Epidemiol. Infect. (2016), **144**, 887–895. © Cambridge University Press 2015
doi:10.1017/S0950268815001958

Is it appropriate to use fixed assay cut-offs for estimating seroprevalence?

G. KAFATOS[1,2]*, N. J. ANDREWS[1], K. J. McCONWAY[2], P. A. C. MAPLE[3], K. BROWN[3] AND C. P. FARRINGTON[2]

[1] *Department of Statistics, Modelling and Economics, Public Health England, London, UK*
[2] *Department of Mathematics and Statistics, The Open University, Milton Keynes, UK*
[3] *Virus Reference Department, Public Health England, London, UK*

Received 12 June 2015; Final revision 3 August 2015; Accepted 4 August 2015; first published online 27 August 2015

SUMMARY

Population seroprevalence can be estimated from serosurveys by classifying quantitative measurements into positives (past infection/vaccinated) or negatives (susceptible) according to a fixed assay cut-off. The choice of assay cut-offs has a direct impact on seroprevalence estimates. A time-resolved fluorescence immunoassay (TRFIA) was used to test exposure to human parvovirus 4 (HP4). Seroprevalence estimates were obtained after applying the diagnostic assay cut-off under different scenarios using simulations. Alternative methods for estimating assay cut-offs were proposed based on mixture modelling with component distributions for the past infection/vaccinated and susceptible populations. Seroprevalence estimates were compared to those obtained directly from the data using mixture models. Simulation results showed that when there was good distinction between the underlying populations all methods gave seroprevalence estimates close to the true one. For high overlap between the underlying components, the diagnostic assay cut-off generally gave the most biased estimates. However, the mixture model methods also gave biased estimates which were a result of poor model fit. In conclusion, fixed cut-offs often produce biased estimates but they also have advantages compared to other methods such as mixture models. The bias can be reduced by using assay cut-offs estimated specifically for seroprevalence studies.

Key words: Assay, cut-off, mixture models, serology, seroprevalence.

INTRODUCTION

Seroepidemiology can be used to assess the burden of infectious diseases and to optimize immunization strategies for vaccine-preventable diseases. This can be achieved by carrying out serological surveys aimed at identifying susceptible cohorts that can subsequently be targeted by public health interventions. Estimating seroprevalence (i.e. the percentage of population that shows markers of having the disease) as opposed to simply prevalence (based on clinical symptoms) has the advantage of more accurate definition of those infected that also includes those vaccinated and with asymptomatic infections [1, 2].

Once serum samples have been collected they can be tested to determine susceptibility to the infection. The amount of antibody present in a sample can be measured in the laboratory using a serological assay that returns a quantitative measurement. Each sample

* Author for correspondence: Dr G. Kafatos, Department of Statistics, Modelling and Economics, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK.
(Email: gkafatos1@hotmail.com)

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

is classified into positive (recent/past infections or vaccinated) or negative (susceptibles) according to a predefined cut-off. Positive measurements close to the cut-off are often further classified into equivocals (low positives); however, for simplicity this will not be considered here [2]. The proportion of positives over the total number of samples included in the serosurvey equals the seroprevalence estimate.

The method described above is used for routine diagnostic testing but it is also used to classify serosurvey samples into susceptibles and non-susceptibles. Therefore, the choice of assay cut-off is crucial since it has a direct impact on seroprevalence estimation [3, 4]. There are different methods for estimating such cut-offs, and well-conducted seroepidemiological studies are expected to include details of how these were calculated [1]. However, regardless of the method used, the optimum value for a cut-off depends on the aims of the study as well as the seroprevalence and type of the infection. For example, an assay used to test prior history of chickenpox in pregnant women needs to be highly specific to ensure that all true negatives are classified as negatives whereas a test for screening population for HIV needs to be highly sensitive to make sure it flags all potential cases. However, the seroprevalence of the disease is also important since for a given sensitivity and specificity, the performance of the test varies for different proportions of positives [i.e. the change in positive predictive value (PPV) and negative predictive value (NPV)]. Finally, the type of the disease is also important, e.g. an HIV assay needs to be highly sensitive to make sure the true positives are tested as positive but also specific (to avoid falsely informing individuals that are positives), whereas the latter might not be so important for a bacterial infection [2].

An alternative method to fixed cut-offs for estimating population seroprevalence is mixture modelling. Mixture models can be used to estimate seroprevalence directly from the quantitative serosurvey data, rather than initially classifying individual samples into positive or negative according to a pre-defined cut-off. This can be achieved by fitting a mixture of distributions and obtaining the proportion of samples belonging to each underlying component [5]. Mixture models have been used in the past to estimate population seroprevalence of various infectious diseases [3, 5–7]. One of the assumptions required for the mixture mode to work is to have reasonably high proportions of data in each underlying component [2]. This was not the case for the blood donor serosurvey presented here where samples were tested for human parvovirus 4 (HP4) as that infection is particularly associated with people who inject drugs (PWID). As an alternative, we propose using serological assay cut-offs to distinguish between HP4 positive and negative samples.

