



Low accuracy of Bayesian latent class analysis for estimation of herd-level true prevalence under certain disease characteristics—An analysis using simulated data

McAloon, Conor G.; Doherty, Michael L.; Whyte, Paul; Verdugo, Cristobal; Toft, Nils; More, Simon J.; O'Grady, Luke; Green, Martin J.

Published in:
Preventive Veterinary Medicine

Link to article, DOI:
[10.1016/j.prevetmed.2018.11.014](https://doi.org/10.1016/j.prevetmed.2018.11.014)

Publication date:
2019

Document Version
Early version, also known as pre-print

[Link back to DTU Orbit](#)

Citation (APA):
McAloon, C. G., Doherty, M. L., Whyte, P., Verdugo, C., Toft, N., More, S. J., O'Grady, L., & Green, M. J. (2019). Low accuracy of Bayesian latent class analysis for estimation of herd-level true prevalence under certain disease characteristics—An analysis using simulated data. *Preventive Veterinary Medicine*, 162, 117-125. <https://doi.org/10.1016/j.prevetmed.2018.11.014>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Title**

2 Low accuracy of Bayesian latent class analysis for estimation of herd-level true
3 prevalence under certain disease characteristics - An analysis using simulated data

4 **Author names and affiliations**

5 Conor G. McAloon^a, Michael L. Doherty^a, Paul Whyte^a, Cristobal Verdugo^b, Nils Toft^c,
6 Simon J. More^a, Luke O'Grady^a, Martin J. Green^d

7 ^aSchool of Veterinary Medicine, University College Dublin, Belfield, Dublin

8 D04 W6F6, Ireland

9 ^bInstituto de Medicina Preventiva Veterinaria, Universidad Austral de Chile, Valdivia,
10 Chile

11 ^cNational Veterinary Institute, Technical University of Denmark, Lyngby, Denmark

12 ^dSchool of Veterinary Medicine and Science, University of Nottingham, Sutton
13 Bonington, United Kingdom

14 **Corresponding author**

15 Conor G. McAloon

16 E-mail: conor.mcaloon@ucd.ie

17 Section of Herd Health and Animal Husbandry, School of Veterinary Medicine,
18 University College Dublin, Belfield, Dublin D04 W6F6, Ireland

19 +353 1 716 6083

20

21 **Abstract**

22 Estimation of the true prevalence of infected individuals involves the application of a
23 diagnostic test to a population and adjusting according to test performance, sensitivity
24 and specificity. Bayesian latent class analysis for the estimation of herd and animal-level
25 true prevalence, has become increasingly used in veterinary epidemiology and is
26 particularly useful in incorporating uncertainty and variability into analyses in a flexible
27 framework. However, the approach has not yet been evaluated using simulated data
28 where the true prevalence is known. Furthermore, using this approach, the within-herd
29 true prevalence is often assumed to follow a beta distribution, the parameters of which
30 may be modelled using hyperpriors to incorporate both uncertainty and variability
31 associated with this parameter. Recently however, the authors of the current study
32 highlighted a potential issue with this approach, in particular, with fitting the
33 distributions and a tendency for the resulting distribution to invert and become
34 clustered at zero. Therefore, the objective of the present study was to evaluate
35 commonly specified models using simulated datasets where the herd-level true
36 prevalence was known. The specific purpose was to compare findings from models
37 using hyperpriors to those using a simple beta distribution to model within-herd
38 prevalence. A second objective was to investigate sources of error by varying
39 characteristics of the simulated dataset. *Mycobacterium avium* subspecies
40 *paratuberculosis* infection was used as an example for the baseline dataset. Data were
41 simulated for 1000 herds across a range of herd-level true prevalence scenarios, and
42 models were fitted using priors from recently published studies. The results
43 demonstrated poor performance of these latent class models for diseases characterised
44 by poor diagnostic test sensitivity and low within-herd true prevalence. All variations of
45 the model appeared to be sensitive to the prior and tended to overestimate herd-level

46 true prevalence. Estimates were substantially improved in different infection scenarios
47 by increasing test sensitivity and within-herd true prevalence. The results of this study
48 raise questions about the accuracy of published estimates for the herd-level true
49 prevalence of paratuberculosis based on serological testing, using latent class analysis.
50 This study highlights the importance of conducting more rigorous sensitivity analyses
51 than have been carried out in previous analyses published to date.

