chlamydia
Ades, AE; Price, Malcolm; Kounali, D.; Akande, V.A.; Wills, G.S.; McClure, M.O.; Muir, P.; Horner, PJ
DOI:
10.1093/aje/kww117
License:
Other (please specify with Rights Statement)

Citation for published version (Harvard):

Publisher Rights Statement:

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
Proportion of Tubal Factor Infertility due to Chlamydia: finite mixture modeling of serum antibody titers.


(*) Correspondence to: Professor AE Ades; School of Social and Community Medicine, 39 Whatley Road, Bristol BS8 2PS, United Kingdom.
Abstract

This study examined whether the proportion of Tubal Factor Infertility (TFI) that is attributable to *Chlamydia trachomatis*, the population excess fraction (PEF), can be estimated from serological data using finite mixture modeling. Whole cell inclusion immune-fluorescence serum antibody titers were recorded in infertile women who were seen at St. Michael's Hospital, Bristol, between 1985-1995 and classified as TFI cases or controls based on laparoscopic examination. Finite mixture models were used to identify the number of component titer distributions and the proportion of samples in each, from which estimates of PEF were derived. Four titer distributions were identified. The component at the highest titer was found only in samples from women with TFI, but there was also an excess of the second highest titer component in TFI cases. Minimum and maximum estimates of the PEF were 28.0% (95% credible interval: 6.9, 50.0) and 46.8% (95% Credible interval: 23.2, 64.1). Equivalent estimates based on the standard PEF formula from case-control studies were 0% and over 65%. Finite mixture modeling can be applied to serological data to obtain estimates of the proportion of reproductive damage attributable to *Chlamydia trachomatis*. Further studies should be undertaken using modern assays in contemporary, representative populations.

<198 words>

Keywords: Tubal Factor Infertility, *Chlamydia trachomatis*, Population Excess Fraction, finite mixture models, antibody titres.
List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>Chlamydia Trachomatis</td>
</tr>
<tr>
<td>CT+</td>
<td>CT infected</td>
</tr>
<tr>
<td>CT-</td>
<td>Not CT infected</td>
</tr>
<tr>
<td>PID</td>
<td>Pelvic Inflammatory Disease</td>
</tr>
<tr>
<td>WIF</td>
<td>Whole cell inclusion immune-fluorescence</td>
</tr>
<tr>
<td>PEF</td>
<td>Population excess fraction</td>
</tr>
<tr>
<td>TFI</td>
<td>Tubal factor infertility</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>CrI</td>
<td>Credible interval</td>
</tr>
</tbody>
</table>
Chlamydia Trachomatis (CT) is a common sexually transmitted infection of young people which if left untreated will cause pelvic inflammatory disease (PID) in around 16% of women (1). PID may then result in adverse reproductive outcomes such as ectopic pregnancy (EP) or tubal factor infertility (TFI) (2). In spite of research using a wide range of study designs, the precise quantitative relationship between *Chlamydia trachomatis* (CT) and reproductive damage remains elusive (2-4). CT, with or without the development of disease, usually resolves spontaneously. Diagnosed infection is treated, so prospective study of untreated infection is not feasible. The major studies of reproductive outcomes in women with PID (5-7) have been restricted to the small proportion of PID (8) that is diagnosed in hospital. In the 1980s and 1990s large numbers of serological case-control studies were carried out, comparing serum antibody levels in women with PID, EP or infertility, with controls (9-19). These studies invariably showed strong associations between detection of CT antibodies and reproductive damage, but it was difficult to draw quantitative conclusions from them, partly because of confounding between CT and other pathogens also implicated in reproductive morbidity (20), and partly because of the poor, and imprecisely known, sensitivity and specificity of the assays used (21, 22).

In this paper we adopt a new analytic approach to this problem: finite mixture modeling (23). Finite mixture models are used when a distribution, in this case of serum antibody titers, is considered to be a mixture of several components, for example “positives” and “negatives”, and where there is an interest in estimating the proportion of samples in diseased and healthy populations in each component. Finite mixture models are often applied to diagnostic tests which lack a “gold standard”, including to serological data (24-28).
In this paper we apply finite mixture models to a previously published dataset (29). Whole cell inclusion immune-fluorescence (WIF) serum antibody titers were recorded in infertile women classified as having Tubal Factor Infertility (cases) or not (controls) following laparoscopy (Table 1). Note that among the cases, a high proportion of the titers that would normally be considered positive (1:32 and above) are at particularly high levels. This has been observed repeatedly in similar studies of TFI (11, 15, 16, 30, 31), and PID (10, 18). In other words, women at higher risk of reproductive damage are more likely to be CT antibody positive, and are more likely to have particularly high titers than antibody-positive controls.