Estimating seroprevalence using assay cut-offs has limitations compared to mixture models; however, it is still widely used mainly because of convenience (once the assay cut-off has been set, it can be used in different serosurveys). The aim of this paper is to compare existing and proposed cut-off estimates and assess their impact on estimating seroprevalence.

METHODS

HP4 dataset

A time-resolved fluorescence immunoassay (TRFIA) developed to test exposure to HP4 was used as an example. In brief, a HP4 IgG TRFIA was developed using a panel of 184 sera from PWID that ensured a high number of positive samples. The aim was to estimate an assay cut-off that could be subsequently applied to a serosurvey of 608 blood donors' sera collected in 1999 and 2009. More information about how the individuals were selected and how samples were tested is given elsewhere [8].

Mixture modelling

A simple mixture model was considered for the PWID dataset. Assuming two underlying populations of positive (past/current infections) and negative samples (individuals with no past infection), a two-component mixture model can be fitted of the type:

$$g(x) = (1-p)f_-(x;\theta_-) + pf_+(x;\theta_+),$$

where f_- and f_+ are the density functions for the negative and positive components, respectively, and θ_- and θ_+ are the parameters defining the shapes of the underlying distributions. For a simple model of this form, the proportion of the sample in the positive underlying component $\hat{p}$ can be used as an estimate of seroprevalence whereas the proportion of susceptibles is given by $1-\hat{p}$ [9].

Cut-off estimation

There are no definite rules for the choice of cut-off point, and it must depend on the reason for performing the test. Reports from good serological studies should always include an account of how the cut-off

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

point was set [1]. A number of methods can be used for setting an assay cut-off depending on the information available when developing the assay. Assuming that the true status of the samples can be estimated by another method (e.g. by an established assay that can serve as gold standard or by questionnaires recording any past infections) then the quantitative assay results can be plotted against the true status using a receiver-operating characteristic (ROC) curve. The optimal combination of sensitivity and specificity can then be chosen [10, 11].

Method 1

Often an estimate of the true status is not available. Then a commonly used cut-off is the mean of the log-transformed titres of a group of samples that are known to be negative plus 2, 3 or 4 standard deviations (S.D.) [2, 11–14]. This is the most commonly used method by the assay manufacturers due to its simplicity [14]. It produces tests with high specificity, disregarding the effect this may have on sensitivity. In terms of seroprevalence, this may lead to some truly positive samples being classified as susceptible [2, 3, 6]. In this paper, the mean of negative samples +2 S.D. (giving assay specificity of 97·5% assuming a normal distribution) will be examined as a commonly used method for estimating the assay cut-off.

Alternatively, once a mixture model has been fitted to the development panel results, a cut-off can be estimated based on the model's parameter estimates [11, 13]. Three methods for cut-off estimation are described here:

Method 2

Assume that from a mixture model, $\hat{p}$ is the seroprevalence estimate and $\hat{\theta}_-$ and $\hat{\theta}_+$ are the parameter estimates for the negative and positive underlying component distributions, respectively. For a given cut-off C, the proportions of falsely classified positive samples (false negatives) can be defined as $F_+(C; \hat{\theta}_+)$ and the false positive as $1 - F_-(C; \hat{\theta}_-)$, where f_- and f_+ are cumulative distributions of the density functions f_- and f_+, respectively, as defined above. One way is to estimate a cut-off that minimizes the total misclassification, i.e. maximizes the following fraction:

$$\frac{1}{(1-\hat{p})(1-F_-(C;\hat{\theta}_-)) + \hat{p}F_+(C;\hat{\theta}_+)}.$$

This is a method that has been proposed and used in the past [13].

Method 3

An alternative cut-off proposed here is to set the number of false-negative samples to be equal to the number of false-positive samples. This occurs when

$$(1-\hat{p})(1-F_-(C;\hat{\theta}_-)) = \hat{p}F_+(C;\hat{\theta}_+),$$

which can be obtained by solving:

$$\left|(1-\hat{p})(1-F_-(C;\hat{\theta}_-)) - \hat{p}F_+(C;\hat{\theta}_+)\right| = 0.$$

This method's cut-off should provide unbiased seroprevalence estimates for a population with the same characteristics.

Method 4

Another cut-off proposed here is to set the number of false-negative samples relative to the negative samples correctly classified to be equal to the number of false positives relative to the number of positive samples correctly classified. This is given by:

$$\frac{(1-\hat{p})(1-F_-(C;\hat{\theta}_-))}{(1-\hat{p})F_-(C;\hat{\theta}_-)} = \frac{\hat{p}F_+(C;\hat{\theta}_+)}{\hat{p}(1-F_+(C;\hat{\theta}_+))},$$

which can be obtained by solving:

$$\left|\frac{1-F_-(C;\hat{\theta}_-)}{F_-(C;\hat{\theta}_-)} - \frac{F_+(C;\hat{\theta}_+)}{1-F_+(C;\hat{\theta}_+)}\right| = 0.$$

The cut-off estimate from method 4 has the advantage of not being affected by the proportion of samples belonging in each underlying component.