52

53 **1. Introduction**

54 Prevalence is an important measurement of disease (or infection) occurrence.
55 Estimation of the true prevalence (P_T) within a population involves the application of a
56 diagnostic test to calculate apparent prevalence (P_A) and adjusting according to test
57 performance, sensitivity (Se) and specificity (Sp) (Rogan and Gladen, 1978).
58 However, there is often uncertainty regarding Se and Sp , and published values may
59 vary. Much of this variation can be attributed to differences among reference
60 populations and sampling strategies that have been used for the test validation
61 procedure (Greiner and Gardner, 2000). In addition, Se and Sp may vary according stage
62 of infection (Nielsen and Toft, 2008), prevalence (Brenner and Gefeller, 1997) and
63 between herds (Greiner and Gardner, 2000). It may therefore be unreasonable to
64 assume a fixed, constant, Se and Sp over different populations (Berkvens et al., 2006).
65 Consequently, the relationship between P_T and P_A can also be expected to vary between
66 populations.

67

68 The use of Bayesian latent class analysis for the estimation of herd (HTP) and animal-
69 level (ATP) true prevalence has become increasingly frequent in veterinary
70 epidemiology (Branscum et al., 2004). Using this approach, all parameters are

71 considered random variables that can be modelled using probability distributions.
72 Uncertainty and variability associated with estimates of test Se and Sp may therefore be
73 incorporated in the analysis. The resulting Bayesian posterior probability distribution
74 will provide inference on prevalence estimates, conditional on both currently observed
75 data and previous knowledge regarding the prevalence of infection.

76

77 To date, many of the studies that have estimated HTP using Bayesian latent class
78 analysis have examined cross sectional test data using models proposed by Hanson et
79 al. (2003). Using this approach, the number of animals testing positive in each herd is a
80 function of the within-herd ATP, and the performance of the test. However, to the
81 authors' knowledge this approach has not yet been evaluated using simulated data for
82 which the HTP is known and this is a fundamental step to assess model performance
83 when no gold standard is available.

84

85 Furthermore, using this approach, the ATP within infected herds is assumed to follow a
86 beta distribution, the parameters of which are estimated from hyperpriors. This method
87 aims to account for both the uncertainty and variability in within-herd ATP between
88 herds (Hanson et al., 2003). Hyperpriors are fitted as beta (μ) and gamma (ψ)
89 distributions to model within-herd ATP in the form Beta ($\mu\psi, \psi(1-\mu)$) (Hanson et al.,
90 2003). However, McAloon et al. (2016) reported a potential issue when using
91 hyperpriors to estimate HTP of paratuberculosis in Irish dairy herds. This related to
92 issues fitting the hyperprior, and a tendency for the resulting beta distribution to invert
93 and become clustered at zero, which is counterintuitive given that it is used to model
94 true prevalence within infected herds, i.e. when prevalence is > 0 by definition. The
95 authors in that study therefore opted to use a simple beta distribution to model within-

96 herd true prevalence which incorporated both the uncertainty and the variability
97 associated with the parameter, assuming an average within-herd ATP distribution over
98 all herds. More recently, other authors have used a logit-normal distribution to model
99 within-herd ATP of digital dermatitis infection in dairy cattle (Yang et al., 2017).

100

101 The consequences of using one approach to model within-herd ATP over another is not
102 clear since HTP remains unknown. However, testing each method against simulated
103 data with a known and fixed HTP would facilitate comparison of these methods whilst
104 also providing an evaluation of the overall method. The first objective of this study
105 therefore was to evaluate a Bayesian latent class analysis model for the estimation of
106 HTP, using simulated datasets over a range of known HTPs and to compare findings
107 from models using beta hyperpriors, logit-normal hyperpriors and those using a simple
108 beta distribution to model within-herd ATP. Model inputs for the base model were
109 based on estimation of paratuberculosis HTP as an example. Paratuberculosis infection
110 is characterised by a poor test Se and generally low within-herd ATP. The second study
111 objective was to investigate how different infection characteristics and test
112 performance influence the accuracy of the model by increasing Se and within-herd ATP
113 in the simulated datasets and in the priors for the corresponding estimating models.