The Lund studies (5-7) established that clinically diagnosed PID was only associated with reproductive damage in women with laparoscopically confirmed salpingitis. Our analysis, therefore, is premised on the assumption that the exceptionally high positive titers seen in TFI cases reflect an inflammatory reaction to CT that is associated with CT-related salpingitis, and that women with titers at these high levels are at risk of CT-related TFI (32). The purpose of the finite mixture analysis is to determine what proportion of the TFI cases have titers at these high levels.

The estimates of the population excess fraction (PEF) formed in this way will be of substantive interest in the many countries where chlamydia and prevention control strategies are in operation, or being considered. However, in view of the limitations inherent in using data collected many years ago for another purpose, this paper should be seen in part as an exploratory, hypothesis forming, exercise into how and whether finite mixture modeling of anti-CT titers can contribute to an understanding of the role of CT in reproductive damage.
METHODS

Data

The primary dataset consisted of WIF titers from 434 TFI cases confirmed on laparoscopy, and 573 controls who were infertile for other reasons, seen in a Reproductive Medicine Clinic, at St. Michael’s hospital in Bristol, between 1985 and 1995. The data were collected in a study exploring the relationship between serum chlamydia antibody titers and detection of tubal damage in infertile women as previously reported (29) (Table 1). Titers of 1:32 or greater would normally be considered positive for CT antibody. Cases had a mean age of 29.3 years (range 18-46) and controls 30.6 years (range 19-47).

A proportion of low titer positives on WIF are likely to be cross-reactions to *Chlamydia pneumonia* (CP) in women with no exposure to CT (33). A secondary dataset provided additional information on the proportion of CT negatives at each WIF titer. Anonymized samples from women undergoing investigation for infertility during 2013 were submitted to Bristol Public Health Laboratories and tested by WIF at Bristol Public Health Laboratories and by the highly specific Pgp3 CT antibody assay (33) at Imperial College. The analyses reported here concern samples from 301 women with WIF titers at or below 1:1024. Causes of infertility and reproductive outcomes were not recorded.

Models

Three models were examined, each characterized by the number of latent distributions assumed to be present (Table 2). For example, the “2-3” model assumed that the control samples were a mixture of two distributions, which we label *CT-* (never infected) and *CT+*...
(previously infected no inflammatory response), while TFI samples may come from either of these distributions or from a third distribution, $CT^{++}$, who have had an inflammatory response to CT infection.

Further models were developed when it was found that the “2-3” model did not fit the data. In the “3-3” model a proportion of control samples is also allowed to belong to the $CT^{++}$ distribution. In the “3-4” model, a further distribution is proposed, $CT^{+++}$, but only TFI samples may belong to it. These labels should be thought of simply as mnemonics, although as the labels suggest, they are listed in order of increasing titer with $CT^{-}$ lowest and representing true CT antibody negatives, and $CT^{+++}$ the highest representing exceptionally high levels of serum antibody.

Statistical methods

Finite Mixture Modeling assumes that each distribution $G$ of titers $y$ is a mixture of say, $D$ underlying latent distributions $f_d(y)$, which we assume are Normal on the log titer scale. We further assume that the only difference between cases ($k=1$) and controls ($k=0$) is in the proportions of samples from each of the latent component distributions, $d=1,\ldots,D$.

$$G_k(y) = \pi_{k1}f_1(y) + \pi_{k2}f_2(y) + \cdots + \pi_{kD}f_D(y)$$

The proportions $\pi_{kd}$ are the proportion of samples that can be attributed to latent distribution $d$, conditional on case / control status $k$. The means and standard deviations of the component distributions remain the same in cases and controls.
Although titers are reported in categories, they are in fact censored observations on a continuous variable. If we designate the lower boundary of categories \(<1:64, 1:64, 1:128, 1:256, 1:512, 1:1024, 1:2048, 1:4096, >1:4096>\) as: \(\hat{l}_{LO} = \{-\infty, 1, 2, 3, 4, 5, 6, 7, 8\}\), and the upper boundaries as \(\hat{l}_{HI} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, \infty\}\), we can express the proportion of distribution \(d\) that falls into the \(i^{th}\) category as:

\[
\theta_{di} = \int_{\hat{l}_{LO}}^{\hat{l}_{HI}} f_d(y) dy
\]

with \(y\) on the natural log titer scale. Finally, the proportions of samples, \(\alpha_{ki}\), in each titer category \(i\) in group \(k\) is obtained as the inner product of the \(\theta_{di}\) and the \(\pi_{kd}\):

\[
\alpha_{ki} = \sum_d \pi_{kd} \theta_{di}
\]

The observed data is the number of samples \(r_{ki}\) in each titer category \(i\) in TFI cases (\(k=1\)) and controls (\(k=0\)); this is multi-nomially distributed, with denominators \(n_k\):

\[
r_{k,i=1...9} \sim \text{Multinomial}(\alpha_{k,1...9}, n_k)
\]

This multinomial distribution is the appropriate choice for the observables i.e. \(n\) trials (samples) with \(k\) possible outcomes (titers) on each trial, and a fixed probability of each outcome over all the trials.

