

Scotland's Rural College

British Escherichia coli O157 in Cattle Study (BECS): to determine the prevalence of E. coli O157 in herds with cattle destined for the food chain

Henry, MK; Tongue, SC; Evans, J; Webster, C; McKendrick, IJ; Morgan, M; Willett, A; Reeves, A; Humphry, RW; Gally, DL; Gunn, GJ; Chase-Topping, ME

Published in:
Epidemiology and Infection

DOI:
[10.1017/S0950268817002151](https://doi.org/10.1017/S0950268817002151)

First published: 19/09/2017

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Henry, MK., Tongue, SC., Evans, J., Webster, C., McKendrick, IJ., Morgan, M., Willett, A., Reeves, A., Humphry, RW., Gally, DL., Gunn, GJ., & Chase-Topping, ME. (2017). British *Escherichia coli* O157 in Cattle Study (BECS): to determine the prevalence of *E. coli* O157 in herds with cattle destined for the food chain. *Epidemiology and Infection*, 145(15), 3168 - 3179. Advance online publication. <https://doi.org/10.1017/S0950268817002151>

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 ?

Take down policy

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.

British E. coli O157 in Cattle Study (BECS): to determine the prevalence of Escherichia coli O157 in herds with cattle destined for the food chain.

Journal:	<i>Epidemiology and Infection</i>
Manuscript ID	HYG-OM-8141-Mar-17.R1
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Henry, Madeleine; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Tongue, Sue; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Evans, Judith; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Webster, Catriona; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) McKendrick, Iain; BIoSS Morgan, Mair; RSK ADAS Ltd. Willett, Alice; RSK ADAS Ltd. Reeves, Aaron; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Humphry, Roger; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Gally, David; The Roslin Institute, University of Edinburgh, Immunity Division, Centre for Infectious Diseases Gunn, George; Scotland's Rural College (SRUC), Epidemiology Research Unit (Inverness Campus) Chase-Topping, Margo; Edinburgh University, Centre for Immunity Infection and Evolution
Keyword:	Epidemiology, Estimating disease prevalence, Escherichia coli (E. coli) O157, Bovine

SCHOLARONE™
Manuscripts

1 Title

2 British *E. coli* O157 in Cattle Study (BECS): to determine the prevalence of
3 *Escherichia coli* O157 in herds with cattle destined for the food chain.

4 Authorship

5 M.K. HENRY¹, S.C. TONGUE¹, J. EVANS¹, C. WEBSTER¹, I.J. McKENDRICK²,
6 M. MORGAN³, A. WILLETT³, A. REEVES¹, R.W. HUMPHRY¹, D.L. GALLY⁴, G.J. GUNN¹,
7 M.E. CHASE-TOPPING⁵

8 ¹ Epidemiology Research Unit (Inverness campus), Scotland's Rural College (SRUC), Kings
9 Buildings, West Mains Road, Edinburgh, EH9 3JG, U.K.

10 madeleine.henry@sruc.ac.uk, sue.tongue@sruc.ac.uk, judith.evans@sruc.ac.uk,
11 catriona.webster@sruc.ac.uk, aaron.reeves@sruc.ac.uk, roger.humphry@sruc.ac.uk,
12 george.gunn@sruc.ac.uk

13 ² Biomathematics and Statistics Scotland, James Hutton Institute, Invergowrie, Dundee,
14 DD2 5DA, U.K.

15 iain@bioss.ac.uk

16 ³ RSK ADAS Ltd., Spring Lodge, 172 Chester Road, Helsby, Cheshire, WA6 0AR, U.K.

17 mair.morgan@adas.co.uk

18 alice.willett@adas.co.uk

19 ⁴ The Roslin Institute, R(D)SVS, University of Edinburgh, Immunity Division, Centre for
20 Infectious Diseases, Easter Bush, EH25 9RG, U.K.

21 dgally@ed.ac.uk

1
2
3 22 ⁵ Centre for Immunity, Infection and Evolution, University of Edinburgh, King's Buildings,
4
5 23 Edinburgh, EH9 3JT, U.K.

6
7
8 24 margo.chase@ed.ac.uk
9

10
11 25

12
13 26 **Running head: *E. coli* O157 cattle prevalence study**
14

15
16 27 **Corresponding author: sue.tongue@sruc.ac.uk**
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

28 Summary

29 *Escherichia coli* O157 are zoonotic bacteria for which cattle are an important reservoir.
30 Prevalence estimates for *E. coli* O157 in British cattle for human consumption are over ten
31 years old. A new baseline is needed to inform current human health risk. The British
32 *E. coli* O157 in Cattle Study (BECS) ran between September 2014 and November 2015 on
33 270 farms across Scotland and England & Wales. This is the first study to be conducted
34 contemporaneously across Great Britain, thus enabling comparison between Scotland and
35 England & Wales. Herd-level prevalence estimates for *E. coli* O157 did not differ significantly
36 for Scotland (0.236, 95% CI 0.166 – 0.325) and England & Wales (0.213, 95% CI 0.156 –
37 0.283) ($P = 0.65$). The majority of isolates were verocytotoxin positive. A higher proportion of
38 samples from Scotland were in the super-shedder category, though there was no difference
39 between the surveys in the likelihood of a positive farm having at least one super-shedder
40 sample. *E. coli* O157 continues to be common in British beef cattle, reaffirming public health
41 policy that contact with cattle and their environments is a potential infection source.

42

43

44 Introduction

45 Human infection with *Escherichia coli* (*E. coli*) O157 is a global concern, as infection can
46 lead to kidney failure, neurological complications and Haemolytic Uraemic Syndrome (HUS).
47 HUS can be fatal, particularly in young, elderly or immunocompromised patients [1].
48 Worldwide, the incidence of HUS due to *E. coli* O157 infection has been reported at
49 approximately 10% [2], with a 3-5% case-fatality rate [3], while the majority of those who
50 survive suffer some degree of chronic renal function impairment [3]. Cattle and their
51 environments are a reservoir of *E. coli* O157 [4-6]. Some strains produce verocytotoxin
52 (verocytotoxigenic *E. coli* (VTEC) O157) and can be excreted in cattle faeces in high
53 numbers, leading to the concept of super-shedding [7, 8]. Certain subtypes of *E. coli* O157,
54 specifically those with the genetic marker encoding toxin *vtx 2*, are more likely to be
55 associated with super-shedding in cattle and these also appear to pose the greatest risk for
56 transmission to humans [8, 9]. There is also evidence that both verocytotoxin type and
57 phage type are linked to, not only excretion levels in cattle but, disease severity in humans
58 [10].

59 In 1998-2000 and 2002-2004, two national cross-sectional surveys in Scotland (SEERAD
60 [11] and IPRAVE [12]) demonstrated the presence of *E. coli* O157 on approximately 20% of
61 farms producing cattle for human consumption. A structured survey in England & Wales
62 during 1999 estimated herd-level VTEC O157 prevalence to be 38.7% [13], whilst a 2003
63 convenience survey in England & Wales identified VTEC O157 on 32.2% of 255 farms [14].
64 Given the poor predictive value of a negative test result due to sporadic faecal shedding [15,
65 16], the advice from public health authorities has been to assume *E. coli* O157 are present in
66 all cattle faeces [17]. Control of shedding from cattle has been suggested as a means to
67 protect public health [9, 18], but is difficult to achieve.

68 Updated prevalence estimates are now required for Scotland and for England & Wales to
69 contextualise the current risk to human health from cattle. As there is evidence that the
70 primary VTEC O157 subtypes are changing in human infections in the UK [10], surveillance

1
2
3 71 of cattle should continue, in order to confirm whether equivalent shifts have occurred in the
4
5 72 cattle VTEC O157 population. If so, this would facilitate the development of measures to
6
7 73 mitigate risk to humans.

8
9
10 74 The study was designed to conduct contemporaneous surveys on equivalent cattle
11
12 75 populations in Scotland and England & Wales. Here we present the study methodology,
13
14 76 descriptive analysis of the sampled farms, the herd-level and pat-level prevalence estimates
15
16 77 obtained for *E. coli* O157 in British cattle destined for the food chain and the vtx frequencies
17
18 78 found. This study provides the essential foundation for a number of further analyses and
19
20 79 future investigative approaches.

21
22
23 80 ~~Additional aims were to repeat sample a subset of Scottish farms for temporal analysis and~~
24
25 81 ~~to collect strains of *E. coli* O157 currently circulating in cattle in both geographical areas for~~
26
27 82 ~~comparison with contemporaneous human strains.~~

31 83 **Methods**

34 84 **Study Design**

35
36
37
38 85 The British *E. coli* O157 in Cattle Study (BECS) described in this manuscript is comprised of
39
40 86 two surveys: one in Scotland and one in England & Wales.