114

115 2. **Materials and Methods**

116 2.1. Study population – data simulation

117 Table 1 shows the list of abbreviations used in the manuscript. Diagnostic test data
118 were simulated for a range of known or actual HTP (aHTP), i.e. the proportion of herds
119 with 1 or more infected cows. At each aHTP, data were simulated for 1000 herds as
120 follows. The number of animals in each herd was drawn from a gamma distribution

121 (rounded to the nearest integer) which had been fitted to herd sizes from an earlier
122 study (McAloon et al., 2016) using the “fitdistrplus” package in R (R Core Team, 2015),
123 and each herd size was rounded to the nearest integer. The number of animals testing
124 positive from each herd was then simulated with the following model;

$$125 \text{ Npos}_i \sim \text{Binomial} (P_{Ai}, \text{herdsize}_i)$$

$$126 P_{Ai} = Se \times ATP_i + (1-Sp) \times (1-ATP_i)$$

$$127 ATP_i = HTP_i \times CWHP_i$$

$$128 HTP_i \sim \text{Bernoulli} (aHTP)$$

$$129 CWHP_i \sim \text{Beta}(\text{alphac}_{WHP}, \text{betac}_{WHP})$$

$$130 Se \sim \text{Beta}(\text{alpha}_{Se}, \text{beta}_{Se})$$

$$131 Sp \sim \text{Beta}(\text{alpha}_{Sp}, \text{beta}_{Sp})$$

$$132 \text{Herdsize}_i \sim \text{Gamma}(S1, S2)$$

133 Where Npos_i was the number of test positive animals in the i -th herd; Npos_i was drawn
134 from a binomial distribution with a probability equal to the within-herd P_{Ai} , and n trials
135 equal to the herdsize_i ; P_A was determined by the ATP in the i -th herd, and the test Se
136 and Sp . Herdsize_i was drawn from a gamma distribution rounded to the nearest integer.

137 ATP was a combination of the HTP and the Conditional Within-Herd Prevalence

138 (CWHP), defined as the within-herd ATP conditional on the herd being infected, i.e.

139 when $HTP > 0$. HTP for the i -th herd was drawn from a Bernoulli distribution with a

140 probability equal to the ‘actual HTP’ (aHTP). In the first instance, datasets were

141 simulated across 3 different HTP scenarios: low HTP, with aHTPs of 0.10, 0.20, 0.30 and

142 0.40; medium HTP with aHTPs of 0.35, 0.45, 0.55 and 0.65; and high HTP, with aHTPs of

143 0.60, 0.70, 0.80 and 0.90. The use of these different HTP scenarios facilitated the use of

144 low, medium and high priors to be used in the estimating model.

145

146 Datasets were simulated for a CWHP beta distribution with a mode of 0.05, and a 95th
147 percentile of 0.15. Parameters of the input distributions are shown in Table 2 and R-
148 code for the simulation of the datasets is provided as Supplementary Material 1.

149 2.2. Prevalence estimation

150 The estimated Herd-level True Prevalence (eHTP) was then found using Bayesian latent
151 class analysis from these datasets. The model had the following model structure;

$$152 N_{pos_i} \sim \text{binomial}(P_{Ai}, \text{herdsize}_i)$$

$$153 P_{Ai} = Se \times ATP_i + (1-Sp) \times (1-ATP_i)$$

$$154 ATP_i = HTP_i \times CWHP_i$$

$$155 HTP_i \sim \text{Bernoulli}(eHTP)$$

$$156 Se \sim \text{beta}(\alpha_{Se}, \beta_{Se})$$

$$157 Sp \sim \text{beta}(\alpha_{Sp}, \beta_{Sp})$$

158 CWHP was modelled in four different ways to compare the outcomes. The first model,
159 represented as BETA, used a simple beta prior distribution (McAloon et al., 2016)
160 whereas the second and third used beta hyperpriors from recently published studies,
161 called BETA-HYP1 (Verdugo et al., 2015) and BETA-HYP2 (Pozzato et al., 2011). These
162 distributions were in the form; Beta ($\mu\psi, \psi(1-\mu)$) where μ is a beta distribution used to
163 model the mean CWHP and ψ is a gamma distribution used to model the variation
164 between herds. In this model structure, the degree of variation between herds is
165 inversely proportional to ψ (Hanson et al., 2003); that is, with higher values of ψ , herds
166 will have more similar CWHP.

167

168 Although BETA-HYP1 and BETA-HYP2 were both originally used as priors to estimate
169 the prevalence of paratuberculosis, they were chosen to reflect the knowledge available
170 on those specific populations at a specific time. For this study, they were chosen as they

171 were relevant to paratuberculosis characteristics i.e. representing low CWHP, however,
172 they also represented two variations of CWHP: one in which the prior for mean CWHP
173 was quite precise, with moderate variation between herds (Verdugo et al., 2015) and
174 the second in which the prior for mean CWHP was imprecise with a greater level of
175 between-herd variation, i.e. with a higher mean ψ (15.8; Pozatto et al., 2011). The fourth
176 model used a logit-normal distribution in the form $\text{logit}(\text{CWHP}_i) = \beta + \alpha_i$, where β is the
177 logit-mean CWHP and α_i is a herd-level random effect modelled as a normal distribution
178 with a mean of 0 and precision τ . This model structure was designated LOGIT-N. The
179 form of each method is shown below and model priors are shown in Table 2.