The secondary data source provides additional information on the proportion of true antibody negatives at each WIF titer (Table 1). This was based on the pgp-3 assay, which we assume to be effectively 100% specific (33). The data in Table 1 provides direct information on the
probability $\omega_{ki}$ of a sample being in the CT-group, conditional on its titer and case/control status. This, in turn, constitute indirect information on $\Theta_{1i}$, the probability of a sample having a specified titer, given that it is CT-. We can use Bayes Rule to relate the quantities, with the index “1” indicating the CT-distribution:

$$\omega_{ki} = \frac{\Theta_{1i} \pi_{k1}}{\sum_d \Theta_{di} \pi_{kd}}$$

Because we do not know the proportion of TFI cases in the secondary dataset, we further define a weighted average of $\omega_{ki}$ for cases and $\omega_{0i}$ for controls, with the proportion which are cases $p^{TFI}$ to be estimated from the data:

$$\omega_i = p^{TFI} \frac{\sum_d \Theta_{di} \pi_{1d}}{\sum_d \Theta_{di} \pi_{0d}} + (1 - p^{TFI}) \frac{\sum_d \Theta_{di} \pi_{0d}}{\sum_d \Theta_{di} \pi_{0d}}$$

The secondary data $r_i$ providing information on the $\omega_i$ represent the numbers of pgp-3 negatives among the sample $n_i$ with WIF titre $i$ (Table 1). These have a Binomial likelihood:

$$r_i \sim \text{Bin}(\omega_i, n_i), \quad i = 1, 2, ..., 6$$

The secondary data covers only the first 6 titer categories, as there are no further pgp-3 negatives in higher WIF categories (Table 1). Estimation. We adopt a Bayesian approach and proceed with computations using Markov Chain Monte Carlo (MCMC), supplying weak or non-informative priors for $\mu_d$, $\pi_{kd}$, and weakly informative priors for the standard deviation parameters $\sigma_d$. Mixture models can be unstable, and technical details of priors and constraints
adopted to achieve stability are supplied in the Web Appendix 1 along with WinBUGS code and datasets. Prior specifications are detailed in Web Appendix 2. A series of sensitivity analyses were conducted testing robustness to changes in priors and structural assumptions: these are also described in the Web Appendix 3. Deviance statistics (34), and comparison of observed and predicted counts at each WIF titer, were used to assess model fit and guide model selection. In a Bayesian MCMC framework the residual deviance (35) is calculated on each iteration. The posterior mean residual deviance is a global measure of goodness of fit, and is expected to be close to the number of observations if the model fits well. A reduction in residual deviance of 3 to 5 points is generally taken as evidence of a better fitting model. Convergence was achieved within 10,000-15,000 iterations using Brooks Gelman statistics (36). Posterior results are based on a total 120,000 samples from four chains, having discarded the first 30,000.

*Estimates of population excess fraction.* In the “2-3” model the distribution with the highest mean titer is only found in the TFI cases. The proportion of TFI samples in the $CT^{++}$ distribution in the “2-3” model is therefore a direct estimate of the PEF

$$PEF^{2-3} = \pi_{1,CT^{++}}$$

In the “3-3” model, a more complicated situation arises: here an estimate of the PEF can be based on the excess proportion of samples in the highest $CT^{++}$ category in TFI cases compared to controls. Thus, it is necessary to take into account that a proportion of the $CT^{++}$ observed in the TFI cases would occur anyway, even if TFI status was unrelated to antibody level. Consider
the number of CT++ samples observed in controls, as a proportion of all CT+ and CT++ in the controls. If there was no excess CT++ in the TFI cases we would expect $\pi_{1,CT++}$ to equal:

$$\left(\frac{\pi_{0,CT+}}{\pi_{0,CT+} + \pi_{0,CT++}}\right)^+$$

The estimate of the PEF is the excess CT++ in the TFI group, which is therefore the difference between the observed CT++ in cases, and what we would predict from the controls:

$$PEF^{3-3} = \pi_{1,CT++} - \left(\pi_{1,CT+} + \pi_{1,CT++}\right)^+$$

For the “3-4” model, we can first follow the same logic as the “2-3” model. The proportion of cases in the CT+++ group is a direct estimate of the PEF: $PEF^{3-4(1)} = \pi_{1,CT+++}$, This must be considered as a lower bound estimate because it ignores the excess proportion of CT++ observed in TFI cases.