41
42 87 In Scotland, the source population for the survey was the holdings that had participated in
43
44 88 both of two earlier Scottish cross-sectional cattle surveys (SEERAD from 1998-2000 [11]
45
46 89 and IPRAVE from 2002-2004 [12]) and still kept cattle aged between one and two years
47
48 90 and/or cattle over two years without offspring – i.e. they were likely to still be producing cattle
49
50 91 for slaughter. These were identified by matching the holding details from all the holdings
51
52 92 sampled in the SEERAD [11] & IPRAVE [12] surveys to determine the subset of holdings
53
54 93 that had been sampled in both. The postcode and farm names were then matched to official
55
56 94 records of cattle numbers (June Agricultural Census 2012 and Cattle Tracing System (CTS)

1
2
3 95 data from June 2013). The holdings sampled in the SEERAD & IPRAVE surveys were
4 originally selected from a list comprising 3,111 farms with cattle, randomly selected from
5
6
7 97 1997 Scottish Agricultural and Horticultural Census data [12].
8

9
10 98 The England & Wales survey was designed to be comparable to the Scottish survey. As
11
12 99 there had been no previous survey, a slightly wider definition of eligible farms was adopted,
13
14 100 to reduce the risk of excluding potentially eligible farms. In England & Wales the source
15
16 101 population comprised holdings containing either at least one (non-dairy breed) female aged
17
18 102 one year or over, or at least one male (any breed) aged one year or over.
19

20
21 103 Sample sizes were estimated using reported prevalence from previous surveys (Scotland
22
23 104 20.5% [12] and England & Wales 39% [13]). Based on the proportion of herds positive and a
24
25 105 sensitivity of 90%, sampling at least 110 farms in Scotland and 160 farms in England &
26
27 106 Wales would provide 96% confidence that the true herd-level prevalence of *E. coli* O157
28
29 107 would fall within a tolerance range of 0.169 of the apparent prevalence estimated in these
30
31 108 surveys. This would be similar to values estimated for SEERAD (0.179) and IPRAVE (0.161)
32
33 109 [12].
34
35

36 110 The final sampling frame for Scotland contained 346 holdings. In England & Wales the
37
38 111 sampling frame was a random selection of 1,280 holdings from a source population of
39
40 112 56,621. This number of holdings would ensure that, if a worst case scenario of a 1:8
41
42 113 participation response was assumed, we would be able to recruit the minimum number of
43
44 114 holdings estimated in the sample size calculations above. Records were assigned a unique
45
46 115 ID and the sampling frames were randomised before recruitment.
47

48
49 116 Recruiters and field samplers were trained according to a standardised protocol. There were
50
51 117 two principal recruiters for each survey, with additional recruiters available if needed. Four
52
53 118 samplers were available in Scotland and ten in England & Wales.
54

55
56 119 Standard notification letters were sent to all farms one month before sampling started. Farms
57
58 120 were then available for telephone recruitment if they had not opted out within two weeks.
59
60

1
2
3 121 To ensure objective recruitment, a recruitment software application was developed; this
4
5 122 randomly selected one farm at a time from all eligible farms. From selection, it was the
6
7 123 recruiter's responsibility to reach one of four potential outcomes: 1. Contact made – further
8
9 124 information requested; 2. Farm recruited – passed to sampler for visit arrangement; 3. Farm
10
11 125 opted out; or 4. Farm could not be reached – moved to a reserve list. The last outcome (4)
12
13 126 followed three unsuccessful contact attempts. The reserve list would become available again
14
15 127 had all farms been phoned without achieving the minimum sample size.

16
17
18 128 Recruited farms received a pack giving information on the study, details of the survey
19
20 129 procedure, confidentiality, use of samples and data, information about *E. coli* O157 and a
21
22 130 consent form. Farms were assigned a new unique ID once a sampling visit was arranged.

23
24
25 131 Sampling visits started in mid-September 2014 and were distributed as evenly as logistically
26
27 132 feasible across geographical regions and over one calendar year. Each farm was visited
28
29 133 once. The sample group was the group of non-breeding cattle closest to slaughter on the
30
31 134 day of the visit. If mixed groups existed, the sampled group contained the cattle that met this
32
33 135 definition. The sampling unit was a fresh faecal pat. **Freshly voided discrete pats were**
34
35 136 **preferentially sampled following the sampling protocol developed for the previous Scottish**
36
37 137 **surveys [9, 17, 18]. The sample teams ensured that they did not sample from the same pat**
38
39 138 **twice, nor from old, dried, or desiccated pats.** The number of pats taken from each group
40
41 139 depended on group size and the sampling schedule from IPRAVE [12, 19, 20]. This gave
42
43 140 90% power to identify a sampled group as positive, if at least one animal were shedding
44
45 141 *E. coli* O157.

46
47
48 142 For each sample, a 30ml universal container was filled to just below the threaded portion
49
50 143 with faeces taken from several locations on a fresh pat. Samplers preferentially targeted
51
52 144 areas on the surface of the pat where mucus was apparent [21]. Samples were labelled and
53
54 145 kept cool during transport to the laboratory.

1
2
3 146 At the sampling visit a questionnaire was completed electronically through face-to-face
4
5 147 interview. The questionnaire (available on request from the corresponding author) was
6
7 148 adapted from the IPRAVE study. Questions covered aspects of farm demographics,
8
9 149 management and health status. Most questions related to the farm although some were
10
11 150 specific to the group of animals that was sampled. There was a different subset of questions
12
13 151 for the sampled group, dependent on whether they were housed, or grazing, at the time of
14
15 152 sampling.

153 Approval

154 The Food Standards Agency approved and authorised informed consent documentation and
155 the questionnaire. Personal data were handled in accordance with the Data Protection Act
156 (1998).

157 Case definition

158 A faecal pat was positive if *E. coli* O157 was detected using the laboratory methods below.
159 A farm was positive if it contained at least one positive pat.

160 Laboratory methods

161 *E. coli* O157 were isolated from 1g of faeces per sample, using immunomagnetic separation
162 methods previously described [22]. Enumeration of *E. coli* O157 was by limiting dilution
163 method on CT-SMac agar plates and was performed in duplicate for each sample [23]. The
164 limit of detection for enumeration was 100 colony-forming units per gram (CFU g⁻¹).
165 Polymerase Chain Reaction [24] was used to confirm the serogroup of the isolates as
166 *E. coli* O157 and further characterise one *E. coli* O157 isolate per pat, according to the
167 presence or absence of genes encoding toxins (*vtx*) 1 and 2. Isolates were sent to SERL
168 (Scottish *E. coli* Reference Laboratory) for confirmation of identity, further subtyping of toxin
169 genes and phage typing (results not included here).

170 Statistical methods

171 Herd-level prevalence and pat-level prevalence were estimated using SAS software version
172 9.4 [25]. Other statistical analyses were performed using R version 3.2.3 [26] and additional
173 R packages [27–31]. Surveys were analysed independently, except when stated otherwise.
174 Univariate statistical comparisons of recruitment and questionnaire data within and between
175 surveys were made using linear, generalised linear regression and ANOVA models,
176 Likelihood Ratio, Mann-Whitney, Fisher's Exact and Pearson's Chi-squared Tests, and
177 Pearson's Product-Moment Correlation, as appropriate. The statistical significance level, α ,
178 was set at a value of 0.05 throughout.

179 Prevalence

180 Herd-level prevalence estimates for Scotland were calculated using generalised linear mixed
181 models with a logit link function, fitted using marginal residual pseudo-likelihood (Proc
182 Glimmix, SAS software [25]). This method was chosen as it provided a consistent framework
183 for ongoing integrated modelling of the current data with the two historical Scottish
184 prevalence surveys; this analysis will need to accommodate the use of different, but inter-
185 related, 'G-side' covariance structures for different subsets of the data, reflecting the different
186 sampling designs in different studies. This issue is of continued relevance because the
187 sample for the current study was selected from the set of farms sampled in both of previous
188 studies, where one of these was not a simple random sample [12]. Thus, it is desirable to
189 produce prevalence estimates for the most recent study which respect the sampling
190 structures applied over the three successive surveys. 'Farm' and an effect to model the
191 effect of spatial-temporal clustering in one of the previous studies were fitted as random
192 effects. Mean estimates and confidence intervals were generated by back transforming from
193 the model output on the logit scale. Scottish pat-level prevalence was modelled in a similar
194 way.