Comparison of the different methods using simulations

Simulations were generated aiming to compare seroprevalence estimates using the different methods. This was done in two steps. The objective of the first step was to estimate the cut-offs. For this, simulations were generated from a dataset based on the PWID data. The process was as follows:

(1) Location (mean μ_- for the negative underlying component and μ_+ for the positive component), dispersion (standard deviation σ_- for the negative component and σ_+ for the positive component), and seroprevalence (p) parameters were set to define a mixture model that data can be simulated from. Four scenarios were chosen with different location parameters for the positive underlying

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

component while all the other parameters in the models remained fixed.

(2) One thousand simulated datasets ($i = 1, \ldots, 1000$) were generated from each of the four mixture models using the following method. For each simulation, $184p$ samples were assumed to have been sampled from the positive (recent or past infection/vaccinated) population and the rest 184 $(1-p)$ samples from the susceptible population. Samples assumed to belong to the positive population were simulated from $N(\mu_+, \sigma_+)$ whereas samples belonging to the susceptible population were simulated from $N(\mu_-, \sigma_-)$.

(3) Following each simulation a mixture model was fitted. Substituting the estimates $\hat{\mu}_{-i}$, $\hat{\sigma}_{-i}$, $\hat{\mu}_{+i}$, $\hat{\sigma}_{+i}$ and $\hat{p}_i$, cut-offs $\hat{C}_{hi}$ (h denotes the four cut-off estimates) were estimated using the methods described above. The mean of the cut-off estimates $\bar{C}_h$ was calculated over the 1000 simulated datasets. The 95% percentile intervals were given as a measure of uncertainty.

The objective of the second step was to apply the estimated cut-offs to the serosurvey, thus obtaining seroprevalence. Specifically, each cut-off was applied to a simulated population of 608 samples (based on the blood donor serosurvey) with varying proportions of positive population. The simulation methodology for this population is described below.

(1) The same location and dispersion parameters that were used to generate the original datasets were selected for the four scenarios. For each model a number of seroprevalence parameters p were chosen ranging from 5% to 50%. One thousand simulations of 608 samples were generated for each of the models and scenarios in the same way as before.

(2) The estimated cut-offs $\bar{C}_h$ were applied to the simulated datasets and hence, a seroprevalence estimate was derived for each simulation by grouping the samples into positive and susceptible. The mean of the seroprevalence estimates was calculated over 1000 simulations. The 95% percentile interval was given as a measure of uncertainty.

The analysis was carried out using Stata statistical software release 12.1 (StataCorp, USA). For the cut-off estimation, to find the root of the continuous monotonic functions, the OPTIMIZE command was used in R version 2.10.1 (R Foundation for Statistical Computing, Vienna, Austria,).

RESULTS

Seroprevalence estimation for PWID and UK blood donors

Following a log transformation, the distribution of the assay development PWID sera and the fitted line of the mixture model are given in Figure 1*a*. The model estimates for the negative component were $\hat{\mu}_- = 4.6$, $\hat{\sigma}_- = 0.3$, and for the positive component were $\hat{\mu}_+ = 6$, $\hat{\sigma}_+ = 0.2$ and the seroprevalence estimate for PWID was 20·6% ($\hat{p} = 0.206$). Based on these model estimates, cut-offs were estimated using different methods. After applying these cut-offs to the UK blood donors' sera resulted in the following seroprevalence estimates: 6·9% [95% uncertainty interval (UI) 5·0–9·2] for method 1, 2·5% (95% UI 1·4–4·0) for method 2, 2·5% (95% UI 1·4–4·0) for method 3 and 2·8% (95% UI 1·6–4·4) for method 4.

Cut-off estimation for simulation scenarios

Based on the estimates derived from PWID sera, four simulation scenarios were defined in Table 1. The underlying components of the mixture distributions together with the results of one simulation are displayed in Figure 2.

For each simulation, cut-offs were estimated using the methods described above. The cut-off estimates averaged over the number of simulations together with their corresponding 95% UIs are given in Table 2. The cut-off estimate using the mean of negative samples +2 S.D. (cut-off 5·2, 95% UI 5·1–5·3) did not vary between the different scenarios. This is because it is solely based on the shape of the underlying distribution of the susceptible individuals that remains fixed across the different scenarios. The cut-off estimates of the three other cut-off methods varied little with the method 3 cut-off being slightly lower than the others. As the location parameter of the positive component was reduced between scenario 1 and scenario 4 the cut-off estimates also became smaller.

Seroprevalence estimation for simulation scenarios

The estimated cut-offs were used to calculate seroprevalence from simulated datasets based on the serosurvey of 608 blood donors. The distribution of the blood donors' serological results following a $\log_{10}$ transformation are given in Figure 1*b*. The estimated seroprevalence together with the corresponding 95% UI for each cut-off method and scenarios are given