Acknowledging an excess in CT++ samples in TFI cases, in addition to the CT+++, we can follow the same argument set out for the “3-3” model. This leads to a second estimate of the PEF for “3-4” model, in which we add the proportion in CT+++ to the excess fraction of the CT++:

$$PEF^{3-4(2)} = \pi_{1,CT+++} + \pi_{1,CT++} - \left(\pi_{1,CT+} + \pi_{1,CT++}\right)^+$$

This can be considered as an upper bound estimate as it ascribes all the excess of CT++ in cases to a causal effect of CT infection. The equations for each estimate are presented in Table 2.
extent to which these different estimates are vulnerable to confounding is taken up in the discussion.

Results

Model selection

Table 3 compares the predicted titer distributions from each model with the observed data, and shows the goodness of fit (Residual Deviance) at each point. Systematic error is evident in the “2-3” and “3-3” models, in that they under-estimate the peak that can be seen in both control and TFI samples at 1:1024, while over-estimating the number of samples at 1:512. The “3-4” models on the other hand fit the distribution well at every point.

The fit of models to data is elaborated further in Figure 1 – Figure 4. These plots, depict the fitted component distributions (in color), for each model and separately for cases and controls and are drawn to the correct scale in order to reflect the fitted proportions of each component. The predicted overall titer distribution (solid black line) is the sum of the components, and can be compared to the observed data shown as a histogram.

In a good fitting model the residual deviance should be no more than the number of data points, which is 8 each in the cases and controls. Therefore the global residual deviance statistics at the foot of Table 3 rule out “2-3” and “3-3” models decisively, with the 3-4 models as an excellent fit.

Parameter estimates and PEF
The estimated mean and standard deviation of the each of the log titer distributions, $CT-$, $CT+$, $CT++$, $CT+++$, are set out in Table 4, along with the proportions of cases and controls in each group. The introduction of a $CT++$ distribution for the controls in “3-3” and “3-4” models has the effect of lowering the mean of the $CT+$ distribution by about 1 log unit, and somewhat lowering its variance. Similarly, the introduction of a $CT+++$ category lowers the mean of the $CT++$ group and also reduces its variance.

The secondary data source provides more information on the $CT-$ distribution: the mean is raised to a somewhat higher titer, and the variance is reduced. Its main effect is to reduce uncertainty in the means and SDs of the $CT-$ and $CT+$ distributions. The model fits the secondary data well, with a residual deviance of 4.0 on 6 observations, 2 of which were zeros.

The central estimates of the PEF (Table 4) lie within the range 28%-48%. Although estimates from the “2-3” model can be discounted due its poor fit, it is interesting that it estimates almost exactly the same PEF as the upper bound estimate from the “3-4” model without secondary data. This shows that when the excess cases in the $CT++$ group are assumed to cause TFI one obtains very similar results whether or not one distinguishes between $CT++$ and $CT+++$ distributions.

The secondary data has a slight impact on the estimates of PEF, lowering them by about 4 percentage points. As expected, the probability that a sample in the secondary dataset was from a TFI and not a control was poorly estimated, 0.54 (95% credible interval 0.04, 0.97), barely different from the prior. We consider the estimates from the “3-4” model with secondary data as the best available from the study, due to their greater precision. The
secondary data improves identification of the boundary between the $CT$- and $CT+$ distributions by eliminating the false positives in $CT+$. This can be observed by comparing the posterior means and 95% credible intervals of the means and standard deviations of the $CT$- and $CT+$ distributions (Table 4).

Sensitivity analyses reported in the Web Appendix 3 (Web Table 1, Web Table 2) showed that the main results were robust to reasonable changes in the priors, to distributional assumptions, and to the proportion of TFI cases in the secondary data.

**Discussion**

Estimates of the proportion of pelvic inflammatory disease, ectopic pregnancy, and infertility that can be attributed to Chlamydia are critical to motivating prevention and control programs for *Chlamydia trachomatis* (CT). A number of authors have attempted to derive estimates from serological case-control studies (13, 37), but these are confounded by other exposures that are likely to occur in women exposed to CT, which are also capable of causing reproductive damage (38). Previously, we attempted to derive estimates from a Dutch case-control study (39), taking account of the sensitivity and specificity of assays. That study was based on a form of the “2-3” mixture model, but utilized reported summary data which did not allow titer distributions to be modeled. However, according to its authors, recruitment to the original study was likely to be subject to selection biases (40), and the resulting estimate of 45% (95% CrI: 28, 62) is likely to be an over-estimate. An estimate of 64% in Scotland was described as an upper bound (41). All
estimates of the PEF are, of course, specific to time and place. For public health purposes, estimates should be based on contemporary local data.