195 Although there were no historical studies for England & Wales to be integrated into an
196 analysis, and hence no requirement to model complex sampling structures, a similar

1
2
3 197 implementation of the same approach to calculating farm and pat-level prevalence was
4
5 198 adopted for these data. For England & Wales, a generalised linear mixed model was fitted,
6
7 199 with a random 'farm' effect to model extra-binomial variability.
8

9
10 200 For all models, seasonal differences were estimated by incorporating 'season' into the model
11
12 201 as a fixed effect, with statistical significance assessed using an F-test in a Type III test of
13
14 202 fixed effects. Season was defined as: spring – March to May; summer – June to August;
15
16 203 autumn – September to November; winter – December to February. Differences between
17
18 204 surveys were assessed by applying a t-test to an appropriate subset of the combined model
19
20 205 outputs.
21

22
23 206 These calculations make no adjustment for the sensitivity and specificity of the assay
24
25 207 therefore estimates can be considered as apparent prevalence throughout.
26

27 28 208 ***E.coli* O157 count data and verocytotoxin genes**

29
30
31 209 Descriptive statistics and count distributions were summarized for positive pats. Where
32
33 210 samples were positive but counts could not be enumerated, these were classified as below
34
35 211 enumeration limits (BEL). The probability of positive pats meeting the definition of super-
36
37 212 shedder was calculated for two classifications – a count of either greater than 10^3 CFU g⁻¹
38
39 213 faeces (SS3) or greater than 10^4 CFU g⁻¹ faeces (SS4) [20] – and compared between
40
41 214 surveys. At pat level, the presence of clustering due to a farm effect was assessed using a
42
43 215 Likelihood Ratio Test to compare models with and without a random 'farm' effect on
44
45 216 outcomes of interest relating to the pat-level descriptive analysis (*vtx* production, SS3 and
46
47 217 SS4 status). The odds of a farm having at least one pat that was SS3, SS4 or *vtx*-producing
48
49 218 were compared between surveys.
50

51 52 219 **Questionnaire data – descriptive analysis**

53
54 220 Questionnaire data were summarised and described. Non-normally distributed continuous
55
56 221 variables were transformed where appropriate. Categorical variables were treated as multi-
57
58 222 level factors; remaining variables were dichotomous. Season was defined as stated earlier.
59
60

1
2
3 223 Cattle management type had four levels: suckler beef (SB), specialist finisher (SF), dairy (D)
4
5 224 and other (Oth).

6
7
8 225 The association between size category – defined as median total cattle greater or less than
9
10 226 the median total cattle on sampled farms – and positive farm status was assessed using
11
12 227 logistic regression.

13 14 228 **Validity**

15
16
17 229 The potential for bias with regard to farm herd size and spatial location was assessed.

18
19 230 Registered herd size (obtained when identifying the sampling frames) was used for this
20
21 231 comparison as data were available for all farms.

22
23
24 232 Median herd size of sampled farms was compared to the same measure for i) the
25
26 233 denominator population; ii) all non-sampled farms; iii) farms that opted out; vi) farms that
27
28 234 were not phoned and v) farms that were reserved. Two definitions of denominator population
29
30 235 were used in England & Wales – a) farms available for phone recruitment and b) all farms in
31
32 236 the original sampling frame.

33
34
35 237 The potential for spatial bias was investigated using Nomenclature of Units for Territorial
36
37 238 Statistics (NUTS) [32]. Based on the distribution of sampling frame farms across NUTS 2
38
39 239 regions, the proportion of sampled farms within each NUTS 2 region was compared with the
40
41 240 expected proportion using Fisher's Exact Test. For England & Wales, many NUTS 2 regions
42
43 241 contained very few farms; a simulated *P*-value was therefore reported for the England &
44
45 242 Wales data. To check whether this might influence England & Wales results, Fisher's Exact
46
47 243 Test with simulated *P*-value was also performed on the Scottish data, to compare with the
48
49 244 calculated *P*-value.

245 Results

246 Farm visits

247 Sampling visits were completed by September 2015 in Scotland and by November 2015 in
248 England & Wales. The England & Wales extension related to recruitment difficulties during
249 spring 2015. One of the 111 Scottish farms visited was excluded from analyses due to
250 ineligibility as was visited in error and had not been sampled in previous surveys. Three of
251 the 163 England & Wales farms visited were excluded because transfer delays affected
252 sample viability.

253 Herd-level prevalence

254 *E. coli* O157 was detected on 26 Scottish farms and 34 farms in England & Wales. The
255 mean herd-level prevalence (95% confidence interval (CI)) of *E. coli* O157 was estimated at
256 0.236 (0.166 – 0.325) and 0.213 (0.156 – 0.283), respectively (Table 1, Fig. 1). This
257 difference was not statistically significant ($P = 0.65$).

258 In Scotland, there was no difference in the number of herds sampled in each season
259 ($P = 0.36$) whereas in England & Wales the seasonal sampling distribution was not uniform
260 ($P = 0.001$), with more samples taken in the autumn, the season with the lowest prevalence
261 estimate (Fig. 1). Within surveys there was no difference in seasonal herd-level prevalence
262 in England & Wales ($P = 0.92$), but in Scotland spring estimates were significantly lower than
263 autumn estimates ($P = 0.02$) (Fig. 1). Between surveys, autumn had the highest herd-level
264 prevalence in Scotland but the lowest in England & Wales ($P = 0.05$) (Table 2, Fig. 1).

265 Pat-level prevalence

266 The mean pat-level prevalence (95% confidence interval (CI)) of *E. coli* O157 was estimated
267 at 0.106 (0.067 – 0.163) for Scotland and 0.069 (0.044 – 0.107) for England & Wales
268 (Table 1). The difference between Scotland and England & Wales was not statistically
269 significant ($P = 0.19$). Within surveys there was no difference in seasonal pat-level

1
2
3 270 prevalence in England & Wales ($P = 0.60$), but in Scotland spring estimates were lower than
4
5 271 estimates for the other seasons ($P < 0.05$) (Fig. 1). Between surveys, the pat-level
6
7 272 prevalence in the autumn was low in England & Wales in comparison to Scotland
8
9 273 ($P = 0.003$) (Table 2 and Fig. 1).

10 11 12 274 ***E. coli* O157 count data and verocytotoxin genes**

13
14
15 275 Counts were determined for 287 *E. coli* O157 positive pats from Scotland and 234 from
16
17 276 England & Wales. The distributions were highly skewed, with the median count in both
18
19 277 surveys BEL. A subset of counts fell within SS3 and SS4 ranges (data not shown). At the
20
21 278 farm level, there was no difference between surveys regarding the odds of a positive farm
22
23 279 having at least one pat in either the SS3 or SS4 category (Table 3, Supplementary
24
25 280 Information (SI)). At the pat level, there was strong evidence of farm-level clustering within
26
27 281 both surveys ($P < 0.001$). There was no difference between surveys in the probability of a
28
29 282 positive pat having super-shedder status once farm-level clustering was accounted for
30
31 283 ($P = 0.97$ for SS3 and $P = 0.74$ for SS4).

32
33
34 284 On 25 of 26 positive Scottish farms, at least one isolate of *E. coli* O157 produced *vtx*,
35
36 285 compared to 29 of 34 positive farms in England & Wales ($P = 0.22$). At the farm level, there
37
38 286 was no difference between surveys regarding the odds of a positive farm having at least one
39
40 287 pat producing *vtx* (Table 3, SI). At the pat level, there was no difference found between
41
42 288 surveys once farm-level clustering was accounted for ($P = 0.84$). In both surveys, the
43
44 289 majority of positive isolates produced *vtx2* alone; *vtx1* appeared only with *vtx2* (Table 4, SI).

45 46 47 48 290 **Descriptive statistics – Questionnaire data**

49
50 291 All Scottish farms completed questionnaires ($n = 110$). One questionnaire from England &
51
52 292 Wales was incomplete ($n = 159$). Tables 5 to 8 (SI) give the univariable summary of
53
54 293 questionnaire results for Scotland and England & Wales. No adjustment for multiple
55
56 294 significance testing has been made.