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

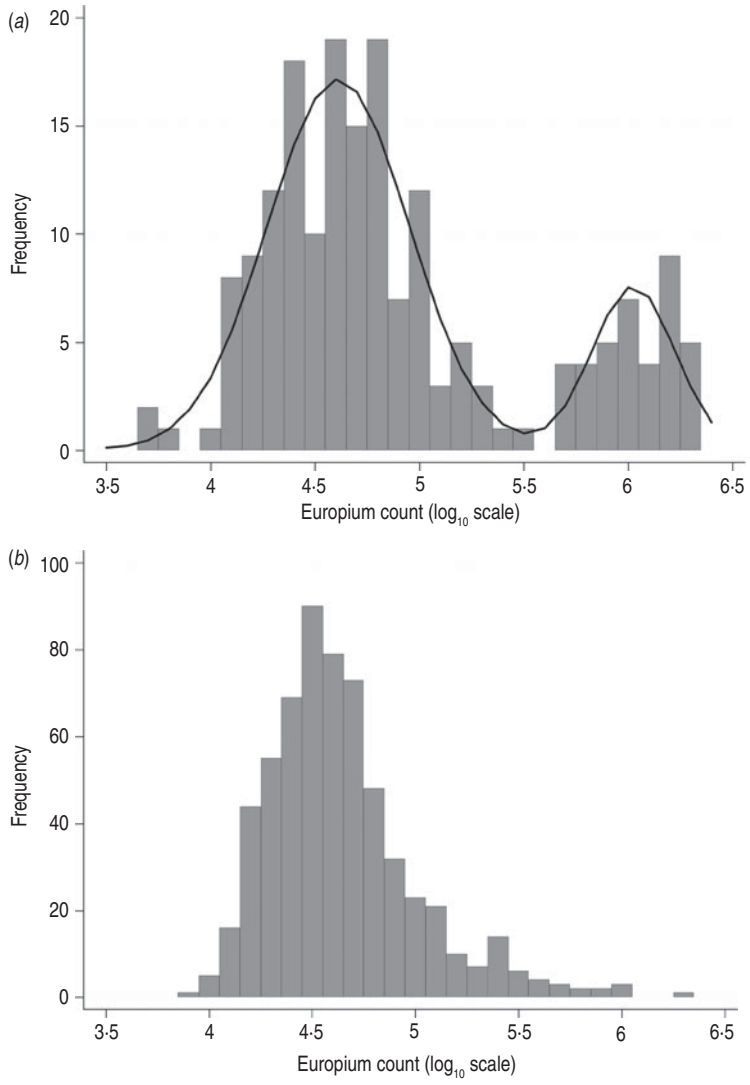

Fig. 1. Distributions of HP4 Europium counts. (*a*) Sera collected from 184 people who inject drugs. (*b*) Sera collected from 608 UK blood donors.

Table 1. *Definition of four mixture model scenarios based on people who inject drugs*

Scenario	μ_- (σ_-)	μ_+ (σ_+)	P
1	4·6 (0·3)	6·0 (0·2)	0·206
2	4·6 (0·3)	5·7 (0·2)	0·206
3	4·6 (0·3)	5·4 (0·2)	0·206
4	4·6 (0·3)	5·1 (0·2)	0·206

in Table 3 generated from models with true seroprevalence of 5%, 15% and 25%. In addition to the estimates obtained by the cut-off methods, seroprevalence was estimated by directly fitting mixture models (proportion of samples belonging to the positive underlying component).

When the true seroprevalence was 15% or 25% (i.e. close to the 20·6% estimated from the development PWID serosurvey) all cut-off estimation methods gave estimates close to the true one for scenarios 1–3. For higher overlap between the underlying components (scenario 4) the estimates deviated from the true value with method 1 giving the most biased estimate. For true seroprevalence of 5%, all cut-off estimation methods gave estimates close to the true one for scenarios 1 and 2, whereas for scenario 3 method 4 gave the most biased estimate ($\hat{p} = 11\%$, 95% UI 5–27). For scenario 4, all methods gave seroprevalence much higher than 5% except for method 1 ($\hat{p} =$ 4%, 95% UI 1–7). It should be noted that about 1% of mixture models failed to converge for scenario 4 and were excluded from further analysis.