This study shows how serum antibody titer distributions from case-control studies can be used to generate estimates of the PEF, based on finite mixture analysis. By attributing the causal mechanisms for TFI to differences between cases and controls in specific components of the titer distributions, rather than to differences in the overall prevalence of antibody, the mixture modeling approach reduces the extent to which PEF estimates are vulnerable to confounding, although it does not eliminate it, as discussed below.

The demonstration that there are four component distributions might appear surprising, rather than the two-component +ves and –ves model that might have been expected. However, the source data used in this exercise (29) has features that were apparent in earlier literature. Histograms suggesting two CT+ve “peaks” in control series have been published previously (11, 12, 16). Similarly, the very high titers seen in women with reproductive damage have also been well-documented for PID / salpingitis (10, 18), and TFI (11, 15, 16, 30, 31). Our analyses suggest the fourth CT+++ component occurs only in TFI cases. We attempted to fit “4-4” models but were unable to achieve stable results. Possibly, evidence for a “4-4” model might be obtained with a larger sample, although our efforts to fit these models suggests that very few controls would be in the CT+++ group.

Interpretation of the different antibody positive groups must be somewhat speculative. Given that salpingitis is a necessary condition for TFI (6, 32, 42), it seems reasonable to regard the CT+++ group, which is observed in cases only, as representing a causal mechanism linked to CT-
related TFI. It is tempting to attribute this to a greater inflammatory response possibly due to a higher infectious load in those who develop TFI, as has been observed at the lower genital tract and in PID (43, 44). The 28.0% (6.9, 50.0) estimate of PEF based on the CT+++ distribution alone can be regarded as a lower bound because it ignores the excess CT++ observed in the cases.

This excess was substantial: 29.4% of TFI cases were in the CT++ group, compared to the 6.5% of the controls (Table 4). We may speculate that the CT++ peak in the women without TFI might represent women who have had upper genital tract infection in whom inflammation has resolved, either following treatment or spontaneously, without causing tubal damage (6, 45), as well as women with recent lower genital tract infections or re-infections, as the decline in antibody over time is far less marked in second infections (46). The excess CT++ seen in cases could simply be due to increased exposure to CT in women whose TFI was in fact caused by other sexually transmitted infections or bacterial vaginosis. Bacterial vaginosis is associated with TFI (47) and with increased likelihood of CT infection and PID (48, 49). Sexual activity may also lead to ascending infection with common respiratory or enteric pathogens that colonize the genital tract, which are also capable of causing reproductive damage (50).

Alternatively, women with TFI are more likely to have been exposed to repeat CT infections, which is associated with both reproductive damage (44, 51) and with higher titers (46). For this reason we may regard the higher PEF estimate (43.0%, 95%CrI: 27.6, 57.5) as an upper bound as it ascribes the entire excess CT++ in cases to a causal mechanism rather than being partly or wholly the result of positive confounding.
The advantage of the mixture model estimates compared to the standard formula for PEF from case-control studies (52) \( PEF = \frac{\pi_{CT}(OR-1)}{\pi_{CT}(OR-1) + 1} \) is that they use the titer distribution as a marker of causal effect. This does not remove vulnerability to confounding, but it does limit it to a proportion of the \( CT^{++} \) distribution, subject of course to our interpretation of the distributions. We can contrast our estimates with those obtained from the standard formula for PEF from case control studies. Using the same WIF data with titres at 1:64 and below as negatives, the Odds Ratio for TFI from Table 1 is \( \frac{380 \times 358}{76 \times 193} = 9.27 \) (95% CI: 6.9, 12.6). A population-based survey of the prevalence of Chlamydia antibody in 16-24 year old women in England, 2007-2010 (53) generated an estimate of 22.9% (95% CrI: 20, 26) in 23-24 year olds. This is most likely an underestimate of \( \pi_{CT} \) in the case-control study because the mean age of women in the WIF data was 30. Applying the formula to these estimates gives a PEF of 65.4%, or 71.3% if \( \pi_{CT} \) is 30%. Both these estimates are upper bounds as they attribute all the excess prevalence to a causal effect; the lower bound is zero, representing the case where CT has no causal role in TFI but exposure to CT is common in those exposed to the true causes.

Nevertheless, both higher and lower estimates generated by the mixture models should be viewed cautiously for two reasons. First, the modeling process was not all pre-planned: each successive model was data-driven, motivated by poor fit in the previous model. The formulae for the PEF estimates were also developed post hoc. Second, our findings are based on an analysis of data collected for a different purpose. Our results therefore need to be confirmed by a specifically planned study, using modern assays, many of which are far more specific (33). The
method could also be extended by testing samples for evidence of other pathogens capable of causing reproductive damage, including *Mycoplasma genitalium* (54), Bacterial vaginosis (47) and possibly *Neisseria gonorrhoeae* (50). The exercise could also be carried out on samples from women with PID and EP. If our results can be confirmed, finite mixture modeling may offer a way of quantifying the role of *Chlamydia* in reproductive damage, and form the basis for monitoring the impact of CT control programs in the population over time.