1
2
3 295 The median ages of the youngest and oldest animals in the sampled groups, at 15 and 22
4 296 months in Scotland and 14 and 20 months in England & Wales, did not differ significantly
5
6 297 ($P = 0.18$ and $P = 0.28$, respectively).
7
8

9
10 298 Scottish farms were larger (median total cattle at sampling) ($P < 0.001$), had more cattle
11 299 aged 12-30 months ($P = 0.015$) and had larger sample groups ($P < 0.001$) than England &
12
13 300 Wales farms. There were within-survey correlations between all three of these measures
14
15 301 (Table 8, SI).
16
17

18
19 302 Few farms held organic status and distribution across management types was similar in both
20 303 surveys (Table 5, SI). There was no difference between Scotland and England & Wales
21 304 regarding health issues in the sampled group in the two weeks before sampling, or treatment
22 305 being given in the three months before sampling (Table 5, SI). Scottish farms were more
23 306 likely than those in England & Wales to have overwintered livestock owned by another
24 307 keeper in the year before sampling ($P = 0.002$) and to employ farm workers ($P < 0.001$).
25
26
27
28
29
30

31
32 308 Fewer Scottish sampled groups were grazing at sampling than in England & Wales
33 309 ($P = 0.003$). Compared to the autumn, sampled groups were more likely to be housed in
34 310 spring in Scotland ($P = 0.007$), and during the winter in both surveys ($P = 0.025$ Scotland,
35 311 $P = 0.002$ England & Wales). Bedding material was used in fewer Scottish housed groups
36 312 than in England & Wales ($P = 0.041$) (Table 6, SI).
37
38
39
40
41

42 313 No differences were found in relation to questions asked specifically for grazing sample
43 314 groups (Table 7, SI).
44
45

46 47 48 315 **Validity**

49
50
51 316 Scottish sampled farms did not differ in median herd size from the denominator population:
52 317 all farms in the original sampling frame (Table 9, SI).
53
54
55
56
57
58
59
60

1
2
3 318 In England & Wales sampled farms had larger median herd sizes than those in either
4
5 319 definition of the denominator population: farms available for phone recruitment (a); or all
6
7 320 farms in the original sampling frame (b) ($P < 0.01$) (Table 9, SI).
8
9
10 321 The 50% of the England & Wales farms that were largest in size, by total cattle numbers (i.e.
11
12 322 above the median), were more likely to test positive for *E. coli* O157 than the 50% that were
13
14 323 smallest in size (OR 3.652, $P = 0.003$). This effect was not seen in Scotland.
15
16
17 324 There was no difference in the proportional spatial distribution of Scottish denominator farms
18
19 325 and sampled farms across NUTS 2 regions ($P = 0.938$). The same was seen for both
20
21 326 definitions of denominator for England & Wales ($P = 0.865$ and $P = 0.781$). There was no
22
23 327 difference in calculated versus simulated P -value for this test on the Scottish data, therefore
24
25 328 it was considered acceptable to report the simulated value for England & Wales.
26
27
28

329 Discussion

30
31 330 In this study, for the first time, contemporaneous surveys have been completed in Scotland
32
33 331 and England & Wales to obtain prevalence estimates for *E. coli* O157 in cattle destined for
34
35 332 the food chain. The mean herd-level prevalence of *E. coli* O157 for the Scotland survey
36
37 333 (0.236 (0.166 – 0.325)) did not differ statistically from that in the England & Wales survey
38
39 334 (0.213 (0.156 – 0.283)). These estimates are similar to previous estimates for *E. coli* O157 in
40
41 335 Scotland [12], but lower than previous estimates for England and Wales [13, 14].
42
43

44 336 The use of randomisation and the recruitment software removed much of the potential for
45
46 337 recruitment selection bias, while the use of two main recruiters per survey with standardised
47
48 338 protocols reduced the potential for recruitment bias due to inter-operator differences. There
49
50 339 was no evidence for participation bias with regard to herd size, spatial location, or sampling
51
52 340 season in the Scotland survey. It can therefore be assumed that this is a valid estimate of
53
54 341 current apparent prevalence for the source population. The original surveys were designed
55
56 342 to be representative of the wider Scottish cattle population; whether this remains the case
57
58
59
60

1
2
3 343 more than a decade later is open to question. Of the 447 Scottish farms that participated in
4
5 344 both historical surveys 346 were still in business, with appropriate cattle **officially** recorded as
6
7 345 present. The **overall** size and geographical distribution of the Scottish National Herd (**SNH**)
8
9 346 has changed [33]. This could distort the current prevalence estimate if those changes are
10
11 347 systematically associated with the likelihood of a farm being *E. coli* O157 positive – or
12
13 348 factors that influence this – or with the reasons for the ineligibility of the no-longer-eligible
14
15 349 subset [34]. The authors consider this to be unlikely, **as there is no reason to believe that**
16
17 350 **changes in the SNH are likely to have affected the survey population differently to the non-**
18
19 351 **survey population, nor for them to be associated with *E. coli* O157 positive status. Firstly,**
20
21 352 **the main change in geographical distribution has been the contraction of the small proportion**
22
23 353 **of the overall number of cattle in the SNH that are within dairy herds, both in numbers and**
24
25 354 **geographically to the south west of Scotland. Secondly, the long term gradual decline in**
26
27 355 **overall cattle numbers has been evident since 1974 [33].** Thus the mean herd-level
28
29 356 prevalence for *E. coli* O157 is considered representative of the Scottish target population,
30
31 357 i.e. those farms keeping cattle destined for the food chain.

32
33
34 358 There has been no previous comparable survey in England & Wales. As the categories and
35
36 359 age groups for which data on cattle numbers are available have changed, the source
37
38 360 population defined was the best achievable approximation to the eligibility requirements of
39
40 361 the original Scottish survey [11]. Some farms included in the sampling frame may not have
41
42 362 had cattle relevant to this study, making them ineligible. Unless they opted out, this would
43
44 363 not have been discovered until they were contacted. Hence, the internal validity of the
45
46 364 survey was assessed against two definitions of denominator farms.

47
48
49 365 As for the Scotland survey, the potential for recruitment bias in the England & Wales survey
50
51 366 was minimised. There was no evidence for a spatial effect on participation, though smaller
52
53 367 farms in England & Wales were both less likely to be randomly selected for phoning and also
54
55 368 less likely to be sampled. Herd size distribution within this group did not differ statistically
56
57 369 significantly from the group of farms **from England & Wales** that opted out **initially**. As it is
58
59
60

1
2
3 370 unlikely that the lower likelihood of being sampled relates to the recruitment process, it is
4
5 371 unfortunate that the reason for opting out when phoned and contacted was not recorded.
6
7 372 This may have provided insight into whether this reflected ineligibility or disinterest. The
8
9 373 mean herd-level prevalence of *E. coli* O157 may have been over-estimated in England &
10
11 374 Wales, given that, as a single variable, larger herds were more likely to test positive for
12
13 375 *E. coli* O157 in this survey. Previously, herd size has been identified as a risk factor for
14
15 376 Scottish farms being positive for *E. coli* O157, where – amongst positive groups – larger
16
17 377 sample groups had lower mean within-group prevalence of shedding [11]. The opposite was
18
19 378 seen in a survey of young cattle in England & Wales [14]. In both surveys presented here,
20
21 379 there was a statistically significant difference in herd size between sampled farms and all
22
23 380 farms that opted out. This highlights a potential recruitment challenge when conducting
24
25 381 cross-sectional surveys that rely on **single time-point records for cattle numbers and**
26
27 382 voluntary farmer participation, **as it has** implications for estimating prevalence of any
28
29 383 condition that is known to be associated with herd size.

31
32 384 The statistically significant difference between the number of herds sampled across seasons
33
34 385 in England & Wales is likely to be a direct result of recruitment issues encountered during
35
36 386 the spring (Fig. 1). This meant that sampling extended into a second autumn period. If the
37
38 387 autumn season were a known risk factor, or should sampling year influence the likelihood of
39
40 388 a farm being positive, then this imbalance may have biased the overall England & Wales
41
42 389 herd-level prevalence estimate. Previously, decreased herd-level prevalence in winter and a
43
44 390 peak during the summer was found in Scottish herds, whilst housed status increased the
45
46 391 mean shedding prevalence at group level [11, 35]. A longitudinal study of young cattle in
47
48 392 England & Wales, however, found that winter was a risk period for shedding; it also
49
50 393 corroborated the reduced risk for cattle at pasture [36]. In this study, winter had the highest
51
52 394 herd-level and pat-level prevalence estimates for England & Wales, though seasonal
53
54 395 differences were not statistically significant within our survey. Seasonal effects can be
55
56 396 confounded by housing status due to management practices in the UK. In this study, the
57
58
59
60

1
2
3 397 lower proportion of farms sampled during the spring in England & Wales may have
4
5 398 decreased the herd-level prevalence estimate if housing were identified as a risk factor for
6
7 399 positive farm status and groups were more likely to be housed during that season (therefore
8
9 400 fewer housed groups were sampled than might have been expected), but this was not the
10
11 401 case.