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

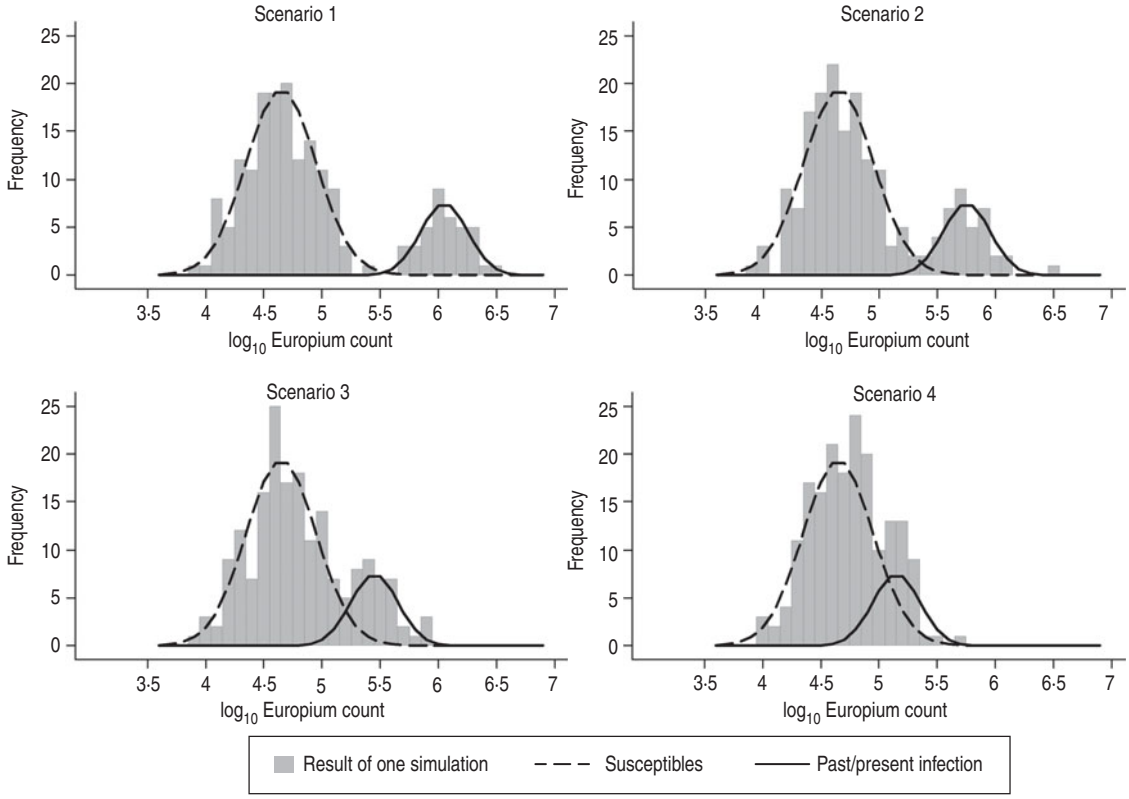

Fig. 2. Mixture model distributions by scenario based on the dataset of 184 people who inject drugs.

Table 2. *Cut-off estimates and 95% uncertainty intervals following 1000 simulations*

Scenario	Method 1	Method 2	Method 3	Method 4
1	5·20 (5·12–5·29)	5·48 (5·48–5·48)	5·50 (5·38–5·61)	5·45 (5·32–5·57)
2	5·20 (5·11–5·29)	5·31 (5·30–5·32)	5·33 (5·19–5·46)	5·27 (5·13–5·40)
3	5·20 (5·12–5·28)	5·15 (5·09–5·22)	5·16 (4·86–5·36)	5·08 (4·82–5·28)
4	5·20 (5·12–5·28)	4·99 (4·65–5·32)	4·96 (4·50–5·40)	4·89 (4·50–5·30)

Table 3. *Seroprevalence estimates and 95% uncertainty intervals following 1000 simulations for each scenario*

Scenario	Mixture model	Method 1	Method 2	Method 3	Method 4
		(*a*) Seroprevalence for $P = 0{\cdot}25$			
1	0·25 (0·25–0·26)	0·27 (0·25–0·29)	0·25 (0·25–0·25)	0·25 (0·24–0·25)	0·25 (0·25–0·26)
2	0·25 (0·24–0·26)	0·27 (0·25–0·28)	0·25 (0·24–0·26)	0·25 (0·22–0·27)	0·26 (0·23–0·28)
3	0·25 (0·20–0·31)	0·23 (0·18–0·27)	0·25 (0·21–0·28)	0·25 (0·14–0·39)	0·28 (0·19–0·42)
4	0·28 (0·07–0·69)	0·10 (0·05–0·15)	0·24 (0·00–0·57)	0·30 (0·02–0·74)	0·34 (0·05–0·73)
		(*b*) Seroprevalence for $P = 0{\cdot}15$			
1	0·15 (0·15–0·15)	0·17 (0·16–0·19)	0·15 (0·15–0·15)	0·15 (0·15–0·16)	0·15 (0·15–0·16)
2	0·15 (0·14–0·16)	0·17 (0·15–0·19)	0·15 (0·14–0·16)	0·15 (0·13–0·17)	0·16 (0·14–0·19)
3	0·15 (0·11–0·22)	0·15 (0·11–0·18)	0·16 (0·13–0·19)	0·16 (0·08–0·31)	0·19 (0·12–0·34)
4	0·21 (0·02–0·79)	0·07 (0·03–0·11)	0·19 (0·00–0·52)	0·26 (0·01–0·70)	0·30 (0·03–0·69)
		(*c*) Seroprevalence for $P = 0{\cdot}05$			
1	0·05 (0·05–0·05)	0·07 (0·06–0·09)	0·05 (0·05–0·05)	0·05 (0·05–0·06)	0·05 (0·05–0·06)
2	0·05 (0·04–0·07)	0·07 (0·05–0·09)	0·06 (0·05–0·06)	0·06 (0·04–0·07)	0·06 (0·05–0·09)
3	0·06 (0·02–0·25)	0·06 (0·04–0·09)	0·08 (0·05–0·10)	0·08 (0·03–0·23)	0·11 (0·05–0·27)
4	0·15 (0·00–0·90)	0·04 (0·01–0·07)	0·14 (0·00–0·46)	0·21 (0·01–0·66)	0·24 (0·02–0·65)