<3712 words>
Table 1. Numbers of Samples From (a) TFI Cases and Controls Seen at Bristol Between 1985-1995 and (b) From Secondary Anonymized Samples submitted at Bristol Public Health Laboratories during 2013 According to WIF Titer.

<table>
<thead>
<tr>
<th>Titer Group Category</th>
<th>WIF titer</th>
<th>(a) Case-control study</th>
<th>(b) Secondary data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Controls</td>
<td>Cases</td>
</tr>
<tr>
<td>1</td>
<td>&lt;1:64</td>
<td>380</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>1:64</td>
<td>61</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>1:128</td>
<td>45</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>1:256</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>1:512</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td>6</td>
<td>1:1024</td>
<td>30</td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>1:2048</td>
<td>9</td>
<td>69</td>
</tr>
<tr>
<td>8</td>
<td>1:4096</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>&gt;1:4096</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>573</td>
<td>434</td>
</tr>
</tbody>
</table>

Abbreviations: WIF, whole cell inclusion immune-fluorescence, Pgp-3, an immunogenic protein secreted by Chlamydia Trachomatis.
Table 2. Component Distributions in the Alternative Models and Estimators of Population Excess Fraction

<table>
<thead>
<tr>
<th>Model</th>
<th>Controls</th>
<th>Cases</th>
<th>Estimator for Population Excess Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CT-</td>
<td>CT+</td>
<td>CT++</td>
</tr>
<tr>
<td>“2-3”</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>“3-3”</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>“3-4”</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

π₁,CT++

π₁,CT++, π₁,CT++⁺, π₁,CT++⁺⁺, \( \frac{π₀,CT+}{π₀,CT+ + π₀,CT++} \)

Abbreviations: Model “2-3”, the control samples are a mixture of two distributions and cases are a mixture of three distributions; Model “3-3”, the control samples are a mixture of three distributions and cases are a mixture of three distributions; Model “3-4”, the control samples are a mixture of three distributions and cases are a mixture of four distributions; CT-, Not infected Chlamydia Trachomatis; CT+, Chlamydia Trachomatis previously infected but with no immune response; CT++, Chlamydia Trachomatis previously infected with immune response; CT+++, Chlamydia Trachomatis previously infected with exceptionally high levels of serum antibody.

Latent group labels should be thought as mnemonics (see Models in Methods section),

[0]: [Minimum, Maximum]
Table 3. Observed and Predicted Frequency Counts of Each Titer, for Each Model, and Residual Deviance.