12
13
14 402 The prevalence estimate for England & Wales is substantially lower than those reported
15
16 403 previously [13, 14]. Possible reasons for this include differences in how previous surveys
17
18 404 defined an eligible farm, their sampling approach, the distribution of herds across
19
20 405 management types and their seasonal distribution of sampling. The true prevalence may
21
22 406 also have genuinely decreased. Having considered the potential differences between the
23
24 407 current and previous approaches, the authors conclude that the estimated mean herd-level
25
26 408 prevalence for *E. coli* O157 can be considered representative of the current England &
27
28 409 Wales target population i.e. those farms with cattle destined for the food chain.

29
30
31 410 This study demonstrates that *E. coli* O157 remains relatively widespread among British
32
33 411 farms with cattle destined for the food chain.

34
35
36 412 No statistically significant difference was found between overall pat-level prevalence in
37
38 413 Scotland and in England & Wales. The mean pat-level prevalence estimates from previous
39
40 414 Scottish surveys were lower [12] than the current Scotland estimate, though this will be
41
42 415 investigated further in another study. There are no previous pat-level estimates for a similar
43
44 416 cattle population in England & Wales, although a sample-level prevalence of 7.7% for
45
46 417 VTEC O157 has also been described in young cattle, based on rectal sampling [36]. Pat-
47
48 418 level estimates will be a function of both the herd-level prevalence and the within-farm
49
50 419 prevalence. This study was not designed to fully explore multi-level risk factors, although
51
52 420 there is the potential for further analyses to investigate possible associations between
53
54 421 demographic or management factors and within-farm prevalence of *E. coli* O157. Given that
55
56
57
58
59
60

1
2
3 422 this study found strong evidence for farm-level clustering of super-shedder status and *vtx*
4
5 423 status, this will be an important question to pursue.

6
7
8 424 Several factors may influence pat-level prevalence: temporal patterns of shedding by
9
10 425 individual animals are known to vary [15, 16]; housing is associated with increased shedding
11
12 426 [35]; there is known heterogeneity of distribution of *E. coli* O157 in pats [22] and it has not
13
14 427 been possible to assess inter-operator differences within the current surveys, let alone
15
16 428 between studies over time. In addition, climatic effects may affect the survival of the
17
18 429 organism within pats [37, 38]. Any, some, or a multi-factorial combination of these may have
19
20 430 contributed to the overall pat-level prevalence estimates observed.

21
22
23 431 Despite the greater number of farms sampled in autumn in the England & Wales survey, the
24
25 432 prevalence estimate for this season remains low compared to Scotland. Overall, the
26
27 433 seasonal differences in herd-level and pat-level prevalence between the two surveys are
28
29 434 interesting, particularly in the autumn and spring seasons. Further investigation of the
30
31 435 *E. coli* O157 subtypes isolated from each survey may provide potential explanations for this
32
33 436 observation.

34
35
36 437 A high proportion of positive farms from this study harboured isolates producing *vtx*, both in
37
38 438 Scotland (0.962, 0.784 – 0.998) and in England & Wales (0.853, 0.682 – 0.945). However
39
40 439 six (1 in Scotland; 5 in England & Wales) did not, which may reflect evolution of the
41
42 440 persisting VTEC, as demonstrated in two Wisconsin dairy farms [39]. Differences in toxin
43
44 441 expression of the dominant strain of *E. coli* O157 on a given farm could also influence within-
45
46 442 farm prevalence. This finding, the lack of a statistical difference in *vtx* status between the
47
48 443 surveys, plus the lack of a statistical difference in super-shedder status warrants more in
49
50 444 depth investigation. The significance of super-shedder status (based on the SS3 definition)
51
52 445 has recently been questioned [40]. There is also discussion about how to define a super-
53
54 446 shedder (SS3 vs. SS4) [20, 21]. Regardless of whether it denotes a persistent characteristic
55
56 447 of the individual animal or a phase through which all colonized cattle pass, super-shedding
57
58
59
60

1
2
3 448 of *E. coli* O157 remains a public health issue through the introduction to the human
4
5 449 environment of potentially harmful bacteria [17]. The classification performed for this study –
6
7 450 into *vtx* 1 and 2 – will be augmented by investigating further subtyping of the toxin genes, the
8
9 451 phage types and genetic structure of *E. coli* isolates collected via whole genome sequencing
10
11 452 (WGS).

12
13
14 453 Over time, data from a 38 month long study in Swedish herds [41] demonstrated that, while
15
16 454 previous positive VTEC O157:H7 status was a predictor for current status, for the majority of
17
18 455 infected herds clearance of infection of occurred within a limited period. Over a matter of
19
20 456 years, data from the previous two Scottish surveys have demonstrated that prior
21
22 457 *E. coli* O157 status at farm level is not a predictor of current status [42]. The design of
23
24 458 BECS, where one of the objectives was to repeat sample a subset of Scottish farms for
25
26 459 temporal analysis, provides a unique opportunity to further extend this investigation, which
27
28 460 will be explored in future analyses.

29
30
31 461 These 2014/15 cattle surveys have obtained isolates of *E. coli* O157 currently circulating in
32
33 462 cattle in both Scotland and England & Wales, resulting in a unique collection. More detailed
34
35 463 classification of collected strains and comparison with those from contemporaneous human
36
37 464 clinical cases will give further insight into the relationship between circulating cattle and
38
39 465 human isolates. With access to historic libraries of both cattle and human isolates for WGS,
40
41 466 there is now the opportunity to investigate the evolution of this clonal type over the last two
42
43 467 decades in the UK and elucidate the genetic determinants underlying zoonotic potential,
44
45 468 such as variation in integrated prophages [43].

46
47
48 469 Only by determining the precise features of *E. coli* O157 that render it dangerous to humans
49
50 470 and establishing the most reliable means of identifying cattle strains that pose the greatest
51
52 471 risk will it be possible to target interventions appropriately within the cattle population and
53
54 472 thus mitigate that risk to human health.

1
2
3 473 While providing the foundation for these further investigations, this work has demonstrated
4
5 474 that *E. coli* O157 remains prevalent on British farms producing cattle for human
6
7 475 consumption. Until further work to identify and characterise circulating strains is completed,
8
9 476 public health messages should continue to outline the potential risk to human health from
10
11 477 contact with cattle and their environment.
12

13
14 478

15
16
17 479 *Figure 1. Mean seasonal prevalence estimates (solid triangles Scotland, solid dots England*
18
19 480 *& Wales) including 95% CI (horizontal bars) for the herd-level and pat-level prevalence of*
20
21 481 *E. coli* O157 in Scotland (blue) and in England & Wales (red) for farms sampled in Scotland
22
23 482 *(n=110) and England & Wales (n=160) between September 2014 and November 2015.*
24
25 483 *Integer values beside each dot indicate the total number of farms or pats, as appropriate,*
26
27 484 *sampled within each survey/season.*
28

29
30 485

31 32 33 34 486 **Acknowledgements**

35
36 487 The research team gratefully acknowledge the vital contribution of the farmers who
37
38 488 participated in this study and consented to having samples collected from their farms. SRUC
39
40 489 staff (J. Brennan, J. Baughan, K. Adam and S. Irvine) and RSK ADAS UK—Ltd. staff
41
42 490 (S. Lloyd, E. Young, D. Webb, K. Wheeler, W. John, C. Smith, S. McMillan, F. Tweedie,
43
44 491 N. Lee, T. Lathwood and E. Jones) are acknowledged for their involvement in and
45
46 492 contribution to recruitment and field sampling. M. Pearce is acknowledged for his advice on
47
48 493 previous Scottish surveys.
49

50 51 52 53 494 **Financial Support**

54
55 495 This work was funded by the Food Standards Agency (FSA) and Food Standards Scotland
56
57 496 (FSS), project number FS101055. MCT was supported through a Wellcome Trust Strategic
58
59
60

1
2
3 497 Award WT095831 to the Centre for Immunity, Infection and Evolution at the University of
4
5 498 Edinburgh.

6
7
8
9 499 **Conflict of Interests**

10
11 500 None.

12
13
14
15 501 **Ethical Standards**

16
17 502 Not applicable.