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

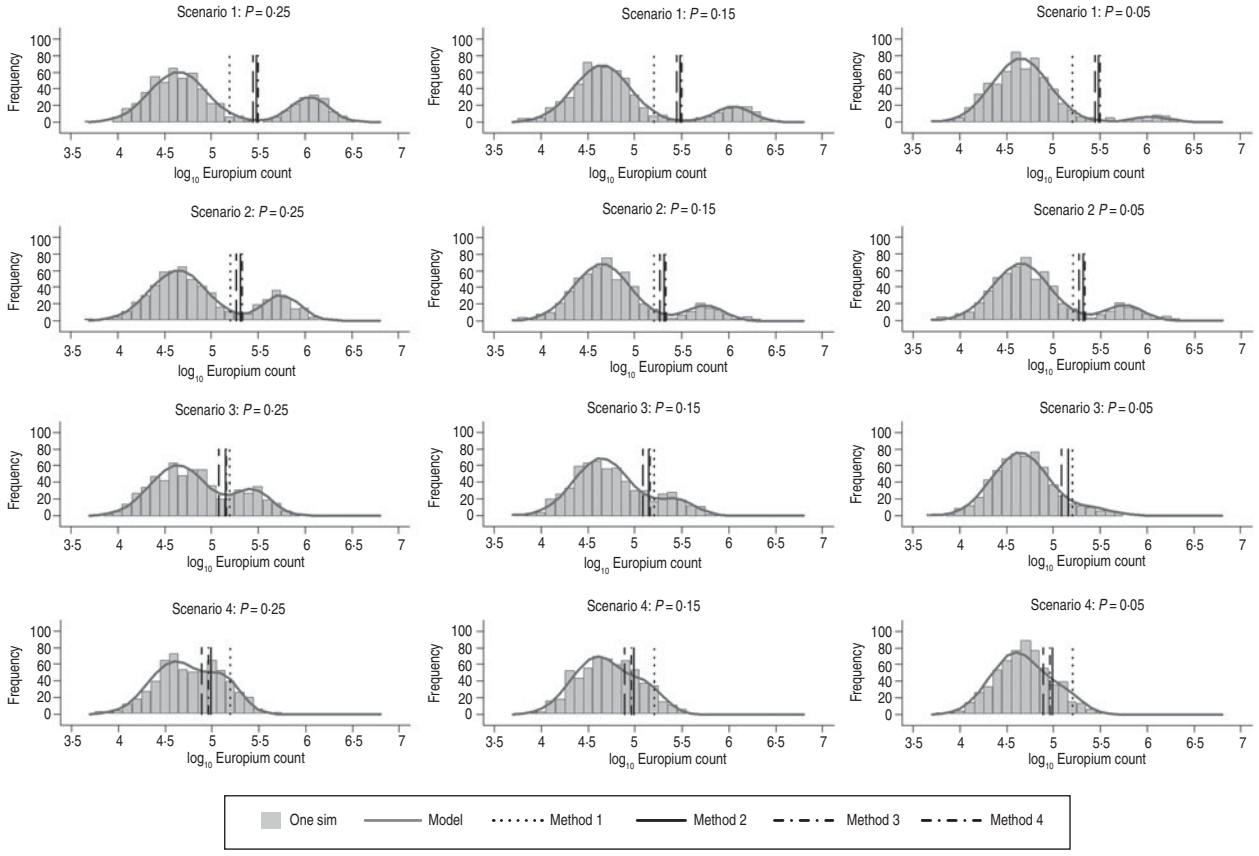

Fig. 3. Distributions of simulated serosurveys of 608 samples and estimated cut-offs for different scenarios.

The mixture model distributions are presented in Figure 3 together with the distribution of one simulated dataset of blood donor results. The cut-off estimates by the four methods used are also displayed. Figure 4 shows the seroprevalence estimates by each scenario and method for a range of seroprevalences (5–50%). All methods seem to work well for scenarios 1 and 2 where there is little overlapping between the negative and positive components. For scenario 3 there are some deviations with the mixture model estimates being closer to the true one followed by methods 2, 3, and 4.

DISCUSSION

Fixed cut-offs are often used in serological studies to estimate seroprevalence [2]. This is preferred by many scientists because it is easier to apply compared to say, a mixture modelling approach. Moreover, once the cut-offs have been estimated, they can be easily applied to samples tested at different times in different centres although caution is advised in such cases to avoid changes in assay testing techniques [15, 16]. An example where fixed cut-offs were used successfully is the European Seroepidemiology Network (ESEN and ESEN2) that included serological samples from eight antigens and 22 national laboratories [17]. Fixed cut-offs can also be applied to surveys where it is difficult to fit a mixture model for a number of reasons such as low proportion of samples belonging to an underlying component. An additional benefit of using a fixed cut-off is that it enables data to be characterized at the individual level as positive/negative for a specific antibody. This allows the possibility of undertaking additional analyses that makes use of other demographic information that may also be available at the individual level. This is less straightforward to do using mixture models by applying the probability of each sample 'belonging to one of the mixture model components' [18, 19].