<table>
<thead>
<tr>
<th>Group</th>
<th>Titers</th>
<th>Observed</th>
<th>Predicted</th>
<th>Residual Deviance</th>
<th>Predicted</th>
<th>Residual Deviance</th>
<th>Predicted</th>
<th>Residual Deviance</th>
<th>Predicted</th>
<th>Residual Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;1:64</td>
<td>380</td>
<td>378.1</td>
<td>1.0</td>
<td>378.0</td>
<td>1.0</td>
<td>378.3</td>
<td>1.0</td>
<td>377.8</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1:64</td>
<td>61</td>
<td>62.2</td>
<td>1.0</td>
<td>60.9</td>
<td>1.0</td>
<td>60.5</td>
<td>0.9</td>
<td>63.3</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>1:128</td>
<td>45</td>
<td>42.2</td>
<td>0.7</td>
<td>44.9</td>
<td>0.5</td>
<td>45.4</td>
<td>0.5</td>
<td>42.9</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>1:256</td>
<td>28</td>
<td>34.1</td>
<td>1.6</td>
<td>29.7</td>
<td>0.7</td>
<td>29.3</td>
<td>0.6</td>
<td>28.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>1:512</td>
<td>20</td>
<td>26.6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>21.4</td>
<td>0.5</td>
<td>19.6</td>
<td>0.6</td>
<td>20.7</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>1:1024</td>
<td>30</td>
<td>17.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>21.3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>28.8</td>
<td>1.0</td>
<td>28.8</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1:2048</td>
<td>9</td>
<td>8.4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.4</td>
<td>13.1</td>
<td>1.8</td>
<td>10.3</td>
<td>0.8</td>
<td>10.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>1:4096</td>
<td>0</td>
<td>3.2</td>
<td>6.4</td>
<td>3.3</td>
<td>6.6</td>
<td>0.7</td>
<td>1.4</td>
<td>0.8</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>&gt;1:4096</td>
<td>0</td>
<td>1.2</td>
<td>2.4</td>
<td>0.4</td>
<td>0.9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Controls&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Total</td>
<td>573</td>
<td>25.4</td>
<td>17.0</td>
<td>7.1</td>
<td>6.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;1:64</td>
<td>76</td>
<td>77.4</td>
<td>1.0</td>
<td>76.6</td>
<td>1.0</td>
<td>76.4</td>
<td>1.0</td>
<td>76.1</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>1:64</td>
<td>26</td>
<td>25.6</td>
<td>0.5</td>
<td>29.5</td>
<td>1.0</td>
<td>29.0</td>
<td>1.0</td>
<td>28.5</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>1:128</td>
<td>34</td>
<td>29.8</td>
<td>1.1</td>
<td>30.5</td>
<td>1.0</td>
<td>31.8</td>
<td>0.7</td>
<td>32.2</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>1:256</td>
<td>33</td>
<td>35.5</td>
<td>0.6</td>
<td>31.0</td>
<td>0.8</td>
<td>31.2</td>
<td>0.9</td>
<td>31.8</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>1:512</td>
<td>48</td>
<td>57.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>59.7&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>52.7</td>
<td>1.4</td>
<td>51.8</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>1:1024</td>
<td>122</td>
<td>105.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>108.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>118.4</td>
<td>1.1</td>
<td>118.5</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>1:2048</td>
<td>69</td>
<td>80.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>77.1</td>
<td>1.6</td>
<td>69.3</td>
<td>0.7</td>
<td>69.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>1:4096</td>
<td>22</td>
<td>20.0</td>
<td>1.2</td>
<td>19.3</td>
<td>1.1</td>
<td>20.1</td>
<td>1.0</td>
<td>20.5</td>
<td>0.9</td>
</tr>
<tr>
<td>TFi&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&gt;1:4096</td>
<td>4</td>
<td>2.7</td>
<td>1.1</td>
<td>2.2</td>
<td>2.2</td>
<td>5.2</td>
<td>1.1</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>434</td>
<td>14.7</td>
<td>15.1</td>
<td>8.8</td>
<td>7.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grand Total</td>
<td>39.2</td>
<td>32.1</td>
<td>15.9</td>
<td>14.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: TFi, tubal factor infertility; Model “2-3”, the control samples are a mixture of two distributions and cases are a mixture of three distributions; Model “3-3”, the control samples are a mixture of three distributions and cases are a mixture of three distributions; Model “3-4”, the control samples are a mixture of three distributions and cases are a mixture of four distribution.

<sup>a</sup> Anonymised samples submitted for infertility investigations at the Bristol Public Health Laboratory during 2013

<sup>b</sup> Women undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael’s hospital, Bristol, during 1985-1995.

<sup>c</sup> Poorly fitting observations.
Table 4. Posterior Summaries From the 4 Models: Mean and Standard Deviation of the log Titers, and Percent in Each Component in the Controls, \( \pi_{0d} \), and TFI Cases \( \pi_{1d} \) along with Estimates of the Population Excess Fraction.

<table>
<thead>
<tr>
<th>Model “2-3”</th>
<th>Model “3-3”</th>
<th>Model “3-4”</th>
<th>Model “3-4” with secondary data&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group Means ( \mu_i )</strong></td>
<td>Mean</td>
<td>95% CrI</td>
<td>Mean</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>-0.61</td>
<td>-1.6, -0.00</td>
<td>-1.09</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>3.48</td>
<td>2.3, 4.5</td>
<td>2.41</td>
</tr>
<tr>
<td>( \mu_3 )</td>
<td>5.84</td>
<td>5.7, 6.0</td>
<td>5.72</td>
</tr>
<tr>
<td>( \mu_4 )</td>
<td>-1.69</td>
<td>-2.41</td>
<td>5.72</td>
</tr>
</tbody>
</table>

| **Group Standard deviations \( \sigma_i \)** | 1.55 | 0.8, 2.2 | 1.54 | 0.51, 2.3 | 1.48 | 0.48, 2.2 | 1.60 | 1.2, 2.0 |
| \( \sigma_2 \) | 1.87 | 1.5, 2.3 | 1.66 | 1.0, 2.2 | 1.55 | 1.0, 2.1 | 1.54 | 1.2, 1.9 |
| \( \sigma_3 \) | 0.83 | 0.6, 1.0 | 0.90 | 0.60, 1.1 | 0.56 | 0.3, 0.8 | 0.56 | 0.3, 0.8 |
| \( \sigma_4 \) | 1.19 | 0.7, 1.9 | 1.11 | 0.7, 1.8 | 1.11 | 0.7, 1.8 |