18
19
20
21
22 503 **References**

- 23 504 1. **Karmali MA, et al.** The association between idiopathic haemolytic uremic syndrome and
24 505 infection by verotoxin-producing *Escherichia coli*. *The Journal of Infectious Diseases* 1985:
25 506 **151**: 775-782.
- 26
27
28 507 2. **Willshaw GA, et al.** Verocytotoxin-producing *Escherichia coli* (VTEC) O157 and other
29 508 VTEC from human infections in England and Wales: 1995-1998. *Journal of Medical*
30 509 *Microbiology* 2001: **50**: 135-142.
- 31
32 510 3. **Schering J, Andreoli SP, Zimmerhacki LB.** Treatment and outcome of Shiga-toxin-
33 511 associated hemolytic uremic syndrome (HUS). *Pediatric Nephrology* 2008: **23**: 1749-1760.
- 34
35 512 4. **Trevena WB, et al.** Transmission of Vero cytotoxin producing *Escherichia coli* O157
36 513 infection from farm animals to humans in Cornwall and west Devon. *Communicable Disease*
37 514 *and Public Health* 1999: **2**: 263-268.
- 38
39 515 5. **O'Brien SJ, Adak GK, Gilham C.** Contact with farming environment as a major risk factor
40 516 for Shiga toxin (Vero cytotoxin)-producing *Escherichia coli* O157 infection in humans.
41 517 *Emerging Infectious Diseases* 2001: **7**: 1049-1051.
- 42
43 518 6. **Friesema I.** Geographical association between livestock density and human shiga toxin-
44 519 producing *Escherichia coli* O157 infections. *Epidemiology & Infection* 2011: **139**: 1081-1087.
- 45
46 520 7. **Matthews L, et al.** Super-shedding cattle and the transmission dynamics of *Escherichia*
47 521 *coli* O157. *Epidemiology & Infection* 2006: **134**: 131-142.
- 48
49 522 8. **Chase-Topping M, et al.** Super-shedding and the link between human infection and
50 523 livestock carriage of *Escherichia coli* O157. *Nature Reviews Microbiology* 2008: **6**: 904-
51 524 912.
- 52
53 525 9. **Matthews L, et al.** Predicting the public health benefit of vaccinating cattle against
54 526 *Escherichia coli* O157. *Proceedings of the National Academy of Sciences of the United*
55 527 *States of America* 2013: **110**: 16265-16270.
- 56
57 528 10. **Dallman TJ, et al.** Applying phylogenomics to understand the emergence of shiga-toxin-
58 529 producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK.

- 1
2
3 530 *Microbial Genomics* 2015: **1**: 1–13.
- 4
5 531 11. **Gunn GJ, et al.** An investigation of factors associated with the prevalence of
6 532 verocytotoxin producing *Escherichia coli* O157 shedding in Scottish beef cattle. *The*
7 533 *Veterinary Journal* 2007: **174**: 554–564.
- 8
9 534 12. **Pearce MC, et al.** Temporal and spatial patterns of bovine *Escherichia coli* O157
10 535 prevalence and comparison of temporal changes in the patterns of phage types associated
11 536 with bovine shedding and human *E. coli* O157 cases in Scotland between 1998-2000 and
12 537 2002-2004. *BMC Microbiology* 2009: **9**: 276.
- 13
14 538 13. **Paiba GA, et al.** Prevalence of faecal excretion of verocytotoxigenic *Escherichia coli*
15 539 O157 in cattle in England and Wales. *Veterinary Record* 2003: **153**: 347–353.
- 16
17 540 14. **Ellis-Iversen J, et al.** Identification of management risk factors for VTEC O157 in young-
18 541 stock in England and Wales. *Preventive Veterinary Medicine* 2007: **82**: 29–41.
- 19
20 542 15. **Smith RP, Paiba GA, Ellis-Iversen J.** Longitudinal study to investigate VTEC O157
21 543 shedding patterns in young cattle. *Research in Veterinary Science* 2010: **88**: 411–414.
- 22
23 544 16. **Robinson SE, et al.** Intermittent and persistent shedding of *Escherichia coli* O157 in
24 545 cohorts of naturally infected calves. *Journal of Applied Microbiology* 2004: **97**: 1045–1053.
- 25
26 546 17. **Locking M, Cowden J.** *Escherichia coli* O157. *British Medical Journal* 2009: **339**:
27 547 b4076.
- 28
29 548 18. **Naylor SW, et al.** Impact of the direct application of therapeutic agents to the terminal
30 549 recta of experimentally colonized calves on *Escherichia coli* O157:H7 shedding. *Applied and*
31 550 *Environmental Microbiology* 2007: **73**: 1493–1500.
- 32
33 551 19. **Halliday JEB, et al.** Herd-level risk factors associated with the presence of phage type
34 552 21/28 *E. coli* O157 on Scottish cattle farms. *BMC Microbiology* 2006: **6**: 99.
- 35
36 553 20. **Chase-Topping ME, et al.** Risk factors for the presence of high-level shedders of
37 554 *Escherichia coli* O157 on Scottish farms. *Journal of Clinical Microbiology* 2007: **45**:1594–
38 555 1603.
- 39
40 556 21. **Naylor SW, et al.** Lymphoid follicle-dense mucosa at the terminal rectum is the principal
41 557 site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host.
42 558 *Infection and Immunity* 2003: **71**: 1505–1512.
- 43
44 559 22. **Pearce MC, et al.** Distribution of *Escherichia coli* O157 in bovine fecal pats and its
45 560 impact on estimates of the prevalence of fecal shedding. *Applied and Environmental*
46 561 *Microbiology* 2004: **70**: 5737–5743.
- 47
48 562 23. **Evans J, et al.** Prevalence of *Escherichia coli* O157: H7 and serogroups O26, O103,
49 563 O111 and O145 in sheep presented for slaughter in Scotland. *Journal of Medical*
50 564 *Microbiology* 2011: **60**: 653–660.
- 51
52 565 24. **ISO/TS, 13136.** Microbiology of food and animal feed. Real-time polymerase chain
53 566 reaction (PCR)-based method for the detection of food-borne pathogens. Horizontal method
54 567 for the detection of Shiga-toxin producing *Escherichia coli* (STEC) and the determination of
55 568 O157, O11. 2012.
- 56
57 569 25. **SAS/STAT®.** 13.1 User's Guide The Glimmix procedure. Cary, NC: SAS Institute Inc,
58 570 2013: pp. 379.
- 59
60 571 26. **R Core Team.** R: A language and environment for statistical computing. R Foundation

- 1
2
3 572 for Statistical Computing, Vienna, Austria. 2015: <https://www.R-project.org/>. Accessed 30
4 573 January 2016.
- 5
6 574 27. **Fox J, Weisberg S.** *An {R} Companion to Applied Regression*, 2nd edn. Los Angeles:
7 575 SAGE Publications, Inc. 2011.
- 8
9 576 28. **Ripley B, Lapsley M.** RODBC: ODBC Database Access. 2015: [https://CRAN.R-](https://CRAN.R-project.org/package=RODBC)
10 577 [project.org/package=RODBC](https://CRAN.R-project.org/package=RODBC) Accessed 30 January 2016.
- 11
12 578 29. **Wickham H.** Reshaping data with the reshape package. *Journal of Statistical Software*
13 579 2007: **21**: 1–20.
- 14
15 580 30. **Wickham H.** *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag,
16 581 2009.
- 17
18 582 31. **Wickham H, Francois R:** dplyr: A Grammar of Data Manipulation. 2015:
19 583 <https://CRAN.R-project.org/package=dplyr> Accessed 30 January 2016.
- 20
21 584 32. **European Parliament, Council of the European Union.** Regulation (EC) No
22 585 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the
23 586 establishment of a common classification of territorial units for statistics (NUTS). *Official*
24 587 *Journal of the European Union L154* 2003: **46**: 1–41
- 25
26 588 33. **National Statistics.** Results from the June 2015 Scottish Agricultural Census. 2015:
27 589 <http://www.gov.scot/Publications/2015/10/6201/downloads> Accessed 30 July 2016.
- 28
29 590 34. **Thrusfield M.** The nature of data - Bias. In: Blackwell Science Ltd (ed). *Veterinary*
30 591 *Epidemiology*. Oxford: Blackwell Publishing, 2007, pp. 160.
- 31
32 592 35. **Synge BA, et al.** Factors influencing the shedding of verocytotoxin-producing
33 593 *Escherichia coli* O157 by beef suckler cows. *Epidemiology & Infection* 2003: **130**: 301–312.
- 34
35 594 36. **Smith RP, Pollitt WJ, Paiba GA.** A longitudinal study of risk factors for shedding of
36 595 VTEC O157 by young cattle in herds with known *E. coli* O157 carriage. *Epidemiology &*
37 596 *Infection* 2016: **144**: 1818–1829.
- 38
39 597 37. **Kudva IT, Blanch K, Hovde CJ.** Analysis of *Escherichia coli* O157:H7 Survival in Ovine
40 598 or Bovine Manure and Manure Slurry. *Applied and Environmental Microbiology* 1998: **64**:
41 599 3166–3174.
- 42
43 600 38. **Fukushima H, Hoshina K, Gomyoda M.** Long-Term survival of shiga toxin-producing
44 601 *Escherichia coli* O26, O111, and O157 in bovine feces. *Applied and Environmental*
45 602 *Microbiology* 1999: **65**: 5177–5181.
- 46
47 603 39. **Park D, et al.** Evolution of the stx2-encoding prophage in persistent bovine *Escherichia*
48 604 *coli* O157:H7 strains. *Applied and Environmental Microbiology* 2013: **79**: 1563-1572
- 49
50 605 40. **Spencer SEF, et al.** “Super” or just “above average”? Supershedders and the
51 606 transmission of *Escherichia coli* O157:H7 among feedlot cattle. *Journal of the Royal Society*
52 607 *Interface* 2015: **12**: 0446.
- 53
54 608 41. **Widren S, et al.** Longitudinal observational study over 38 months of verotoxigenic
55 609 *Escherichia coli* O157:H7 status in 126 cattle herds. *Preventive Veterinary Medicine* 2015:
56 610 **121**: 343-352
- 57
58 611 42. **Herbert LJ, et al.** *E. coli* O157 on Scottish cattle farms: evidence of local spread and
59 612 persistence using repeat cross-sectional data. *BMC Veterinary Research* 2014: **10**: 95.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