180

181 *Model - BETA*

182 $\text{CWHP}_i \sim \text{beta}(\text{alpha}, \text{beta})$

183

184 *Model - BETA-HYP1/BETA-HYP2*

185 $\text{CWHP}_i \sim \text{beta}(\mu_i \psi_i, \psi_i (1 - \mu_i))$

186 $\mu_i \sim \text{beta}(\text{alpha}, \text{beta})$

187 $\psi_i \sim \text{gamma}(S1, S2)$

188

189 *Model - LOGIT-N*

190 $\text{logit}(\text{CWHP}_i) = \beta + \alpha_i$

191 $\alpha_i \sim \text{norm}(0, 1/\tau)$

192

193 2.3. Sensitivity analysis

194 Sensitivity analysis was conducted by simulating and analysing a number of scenarios.

195 2.3.1. eHTP prior

196 In each case, aHTPs were simulated across 3 different HTP scenarios (low, medium and
197 high) as described above. For each of these scenarios, two different eHTP priors were
198 trialled: firstly, a uniform beta(1,1) distribution was used as the prior for eHTP. Next, a
199 beta prior which corresponded to the HTP scenario being simulated was also trialled. In
200 the low HTP scenario, a beta prior with a mode of 0.25 was used, in the medium HTP
201 scenario, a beta prior with a mode of 0.50 was used, and in the high HTP scenario a beta
202 prior with a mode of 0.75 was used (Table 2).

203

204 2.3.2. CWHP simulation method

205 In the base dataset, CWHP was simulated using a simple beta distribution. To assess the
206 sensitivity of this method to the method used to simulate the data, alternative datasets
207 were simulated in which CWHP was modelled using exactly the same model structure
208 and inputs as the analytical model used for the estimation. For example, when assessing
209 the accuracy of BETA-HYP1, this model was trialled on a dataset in which CWHP was
210 simulated using a simple beta distribution, and a second dataset in which CWHP was
211 modelled using the same model structure as the analytical model. In each case μ and ψ
212 were specified as distributions for the overall population. The CWHP for the *i-th* herd
213 was then simulated by first drawing separately from these two distributions. These
214 drawn values were used to generate parameters for a beta distribution, from which a
215 single value was simulated as the CWHP of the herd. The datasets generated using the
216 simple beta distribution and the dataset simulated according to the form of the
217 estimating model were designated “Simple” and “Model Form” datasets respectively.
218 The same approach was taken for BETA-HYP2 and LOGIT-N.

219 2.3.3. Test and disease characteristics

220 For the second objective, we investigated how the accuracy of the prevalence estimates
221 changed according to CWHP and test performance. The steps above were repeated
222 under alternative infection scenarios with medium (mode, 0.5, 95% less than 0.6) and
223 high (mode 0.8, 95% greater than 0.7) test Se; and for medium and high CWHP. For the
224 CWHP sensitivity analysis, the distributions dictating the variability between herds, i.e.
225 the gamma components for BETA-HYP1, BETA-HYP2 and LOGIT-N, were maintained
226 from the base model, and only the parameters dictating the mean of the overall
227 distribution were varied, i.e. the beta distributions for BETA-HYP1 and BETA-HYP2 and
228 the normal distribution for LOGIT-N (Table 2).

229

230 Models were implemented in WinBUGS 4.3.1 (Lunn et al., 2000) with the first 5,000
231 iterations discarded as burn in and 15,000 iterations used for posterior inference.
232 Convergence was assessed by visual inspection of the time series trace plots and by
233 running multiple ($n = 3$) chains from different starting values. In all cases, chains
234 reached stationary distributions within 5,000 iterations. A number of models were also
235 run for 100,000 iterations check for identifiability issues.

236

237 **3. Results**

238 Figure 1 shows the distributions of CWHP simulated from each of the model structures.
239 BETA-HYP2 in particular demonstrates significant clustering at zero as occurs when the
240 alpha parameter of the beta distribution is <1 .