| **Mixing proportions: Controls<sup>b</sup>** | 74.5 | 57, 86 | 67.9 | 45, 91 | 66.5 | 41, 87 | 74.0 | 66, 83 |
| \( \pi_{01} \) | 25.5 | 14, 43 | 25.4 | 18, 48 | 26.6 | 5.8, 52 | 19.4 | 10, 28 |
| \( \pi_{02} \) | 6.65 | 2.6, 11 | 6.92 | 3.7, 10 | 6.54 | 3.3, 10 |

| **Mixing proportions: Cases<sup>b</sup>** | 16.8 | 6.3, 26 | 12.7 | 1.3, 25 | 13.1 | 1.4, 24 | 16.4 | 9.7, 22 |
| \( \pi_{11} \) | 35.5 | 25, 47 | 30.0 | 17, 44 | 26.9 | 15, 40 | 26.3 | 15, 37 |
| \( \pi_{12} \) | 47.8 | 34, 58 | 57.3 | 35, 66 | 28.3 | 8.7, 50 | 29.4 | 9.5, 50 |
| \( \pi_{13} \) | - | - | - | - | 31.7 | 8.7, 53 | 28.0 | 6.9, 50 |
| \( \pi_{14} \) | 47.7 | 34.2, 57.8 | 35.9 | -21.3, 52.7 | 31.7<sup>c</sup> | 8.7, 52.8 | 28.0<sup>a</sup> | 6.9, 50.0 |

| Population Excess Fraction | 46.8<sup>d</sup> | 23.2, 64.1 | 43.0<sup>b</sup> | 27.6, 57.5 |

Abbreviations: TFI, Tubal factor infertility; CrI, Credible interval; Model “2-3”, the control samples are a mixture of two distributions and cases are a mixture of three distributions; Model “3-3”, the control samples are a mixture of three distributions and cases are a mixture of three distributions; Model “3-4”, the control samples are a mixture of three distributions and cases are a mixture of four distributions.

<sup>a</sup> Anonymised samples submitted for infertility investigations at the Bristol Public Health Laboratory during 2013

<sup>b</sup> Women undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael’s hospital, Bristol, during 1985-1995.

<sup>c</sup> Lower bound/minimum estimate for the population excess fraction

<sup>d</sup> Upper bound/maximum estimate for the population excess fraction
Acknowledgements

Author affiliations: School of Social and Community Medicine, University of Bristol, Bristol, UK (AE. Ades, D. Kounali, PJ. Homer); Bristol Sexual Health Centre, University Hospitals Bristol NHS Trust, Bristol, UK (PJ. Homer); Institute of Applied Health Research, University of Birmingham, Birmingham, UK (MJ. Price); Jefferiss Trust Laboratories, Wright-Fleming Institute, Imperial College London, London, UK (GS. Wills, MO., McClure); Public Health Laboratory Bristol, National Infection Service, Public Health England, Myrtle Road, Bristol, UK. (P. Muir); School of Clinical Sciences, University of Bristol and Bristol Centre for Reproductive Medicine, North Bristol NHS Trust, Bristol, UK (VA. Akande).

This work was supported by the NIHR Health Protection Research Unit in Evaluation of Interventions at the University of Bristol under award number HPRU-2012-10026 and in partnership with Public Health England (PHE). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Conflict of interest: none declared.
Figure 1. Fitted component distributions for controls (Panel A) and cases (Panel B) undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael's hospital, Bristol, during 1985-1995, based on the model assuming two component distributions for cases and three for controls\(^a\).

Figure 2. Fitted component distributions for controls (Panel A) and cases (Panel B) undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael’s hospital, Bristol, during 1985-1995, based on the model assuming three component distributions for cases and three for controls\(^a\).

Figure 3. Fitted component distributions for controls (Panel A) and cases (Panel B) undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael’s hospital, Bristol, during 1985-1995, based on the model assuming three component distributions for cases and four for controls\(^a\).

Figure 4. Fitted component distributions for Controls (Panel A) and cases (Panel B) undergoing infertility investigations at the Reproductive Medicine Clinic, at St. Michael’s hospital, Bristol, during 1985-1995, based on the model assuming three component distributions for cases and four for controls and which also made use of secondary data on Anonymised samples submitted for infertility investigations at the Bristol Public Health Laboratory during 2013.

\(^a\) The figure is drawn to a scale reflecting the mixing proportions, the predicted overall titer distribution (solid black line), and the observed data (dashed line histogram). The left-most histogram bar comprises titers below 1:64. This has been plotted to cover the area -5 to +1 on the log titer scale; its area corresponds to the proportion of data at these titers.
References