613 43. **Lupolova N, et al.** Support vector machine applied to predict the zoonotic potential of
614 *E. coli* O157 cattle isolates. *Proceedings of the National Academy of Sciences of the United*
615 *States of America* 2016; **113**: 11312-11371

616

For Review Only

Table 1: Estimates for mean herd-level and pat-level prevalence of *E.coli* O157 for cattle farms sampled in Scotland and England & Wales between September 2014 and November 2015.

Analysis level		Estimate	SE	Mean prevalence*	95% CI**	
		N				
Herd	Scotland	110	-1.174	0.225	0.236	0.166 – 0.325
	England & Wales	160	-1.310	0.193	0.213	0.156 – 0.283
Pat	Scotland	2763	-2.136	0.257	0.106	0.067 – 0.163
	England & Wales	2913	-2.598	0.241	0.069	0.044 – 0.107

* $1/(1+\text{EXP}(-\text{estimate}))$

** CI, Confidence interval; $1/(1+\text{EXP}(-\text{estimate} \pm (1.96*\text{SE})))$

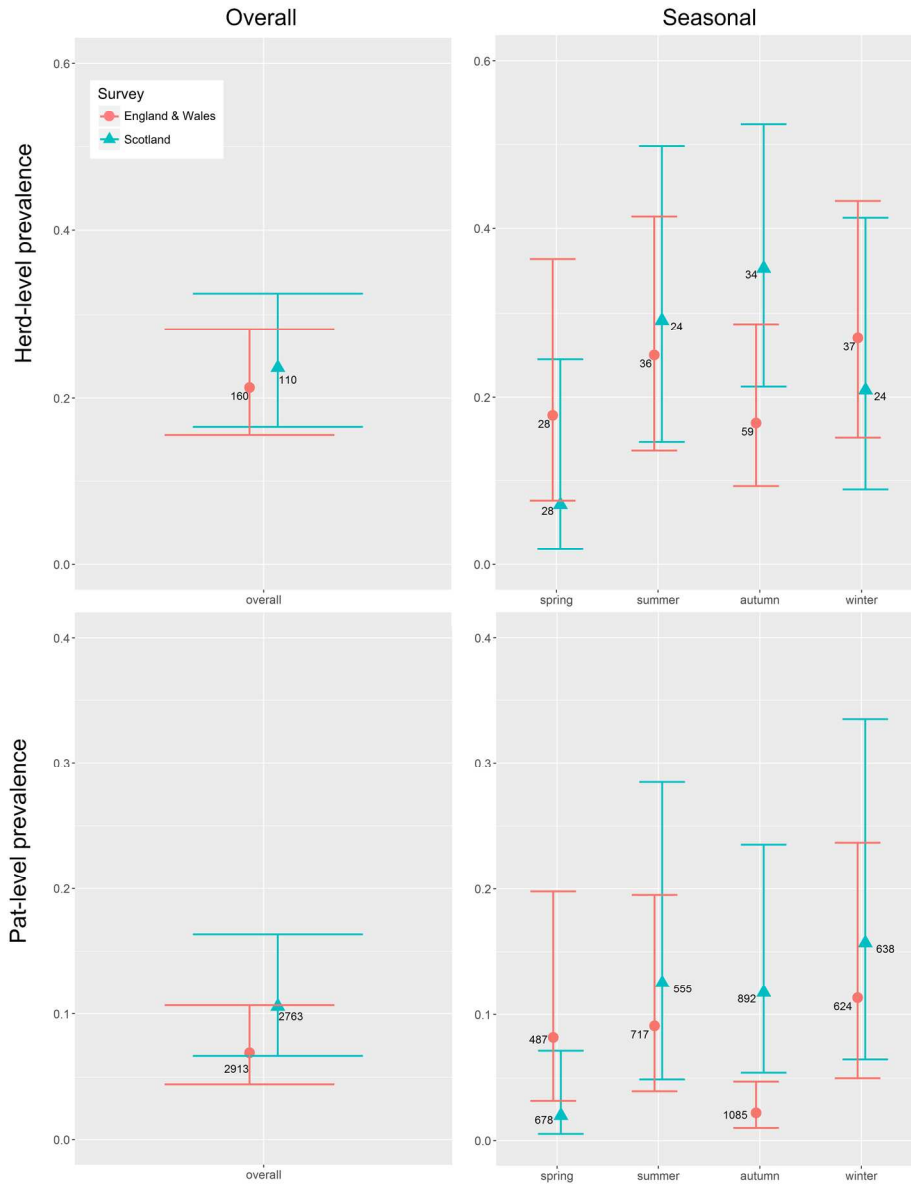


Figure 1. Mean seasonal prevalence estimates (solid triangles Scotland, solid dots England & Wales) including 95% CI (horizontal bars) for the herd-level and pat-level prevalence of E. coli O157 in Scotland (blue) and in England & Wales (red) for farms sampled in Scotland (n=110) and England & Wales (n=160) between September 2014 and November 2015. Integer values beside each dot indicate the total number of farms or pats, as appropriate, sampled within each survey/season.

173x224mm (300 x 300 DPI)

1 Table 2: Estimates for mean seasonal herd-level and pat-level prevalence of *E.coli* O157 for
 2 cattle farms sampled in Scotland (N=110) and England & Wales (N=160) between
 3 September 2014 and November 2015. P-value shown is for t-test of the difference between
 4 surveys.

Analysis level	Mean prevalence - proportion			P-value
	Season	Scotland	England & Wales	
Herd	Spring	0.071 [0.017 – 0.250]	0.179 [0.076 – 0.364]	0.248
	Summer	0.292 [0.144 – 0.502]	0.250 [0.136 – 0.415]	0.724
	Autumn	0.353 [0.210 – 0.528]	0.169 [0.094 – 0.287]	0.053
	Winter	0.208 [0.088 – 0.418]	0.270 [0.152 – 0.433]	0.589
Pat	Spring	0.020 [0.005 – 0.072]	0.082 [0.031 – 0.198]	0.085
	Summer	0.125 [0.049 – 0.284]	0.091 [0.040 – 0.195]	0.604
	Autumn	0.117 [0.054 – 0.235]	0.022 [0.001 – 0.047]	0.003
	Winter	0.157 [0.065 – 0.335]	0.113 [0.050 – 0.237]	0.577

5

Table 3 - Supplementary Information: Odds ratio (comparison between surveys: for Scotland compared to England & Wales) of an *E. coli* O157 positive farm having at least one *E. coli* O157 supershedder pat (SS3* or SS4** definition) or at least one pat producing vtx.