241

242 Figure 2 plots the range of aHTP against the estimated HTP (eHTP) for low, medium and
243 high HTP scenarios. Four main conclusions can be drawn from these figures: 1, in
244 general, models were poor at estimating aHTP; 2. this estimation was not substantially

245 improved by varying the method used to model CWHP in the analytical model; 3, using
246 exactly the same model structure to simulate CWHP as that used for the analytical
247 model did not improve estimates, in fact, in many cases it appeared to make the
248 estimates worse; and 4, the estimates tended to be quite sensitive to the HTP prior used,
249 particularly with high HTPs. In the low HTP scenario, all the models tended to
250 overestimate HTP, with the exception of the BETA model which underestimated
251 prevalence for HTPs of 0.3 and 0.4, regardless of the prior used. Similarly, in the
252 medium HTP scenario, all models with the exception of the BETA model overestimated
253 HTP. In the high HTP scenario, estimates tended to cluster close to the HTP prior when
254 this was used, leading to overestimation of lower HTPs and under estimation of the 0.8
255 and 0.9 HTPs.

256

257 Figures 3 and 4 show the effect of varying the diagnostic test to medium and high Se
258 respectively. In general, accuracy of estimates are improved considerably with
259 increasing Se across all of the methods used to model CWHP. Both figures show
260 substantially improved HTP estimates and a much-reduced sensitivity to the prior for
261 HTP. Overall, there is still a tendency for models to overestimate HTP, particularly
262 models BETA-HYP1 and BETA-HYP2 and this tendency is reduced as test Se is
263 increased. The accuracy of the models are substantially improved at higher aHTPs,
264 particularly in the simple dataset. In contrast to the base model, there appears to be a
265 small improvement in using the same model structure for the simulation.

266 Figures 5 and 6 show the effect of increasing CWHP on the accuracy of the model. In
267 general, estimates were improved relative to the base scenario. However, in the
268 medium CWHP scenario, some large positive deviations in eHTP relative to aHTP may
269 be observed. This appears to be particularly evident at low aHTPs in the BETA-HYP2

270 model and in the model form scenarios, which could be related to the fact that the
271 CWHP distributions used to model this scenario include a large amount of between-
272 herd variability in CWHP.

273

274 **4. Discussion**

275 The use of simulated data to assess and compare the effectiveness of mathematical
276 models is a useful method of model evaluation that is commonly used within the field of
277 genetics (Stephens and Donnelly, 2003; Wilson and Rannala, 2003; Faubet et al., 2007))
278 and has gained increasing popularity with the field of veterinary epidemiology
279 (Denwood et al., 2010; Singleton and Breheny, 2016). Similarly, in veterinary
280 epidemiology, the use of Bayesian models to estimate prevalence has also increased in
281 recent years and is often used to estimate the prevalence of paratuberculosis, because
282 of uncertainty around the performance of diagnostic tests (Liapi et al., 2011; Pozzato et
283 al., 2011; Verdugo et al., 2015; McAloon et al., 2016). However, to the authors'
284 knowledge this is the first study that has used simulated data to evaluate the overall
285 accuracy of Bayesian latent class analysis for the estimation of HTP, and to evaluate the
286 effect of varying components within the model, for example the use of hyperpriors for
287 modelling CWHP.

288

289 This study raises substantive concerns about the effectiveness of conventional Bayesian
290 latent models to estimate paratuberculosis HTP and this may apply to other infections
291 or diseases with similar diagnostic test characteristics and where within-herd
292 prevalence is often very low. Irrespective of the method used to model CWHP, our
293 models tend to overestimate HTP. The HYP1, HYP2 and LOGIT-N models produced
294 estimates with larger probability intervals, whereas the BETA model produced median

295 values that were closer to aHTP, but with much narrower probability intervals. There
296 was little difference between the two hyperprior methods of modelling CWHP, however,
297 HYP2 tended to produce less predictable estimates in response to increasing aHTP in
298 comparison to HYP1 (Figure 5).

