



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Clark, NJ;Tozer, S;Wood, C;Firestone, SM;Stevenson, M;Caraguel, C;Chaber, AL;Heller, J;Soares Magalhães, RJ

Title:

Unravelling animal exposure profiles of human Q fever cases in Queensland, Australia, using natural language processing

Date:

2020-09-01

Citation:

Clark, N. J., Tozer, S., Wood, C., Firestone, S. M., Stevenson, M., Caraguel, C., Chaber, A. L., Heller, J. & Soares Magalhães, R. J. (2020). Unravelling animal exposure profiles of human Q fever cases in Queensland, Australia, using natural language processing. *Transboundary and Emerging Diseases*, 67 (5), pp.2133-2145. <https://doi.org/10.1111/tbed.13565>.

Persistent Link:

<https://hdl.handle.net/11343/275661>

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

DR. NICHOLAS CLARK (Orcid ID : 0000-0001-7131-3301)

Article type : Original Article

## Unravelling animal exposure profiles of human Q fever cases in Queensland, Australia using natural language processing

### Running title

Text mining to deduce Q fever exposure pathways

Nicholas J Clark<sup>1</sup>, Sarah Tozer<sup>3</sup>, Caitlin Wood<sup>1</sup>, Simon M. Firestone<sup>4</sup>, Mark Stevenson<sup>4</sup>, Charles Caraguel<sup>5</sup>, Anne-Lise Chaber<sup>5</sup>, Jane Heller<sup>6</sup>, Ricardo J. Soares Magalhães<sup>1,2</sup>

<sup>1</sup> UQ Spatial Epidemiology Laboratory, School of Veterinary Science, The University of Queensland, Gatton 4343, Queensland, Australia

<sup>2</sup> Children Health and Environment Program, Child Health Research Centre, The University of Queensland, South Brisbane 4101, Queensland, Australia

<sup>3</sup> Queensland Centre for Gynaecological Cancer, The University of Queensland, Herston 4029, Queensland, Australia

<sup>4</sup> Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville 3010, Victoria, Australia

<sup>5</sup> School of Animal and Veterinary Science, The University of Adelaide, Roseworthy 5371, Adelaide, South Australia, Australia

<sup>6</sup> Graham Centre for Agricultural Innovation, School of Animal and Veterinary Sciences, Charles Sturt University, Wagga 2650, New South Wales, Australia

### Abstract

Q fever, caused by the zoonotic bacterium *Coxiella burnetii*, is a globally distributed emerging infectious disease. Livestock are the most important zoonotic transmission sources, **This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/TBED.13565](https://doi.org/10.1111/TBED.13565)**

33 yet infection in people without livestock exposure is common. Identifying potential exposure  
34 pathways is necessary to design effective interventions and aid outbreak prevention. We used  
35 natural language processing and graphical network methods to provide insights into how Q  
36 fever notifications are associated with variation in patient occupations or lifestyles. Using an  
37 18-year time-series of Q fever notifications in Queensland, Australia, we used topic models  
38 to test whether compositions of patient answers to follow-up exposure questionnaires varied  
39 between demographic groups or across geographical areas. To determine heterogeneity in  
40 possible zoonotic exposures, we explored patterns of livestock and game animal co-  
41 exposures using Markov Random Fields models. Finally, to identify possible correlates of Q  
42 fever case severity, we modelled patient probabilities of being hospitalised as a function of  
43 particular exposures. Different demographic groups consistently reported distinct sets of  
44 exposure terms and were concentrated in different areas of the state, suggesting the presence  
45 of multiple transmission pathways. Macropod exposure was commonly reported among Q  
46 fever cases, even when exposure to cattle, sheep or goats was absent. Males, older patients  
47 and those that reported macropod exposure were more likely to be hospitalised due to Q fever  
48 infection. Our study indicates that follow-up surveillance combined with text modelling is  
49 useful for unravelling exposure pathways in the battle to reduce Q fever incidence and  
50 associated morbidity.