However, fixed cut-offs have limitations which can question their appropriateness for seroepidemiological studies. The main limitation is that cut-offs are often defined by the assay manufacturer for diagnostic purposes and therefore, are not appropriate for seroepidemiology [7]. In this paper, the method for estimating

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

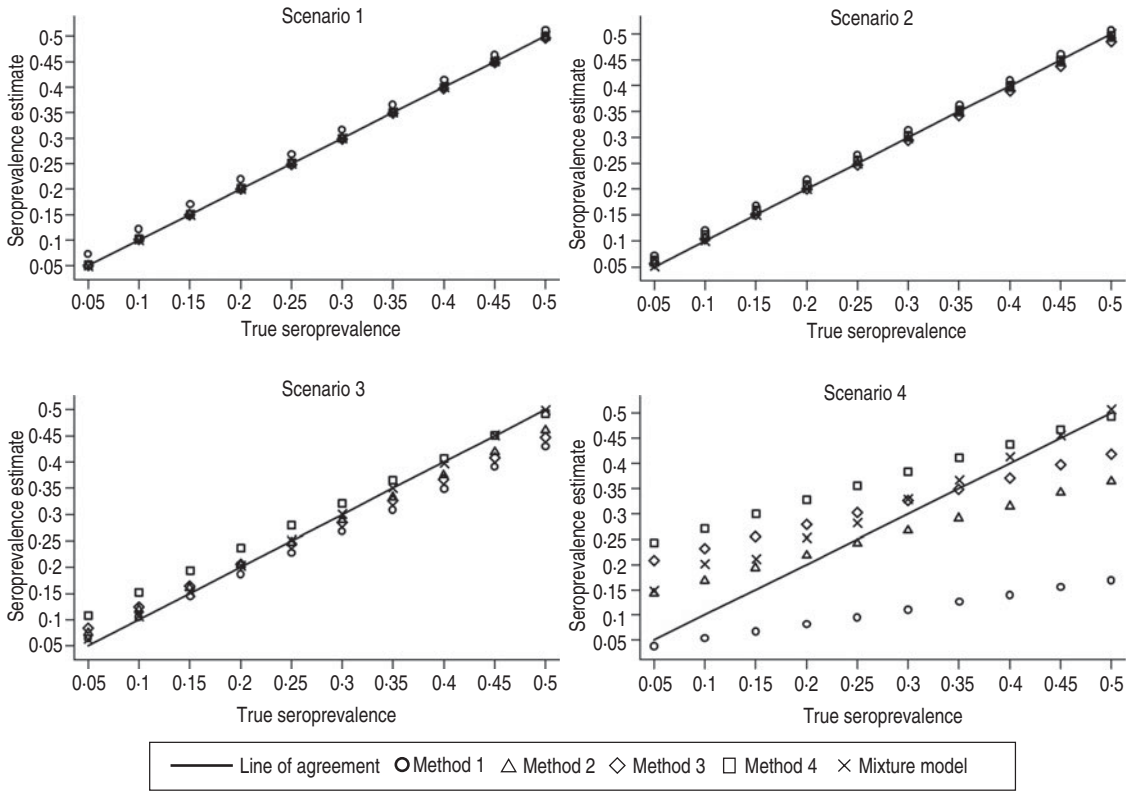

Fig. 4. Comparison of seroprevalence estimates between different methods and scenarios.

the diagnostic assay cut-off was compared to alternative cut-offs specifically estimated for seroepidemiology. In the simulation scenarios used, the seroprevalence based on the manufacturers' approach gave, in general, estimates further from the true value compared to the other methods shown especially for high overlap between susceptible and positive underlying populations. The other three cut-off estimation methods presented gave very similar results. Given that method 4 is not being affected by the proportion of samples belonging to each underlying component, this method may be preferred over the other two methods.

One important assumption needs to hold for the estimated cut-off to be applicable for seroprevalence estimation. The development serosurvey where the cut-off is estimated must have the same characteristics as the serosurvey the cut-off is applied to. Given that there is no change in assay method or testing techniques this assumption should hold for the location and dispersion parameters. However, the proportion of samples attributed to each underlying component can vary like the example given here with higher seroprevalence of HP4 in PWID compared to blood donors. From the simulation examples when the seroprevalence in the development serosurvey is similar to the serosurvey the cut-off is applied to, the final estimates were closer to the true value.

Another limitation of fixed cut-offs is that they cannot adjust for important confounders [3, 7]. This can be taken into account by estimating separate cut-offs for each level of the confounding variable (e.g. separate cut-offs by age group). Although age-specific cut-offs are meaningless from a diagnostic point of view they can be useful in seroepidemiology if they are used as a vehicle for estimating seroprevalence. An application of such age-specific cut-off has been shown previously [2].

Often antibody prevalence data using serum samples reflecting populations who are either susceptible or have experienced natural infection can be differentiated into clearly distinguishable distributions. These describe those previously infected who generally elicit strong antibody responses that are clearly distinct from those remaining susceptible where antibody is absent. This greatly facilitates employing a fixed cut-off with any potential resulting misclassification that may occur of relatively limited impact. Such circumstances therefore greatly favour employing a fixed cut-off. By contrast, however, this may not always be the case and distributions may sometimes significantly overlap where it can be much more difficult to visually

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press

distinguish those susceptible from those with evidence of antibody. For example, such situations may be encountered when considering highly vaccinated populations where significant proportions may demonstrate only lower antibody levels or when investigating infectious diseases where antibody levels wane significantly over time. This can result in larger proportions of samples with only low levels of antibody which tend to aggregate around proposed fixed cut-off points. Justifying any fixed cut-off point selected is therefore much more challenging and less intuitive in these instances, and may compromise the extent to which antibody prevalence can ultimately be meaningfully estimated. Such scenarios might therefore be argued to favour a mixture modelling approach to estimate seroprevalence, stratifying by age/time [3].