299

300 Importantly, when used in the paratuberculosis scenario, all models appeared to be
301 overly sensitive to the prior used for HTP, particularly when a high HTP prior was used.
302 Interestingly, in the worked example in Branscum (2004), we note that the median and
303 95th percentile of the posterior estimate for HTP (0.58, 0.83 respectively) were also
304 notably close to the median and 95th percentile from the prior distribution (0.59, 0.85
305 respectively). Similarly, in published examples of the method, Pozatto (2012) found that
306 the HTP (median, 95% credible intervals) in 2 regions in Italy was 0.70, 0.50-0.87 and
307 0.71, 0.54 – 0.87, whilst the prior distribution used for HTP in this study was 0.69, 0.50-
308 0.84. Liapi et al., (2011) used a prior of 0.65 with a 5th percentile of 0.40 and found a
309 posterior estimate of 0.61 and 0.42 respectively. In Bayesian analyses, when posterior
310 estimates closely reflect prior distributions, there is cause for concern that the data are
311 having little impact on the results, which suggests models may not be appropriately
312 specified. A greater difference between prior and posterior estimates was found in
313 Verdugo et al. (2015) who reported posterior estimates for HTP of 0.92 (0.87-0.96),
314 0.78 (0.74-0.83) and 0.75 (0.71-0.78) with a prior of 0.86 (0.59 – 0.95), however this
315 model used a different approach which allowed for an age-specific sensitivity for each
316 animal which were higher than the Se estimates used in other analyses. This study was
317 based on a larger sample size, however, our analyses have shown that the problems
318 identified with this method cannot be overcome by increasing sample size (data not
319 shown).

320

321 Figures 5 and 6 show large deviations of eHTP relative to aHTP at specific aHTP values,
322 for example in the BETA-HYP2 model on the Model Form dataset, under the low HTP
323 scenario (Figure 5). In these cases, the posterior distribution for eHTP was very high
324 relative to the aHTP, whereas the posterior estimate for Se was very low, approaching
325 zero. Repeat analysis with multiple chains showed stability of separate chains at two
326 different parameter spaces suggesting a problem with model identifiability. These
327 issues were not resolved by running the model for more (n=100,000) iterations or by
328 reducing the uncertainty around the Se prior but could be 'fixed' by varying the initial
329 starting values. In practice it may not be possible to know what the 'true' model is,
330 therefore for future studies, it is particularly important that multiple chains are run
331 from a variety of initial values, to check for identifiability issues. In addition,
332 examination and reporting of the posterior distributions for the rest of the parameters
333 in the model is also recommended, including those parameters that are not specifically
334 of interest.

335

336 Studies using simulation to assess model accuracy often generate a reasonably large
337 number of datasets from a particular model with particular parameters. Each of these
338 datasets is analysed, and the results used to examine the performance of the estimation
339 method. For example, Singleton et al. (2016) used simulated data to assess the utility of
340 a non-linear hierarchical model applied to experimental infection data. Three sample
341 sizes were chosen, and 5,000 datasets generated for each set of parameters with each
342 dataset analysed by the proposed model. In the case of our study, the outcome of
343 interest at each aHTP was a known point prevalence which would not change if
344 additional datasets were generated. For each aHTP however, 1,000 herds were

345 simulated for each set of parameter values, representing the replicated datasets to
346 assess the method.

347

348 The use of hyperpriors to model within-herd ATP is commonly advocated in the use of
349 latent class estimation of HTP. Using this method, hyperpriors are fitted as beta (μ) and
350 gamma (ψ) distributions to model within-herd ATP in the form Beta ($\mu\psi, \psi(1-\mu)$)
351 (Hanson et al., 2003). The potential advantage of this method is that it facilitates the
352 incorporation of both uncertainty regarding the parameter as well as the between-herd
353 variability. The distributions are fitted through the elicitation of expert opinion, who are
354 asked to specify the mean and confidence intervals of the within-herd ATP across herds,
355 which is fitted as a beta distribution (μ). Then, conditional on the mean, experts are
356 asked to specify the value below which they are 95% sure that 90% of the within-herd
357 ATP are below. These values are then used to fit the gamma distribution (ψ). However,
358 whilst this method has obvious theoretical advantages, we argue that the data required
359 from expert elicitation may be restrictively complex. Furthermore, McAloon et al.
360 (2016) highlighted inconsistencies in published literature between values elicited from
361 experts and those same percentiles based on simulation of the hyperprior distributions.
362 Finally, given that within this method, distributions are fitted conditional on a mean,
363 rather than mode, the distribution often becomes inverted, and very often the median
364 prevalence within infected herds may be less than 0.01. This is potentially problematic
365 with small to medium herd sizes as herds may be deemed infected yet have less than 1
366 infected cow in the herd. We hypothesised that this may result in overestimation of
367 HTP. The present study seems to suggest that the use of beta hyperpriors does appear
368 to overestimate the HTP more so than the BETA or LOGIT-N models. This

369 overestimation is particularly evident with priors that incorporate increased variability
370 in CWP, for example the BETA-HYP1 model.