Analysis level	OR	P-value
	[95% C.I.]	
At least one SS3* pat present	1.0	1
	[0.3 – 3.6]	
At least one SS4** pat present	2.0	0.252
	[0.6 – 7.7]	
At least one pat producing vtx	5.2	0.126
	[0.6 – 255.7]	

*an *E. coli* O157 count of greater than 10^3 CFU g⁻¹ faeces (SS3); ** an *E. coli* O157 count of greater than 10^4 CFU g⁻¹ faeces (SS4) as per [20]

Table 4 – Supplementary Information: Description of positive *E. coli* O157 isolates according to *vtx* production, by survey.

Analysis level	<i>vtx</i> type	Number (proportion) that had this <i>vtx</i> type [95% CI]	
Survey		Scotland	England & Wales
<i>E. coli</i> O157 positive farms		26	34
<i>E. coli</i> O157 positive farms with isolates	<i>vtx</i> negative	1 (0.038) [0.002 – 0.216]	6 (0.176) [0.074 – 0.352]
	Any <i>vtx</i> present	25 (0.962) [0.784 – 0.998]	29 (0.853) [0.682 – 0.945]
	<i>vtx1</i> only	0 (0) [0.000 – 0.160]	0 (0) [0.000 – 0.126]
	<i>vtx2</i> only	22 (0.846) [0.643 – 0.950]	23 (0.676) [0.494 – 0.820]
	<i>vtx1</i> and <i>vtx2</i>	5 (0.192) [0.073 – 0.400]	7 (0.206) [0.087 – 0.379]
	<i>E. coli</i> O157 isolates		287
<i>E. coli</i> O157 isolates	<i>vtx</i> negative	1 (0.003) [0.000 – 0.022]	40 (0.171) [0.126 – 0.227]
	Any <i>vtx</i> present	286 (0.997) [0.978 – 1.000]	194 (0.829) [0.773 – 0.874]
	<i>vtx1</i> only	0 (0) [0.000 – 0.016]	0 (0) [0.000 – 0.020]

	<i>vtx2</i> only	210 (0.732)	133 (0.568)
		[0.676 – 0.781]	[0.502 – 0.632]
	<i>vtx1</i> and <i>vtx2</i>	76 (0.265)	61 (0.261)
		[0.215 – 0.321]	[0.207 – 0.323]

Note: proportion of farms does not sum to 1 per survey as farms could have several isolates producing different vtx types.

For Review Only

Table 5 – Supplementary Information: Description by survey and comparison of questionnaire data for the variables common to *all sampled* groups.

Survey	Number (proportion) of farms		P-value for difference between surveys	
	Scotland N=110	England & Wales N=159		
Management type	Suckler Beef	73 (0.663)	109 (0.681)	0.690
	Dairy	14 (0.127)	17 (0.106)	0.699
	Specialist Finisher	13 (0.118)	18 (0.113)	1
	Other	10 (0.091)	15 (0.094)	1
Sample season*	Spring	28 (0.255)	28 (0.175)	0.128
	Summer	24 (0.218)	37 (0.225)	1

	Autumn	34 (0.309)	59 (0.369)	0.362
	Winter	24 (0.218)	37 (0.231)	0.883
Farm has organic status		5 (0.045)	3 (0.019)	0.278
Cattle moved onto farm in the past 12 months		87 (0.791)	121 (0.756)	0.657
Farm has shared a breeding bull in the past 12 months		19 (0.173)	16 (0.100)	0.098
Livestock other than cattle purchased in the past 12 months		60 (0.545)	88 (0.550)	0.902
Livestock overwintered in the past 12 months		33 (0.300)	22 (0.138)	0.002
Livestock currently present on farm that are not owned by the farmer		19 (0.173)	22 (0.138)	0.492
Organic waste from own farm spread in the past 12 months		85 (0.773)	121 (0.756)	0.884
Organic waste from other farm(s) spread in the past 12 months		5 (0.045)	9 (0.056)	0.786
Cows calve on the farm		97 (0.882)	133 (0.831)	0.379
Cattle known to have access to water from a natural water source		75 (0.682)	108 (0.675)	1

Employ farm workers in addition to main household	64 (0.582)	51 (0.319)	<0.001
Sample group had access to grazing	27 (0.245)	68 (0.425)	0.003
Health problems seen in the sample group in the past 2 weeks	7 (0.064)	7 (0.044)	0.579
Treatments used on cattle in the sample group in the past 3 months	41 (0.373)	56 (0.350)	0.796

*Values for sample season in England & Wales sum to 160 because this was known without completion of the questionnaire, as it relates to sampling date.

Table 6 – Supplementary Information: Description by survey and comparison of questionnaire data for the variables that relate to the **housed** sample groups.

Variable	Number (proportion) of farms		P-value for difference between surveys	
	Scotland N=83	England & Wales [^] N=92		
Type of housing				
	Straw courts	63 (0.759)	79 (0.859)	0.121
(more than one could be selected)	Slats	11 (0.133)	4 (0.043)	0.056
	Byre	7 (0.084)	3 (0.033)	0.195
	Other	4 (0.048)	7 (0.076)	0.542
Feeding changed in the past 2 weeks		21 (0.253)	18 (0.196)	0.371
Location changed in the past 2 weeks		15 (0.181)	16 (0.174)	1
New animals added in the past 2 weeks		10 (0.120)	8 (0.087)	0.619
Bedding used where sample group housed*		73 (0.880)	89 (0.967)	0.041
	All old bedding removed prior to housing (N= *)	63 (0.863)	69 (0.775)	0.163
	Wet bedding removed since housing (N= *)	35 (0.479)	53 (0.596)	0.156
	New bedding added since housing (N= *)	69 (0.945)	84 (0.944)	1

Group had nose-to-nose contact with other cattle under 12 months	26 (0.313)	20 (0.217)	0.080
--	------------	------------	-------

[^] *England & Wales data includes one group that had access to both housing and grazing*

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 7 – Supplementary Information: Description by survey and comparison of questionnaire data for the variables that relate to the **grazing** sampled groups

Variable	Number (proportion) of farms		P-value for difference between surveys
	Scotland N=27	England & Wales^ N =68	
Feeding changed in the past 2 weeks	6 (0.222)	12 (0.176)	0.578
Location changed in the past 2 weeks	6 (0.222)	15 (0.221)	1
Grazing shared with other cattle	5 (0.185)	8 (0.118)	0.509
Grazing shared with other livestock species	4 (0.148)	15 (0.221)	0.573
Grazing ground had been cut in the past 2 weeks	2 (0.074)	3 (0.044)	0.621

[^]England & Wales data includes one group that had access to both housing and grazing.

Table 8 – Supplementary Information. Median values and correlation estimates for numbers of cattle in three groups: total cattle on farm, total cattle aged between 12 and 30 months, total cattle in the sample group. Pearson's Product-Moment Correlation calculated on (log-transformed values +1) for each of the three independent variables.

	Scotland	England & Wales
	N=110	N=159
Median cattle (range)		
on farm at sampling	176 (6 – 849)	85 (2 – 990)
12-30 months	31 (0 – 400)	17 (0 – 260)
in sample group	17 (1 – 90)	14 (1 – 125)
Correlation [95% C.I.] (P)		
between total cattle on farm and total cattle 12-30 months	0.492 [0.338 – 0.622] (<0.001)	0.580 [0.467 – 0.675] (<0.001)
between total cattle on farm and total cattle in sample group	0.284 [0.103 – 0.448] (0.003)	0.535 [0.414 – 0.638] (<0.001)
between total cattle 12-30 months and total cattle in sample group	0.202 [0.015 – 0.375] (0.034)	0.485 [0.356 – 0.598] (<0.001)

Table 9 – Supplementary Information. Validity: comparison of registered herd size for various groups in each survey.

Analysis level	Group	n	Median herd size	P-value for within-survey difference in herd size compared to sampled farms
Scotland	Sampled farms	110	186	-
	(i) Denominator population	346	155	0.198
	(ii) All non-sampled farms	236	140	0.074
	(iii) Farms that opted out at either the preliminary letter, or phone stage	99	103	0.002
	(iv) Farms that were not phoned	100	162	0.775
	(v) Farms that were reserved	37	166	0.839
England & Wales	Sampled farms	160	81	-
Wales	(i) Denominator (a): Farms available for phone recruitment	848	56	0.008
	(i) Denominator (b): All farms in the original sampling frame	1264	55	0.004
	(ii) All non-sampled farms	1101	51	0.001
	(iii) Farms that opted out at either the preliminary letter, or phone stage	668	52	0.001
	(iv) Farms that were not phoned	282	44	<0.001

(v) Farms that were reserved	151	87	0.937
Farms that were excluded from analyses due to sample transfer delay	3	-	-

For Review Only