Apart from serum samples oral fluid is another important type of sample with significant compliance advantages that has often been employed for large antibody prevalence studies. However, IgG is present at much lower levels in oral fluids with large proportions of data often aggregating in those regions where a fixed cut-off is most likely to be placed. This makes the choice of a fixed cut-off challenging and therefore, a mixture modelling approach may be more appropriate [20].

In conclusion, fixed cut-offs have certain advantages compared to mixture modelling and provide valid seroprevalence estimates when there is little overlap between the underlying positive and susceptible populations. The bias can be reduced by using fixed cut-offs estimated specifically for seroprevalence studies.

DECLARATION OF INTEREST

The lead author (G.K.) declares no conflict of interest at the time the research was carried out. He is currently employed by a biotechnology company (Amgen Ltd) and owns shares in a biotechnology company (Amgen Ltd). The remaining authors declare no conflict of interest.

REFERENCES

1. **Giesecke J.** *Modern Infectious Disease Epidemiology*, 2nd edn. London: Arnold, 2002.
2. **Kafatos G.** Statistical analysis of quantitative seroepidemiological data (PhD thesis). The Open University, 2011.
3. **Vyse AJ, *et al.*** Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century. *Epidemiology and Infection* 2006; **134**: 1303–1312.
4. **Vyse AJ, *et al.*** The burden of parvovirus B19 infection in women of childbearing age in England and Wales. *Epidemiology and Infection* 2007; **135**: 1354–1362.
5. **Gay NJ.** Analysis of serological surveys using mixture models: application to a survey of parvovirus B19. *Statistics in Medicine* 1996; **15**: 1567–1573.
6. **Hardelid P, *et al.*** Analysis of rubella antibody distribution from newborn dried blood spots using finite mixture models. *Epidemiology and Infection* 2008; **36**: 1698–1706.
7. **Vyse AJ, *et al.*** Seroprevalence of antibody to varicella zoster virus in England and Wales in children and young adults. *Epidemiology and Infection* 2004; **132**: 1129–1134.
8. **Maple PAC, *et al.*** Testing UK blood donors for exposure to human parvovirus 4 using a time-resolved fluorescence immunoassay to screen sera and Western blot to confirm reactive samples. *Transfusion* 2013; **53**: 2575–2584.
9. **McLachlan G, Peel D.** *Finite Mixture Models*. New York: John Wiley & Sons Inc., 2000.
10. **Bland M.** *An Introduction to Medical Statistics*. Oxford: Oxford University Press, 2000.
11. **Maple PA, *et al.*** Application of a noninvasive oral fluid test for detection of treponemal IgG in a predominantly HIV-infected population. *European Journal of Clinical Microbiology and Infectious Diseases* 2006; **25**: 743–749.
12. **Baughman AL, *et al.*** Mixture model analysis for establishing a diagnostic cut-off point for pertussis antibody levels. *Statistics in Medicine* 2006; **25**: 2994–3010.
13. **Tong DDM, Buxser S, Vidmar TJ.** Application of a mixture model for determining the cutoff threshold for activity in high-throughput screening. *Computational Statistics and Data Analysis* 2007; **51**: 4002–4012.
14. **Ridge SE, Vizard AL.** Determination of the optimal cut-off value for a serological assay: an example using the Johne's absorbed EIA. *Journal of Clinical Microbiology* 1993; **31**: 1256–1261.
15. **Andrews N, *et al.*** The European Sero-Epidemiology Network: standardising the enzyme immunoassay results for measles, mumps and rubella. *Epidemiology and Infection* 2000; **125**: 127–141.
16. **Kafatos G, Andrews N, Nardone A.** Model selection methodology for inter-laboratory standardisation of antibody titres. *Vaccine* 2005; **23**: 5022–5027.
17. **Osborne K, Weinberg J, Miller E.** The European Sero-Epidemiology Network. *Eurosurveillance* 1997; **2**: 29–31.
18. **Hardelid P, *et al.*** Rubella seroprevalence in pregnant women in North Thames: estimates based on newborn screening samples. *Journal of Medical Screening* 2009; **16**: 1–6.
19. **Vyse AJ, Hesketh LM, Pebody RG.** The burden of infection with cytomegalovirus in England and Wales: how many women are infected in pregnancy? *Epidemiology and Infection* 2009; **137**: 526–533.
20. **Gay NJ, *et al.*** Improving sensitivity of oral fluid testing in IgG prevalence studies: application of mixture models to a rubella antibody survey. *Epidemiology and Infection* 2003; **130**: 285–291.

https://doi.org/10.1017/S0950268815001958 Published online by Cambridge University Press