371

372 A possible explanation for this finding is that this method fits the 90th percentile
373 conditional on a fixed mean. However, given a beta distribution with a fixed mean,
374 increasing the variance in order to increase the 90th percentile leads to a shift in the
375 median in the opposite direction creating an increasing skewed distribution. If the mean
376 is low as is the case in paratuberculosis, the median moves very close to 0 as the alpha
377 parameter becomes < 1 . With very low CWHP, the probability of herds being infected
378 with an AP of 0 increases potentially leading to this herd being “infected” across more
379 iterations. In contrast, the LOGIT-N method, facilitates increased variation but still
380 retains a distribution shape that is possibly more reflective of the likely distribution
381 (Figure 1). However, the overall effect of this problem with the method of modelling
382 CWHP was relatively minor when compared with the problems associated with the
383 overall use of the model to estimate HTP of paratuberculosis, with relatively poor Se
384 and low CWHP. Increasing the Se to 0.5 and 0.8 led to increases in the accuracy of the
385 estimates. Similarly, increasing the mode of the distribution used to model mean CWHP
386 to 0.3 and 0.7 also led to increased accuracy of the estimates and a decreased sensitivity
387 to the HTP prior used, across all of the models used. Therefore, these models may be
388 reasonably accurate when used to estimate prevalence for infections or diseases with
389 poor Se *or* low CWHP but not when both of these are present.

390

391 In addition, it is important to note that during the simulation stage of this study, the
392 “design” aHTP used to generate the simulated dataset may have differed from the actual
393 proportion of herds in the simulated dataset with one or more infected animals. This

394 occurred because herds were first simulated as infected by drawing from a Bernoulli
395 distribution with a probability equal to the aHTP. Within those herds deemed infected,
396 the number of infected individuals was then drawn from a binomial distribution with a
397 probability equal to CWHP drawn for that herd. However, with moderate herd sizes and
398 low CWHP, the probability of drawing zero infected individuals in an “infected” herd is
399 >0 and increases with decreasing CWHP. Within the low CWHP datasets, the difference
400 between the design and actual datasets was greatest for the BETA-HYP1 and BETA-
401 HYP2 models compared to the BETA and LOGIT-N models, probably because of the
402 greater tendency for this model structure to become clustered at zero. All of the models
403 in general tended to overestimate aHTP, and aHTP may be an overestimate of the actual
404 proportion of infected herds.

405

406 **5. Conclusion**

407 Our results suggest poor accuracy of commonly specified Bayesian latent class models
408 for paratuberculosis herd-level true prevalence estimation. All variations of the model
409 appeared to be sensitive to the prior and tended to overestimate herd-level true
410 prevalence, raising questions about whether previous estimates of paratuberculosis
411 HTP reported in the literature may be inaccurate. Estimates were substantially
412 improved in different infection scenarios by increasing test sensitivity and within-herd
413 true prevalence. This study highlights the importance of conducting more rigorous
414 sensitivity analyses than have been carried out in previous analyses published to date.
415 In addition, we advocate increased use of simulation as an initial stage in conducting
416 future analyses and also suggest that new model methodologies be explored, to
417 determine whether alternative approaches might perform better than conventional
418 latent class models.

419

420 **Acknowledgements**

421 Conor McAloon is supported by a UCD Wellcome Institutional Strategic Support Fund,

422 which was financed jointly by University College Dublin and the SFI-HRB-Wellcome

423 Biomedical Research Partnership (ref 204844/Z/16/Z).

424 **6. References**

- 425 Berkvens, D., Speybroeck, N., Praet, N., Adel, A., Lesaffre, E., 2006. Estimating disease
426 prevalence in a Bayesian framework using probabilistic constraints. *Epidemiology* 17,
427 145-153.
- 428 Branscum, A., Gardner, I., Johnson, W., 2004. Bayesian modeling of animal-and herd-
429 level prevalences. *Preventive veterinary medicine* 66, 101-112.
- 430 Brenner, H., Gefeller, O., 1997. Variation of sensitivity, specificity, likelihood ratios and
431 predictive values with disease prevalence. *Statistics in medicine* 16, 981-991.
- 432 Denwood, M.J., Reid, S.W.J., Love, S., Nielsen, M.K., Matthews, L., McKendrick, I.J.,
433 Innocent, G.T., 2010. Comparison of three alternative methods for analysis of equine
434 faecal egg count reduction test data. *Preventive veterinary medicine*, 93, 316-323.
- 435 Dorshorst, N.C., Collins, M.T., Lombard, J.E., 2006. Decision analysis model for
436 paratuberculosis control in commercial dairy herds. *Preventive veterinary medicine* 75,
437 92-122.
- 438 Espejo, L., Godden, S., Hartmann, W., Wells, S., 2012. Reduction in incidence of Johne's
439 disease associated with implementation of a disease control program in Minnesota
440 demonstration herds. *Journal of dairy science* 95, 4141-4152.
- 441 Faubet, P., Waples, R.S., Gaggiotti, O.E., 2007. Evaluating the performance of a multilocus
442 Bayesian method for the estimation of migration rates. *Molecular ecology*, 16, 1149-
443 1166.
- 444 Ferrouillet, C., Wells, S., Hartmann, W., Godden, S., Carrier, J., 2009. Decrease of Johne's
445 disease prevalence and incidence in six Minnesota, USA, dairy cattle herds on a long-
446 term management program. *Preventive veterinary medicine* 88, 128-137.
- 447 Greiner, M., Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary
448 diagnostic tests. *Preventive veterinary medicine* 45, 3-22.

449 Hanson, T., Johnson, W.O., Gardner, I.A., 2003. Hierarchical models for estimating herd
450 prevalence and test accuracy in the absence of a gold standard. *Journal of Agricultural,*
451 *Biological, and Environmental Statistics* 8, 223-239.

452 Johnson-Ifeorulundu, Y., Kaneene, J., Sprecher, D., Gardiner, J., Lloyd, J., 2000. The effect
453 of subclinical *Mycobacterium paratuberculosis* infection on days open in Michigan, USA,
454 dairy cows. *Preventive veterinary medicine* 46, 171-181.

455 Liapi, M., Leontides, L., Kostoulas, P., Botsaris, G., Iacovou, Y., Rees, C., Georgiou, K.,
456 Smith, G.C., Naseby, D.C., 2011. Bayesian estimation of the true prevalence of
457 *Mycobacterium avium* subsp. *paratuberculosis* infection in Cypriot dairy sheep and goat
458 flocks. *Small Ruminant Research* 95, 174-178.

459 Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS — a Bayesian
460 modelling framework: concepts, structure, and extensibility. *Statistics and Computing,*
461 10:325–337.

462 McAloon, C.G., Doherty, M.L., Whyte, P., O’Grady, L., More, S.J., Messam, L.L.M., Good, M.,
463 Mullooney, P., Strain, S., Green, M.J., 2016. Bayesian estimation of prevalence of
464 paratuberculosis in dairy herds enrolled in a voluntary Johne’s Disease Control
465 Programme in Ireland. *Preventive veterinary medicine* 128, 95-100.

466 Nielsen, S.S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: a review of
467 accuracies of ELISA, interferon- γ assay and faecal culture techniques. *Veterinary*
468 *microbiology* 129, 217-235.

469 Pillars, R., Grooms, D., Gardiner, J., Kaneene, J., 2011. Association between risk-
470 assessment scores and individual-cow Johne's disease-test status over time on seven
471 Michigan, USA dairy herds. *Preventive veterinary medicine* 98, 10-18.

472 Pozzato, N., Capello, K., Comin, A., Toft, N., Nielsen, S.S., Vicenzoni, G., Arrigoni, N., 2011.
473 Prevalence of paratuberculosis infection in dairy cattle in Northern Italy. Preventive
474 veterinary medicine 102, 83-86.

475 Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test.
476 American journal of epidemiology 107, 71-76.

477 Singleton, M.D., Breheny, P.J., 2016. Nonlinear hierarchical modeling of experimental
478 infection data. Preventive veterinary medicine, 130, 129-136.

479 Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype
480 reconstruction from population genotype data. The American journal of human
481 genetics, 73, 1162-1169.

482 Verdugo, C., Toft, N., Nielsen, S.S., 2015. Within-and between-herd prevalence variation
483 of Mycobacterium avium subsp. paratuberculosis infection among control programme
484 herds in Denmark (2011–2013). Preventive veterinary medicine 121, 282-287.

485 Wilson, G.A., Rannala, B., 2003. Bayesian inference of recent migration rates using
486 multilocus genotypes. Genetics, 163, 1177-1191.

487 Yang, D.A., Heuer, C., Laven, R., Vink, W.D., Chesterton, R.N., 2017. Farm and cow-level
488 prevalence of bovine digital dermatitis on dairy farms in Taranaki, New Zealand. New
489 Zealand veterinary journal, 65(5), pp.252-256.