51

## 52 **Keywords**

53 Australia, *Coxiella burnetii*, Markov Random Fields, text mining, topic models, Q fever,  
54 zoonosis

## 55 **Introduction**

56 Q fever is a globally distributed emerging infectious disease caused by the bacterium  
57 *Coxiella burnetii* (Allan-Blitz, Sakona, Wallace, & Klausner, 2018; Bond, Franklin, Sutton,  
58 Stevenson, & Firestone, 2018; Gyuranecz et al., 2014; Van der Hoek et al., 2010). Acute  
59 infection with *C. burnetii* is commonly described as a flu-like illness with symptoms  
60 including high fevers, headaches or pneumonia, as well as atypical symptoms such as  
61 hepatitis or myocarditis (Didier Raoult & Marrie, 1995; Sellens et al., 2018). However, up to  
62 60% of human cases are thought to be asymptomatic (Roest et al., 2011). Infection by *C.*  
63 *burnetii* rarely causes mortalities but can manifest as a wide spectrum of recurrent, focalized  
64 morbidities that result in debilitating conditions involving the cardiovascular system or lungs  
65 (Fenollar et al., 2001; Million & Raoult, 2017).

66 Infected livestock, particularly goats and sheep, are the most important sources of  
67 zoonotic Q fever outbreaks in humans (Arricau-Bouvery & Rodolakis, 2005; N. Clark &  
68 Soares Magalhães, 2018). Inhalation of robust and infective small cell variants (SCVs;  
69 sometimes referred to as ‘spores’) is the primary mode of animal-to-human transmission for  
70 *C. burnetii* (D Raoult, Marrie, & Mege, 2005). Large quantities of SCVs may occur in animal  
71 faeces, vaginal mucus, products of conception and unpasteurized dairy products (Guatteo et  
72 al., 2006). People whose occupations involve close livestock contact (e.g. abattoir workers,  
73 livestock transporters and veterinarians) are considered at highest risk of infection (Graves &  
74 Islam, 2016; Karagiannis et al., 2009; Mori & Roest, 2018; Van der Hoek et al., 2010).

75 In Australia, human Q fever infection has been a notifiable disease in all states and  
76 territories since 1977. Human notification rates of Q fever in Australia are amongst the  
77 highest in the world (Gidding, Wallace, Lawrence, & McIntyre, 2009; Lindsay, Rohailla, &  
78 Miyakis, 2018; Sloan-Gardner, Massey, Hutchinson, Knope, & Fearnley, 2017). Australia  
79 also has the world’s only licensed human vaccine for Q fever (Q-Vax®, Seqirus Limited,  
80 VIC, Australia). Across the years 2001 – 2006 inclusive, the Australian Government funded a  
81 National Q fever Management Program, which involved screening and vaccination for  
82 specific at-risk populations including abattoir workers and livestock farmers (Gidding et al.,  
83 2009). This led to notable decreases in human Q fever notifications in subsequent years,  
84 particularly for the states of Queensland and New South Wales where the majority of  
85 Australian notifications occur (Karki, Gidding, Newall, McIntyre, & Liu, 2015; Sloan-  
86 Gardner et al., 2017).

87 Despite commendable vaccination and education efforts, Q fever persists as a public  
88 health concern in Australia (Lindsay et al., 2018; Sivabalan, Saboo, Yew, & Norton, 2017; S.  
89 Tozer, Lambert, Sloots, & Nissen, 2011). Moreover, recent notifications are now commonly  
90 attributed to people with no previous record of occupational exposure to risks associated with  
91 regular livestock contact, suggesting other transmission pathways may play roles in the  
92 epidemiology of the disease (N. Clark & Soares Magalhães, 2018; Reedijk, Van Leuken, &  
93 Van Der Hoek, 2013; Sloan-Gardner et al., 2017; S. Tozer et al., 2011). These underexplored  
94 transmission routes may differ substantially among people that live or work in different  
95 sectors (Clutterbuck, Eastwood, Massey, Hope, & Mor, 2018). *Coxiella burnetii* can persist  
96 in the environment, is resistant to harsh conditions and may be transported over long  
97 distances on prevalent winds (N. Clark & Soares Magalhães, 2018; Fitzpatrick, Kersh, &  
98 Massung, 2010; Reedijk et al., 2013). Coupled with the bacterium’s long incubation period of  
99 up to 4 – 6 weeks (Didier Raoult & Marrie, 1995), these aspects of Q fever epidemiology

100 make it difficult to investigate relevant exposure pathways. Nevertheless, a diversity of  
101 possible wildlife reservoirs has been identified through molecular and serological surveys,  
102 including wild and domestic mammals, birds and even ticks (Alanna Cooper, Barnes, Potter,  
103 Ketheesan, & Govan, 2012; A Cooper, Stephens, Ketheesan, & Govan, 2013; Flint et al.,  
104 2016; Webster, Lloyd, & Macdonald, 1995). Among these, macropods (including kangaroos,  
105 wallabies and wallaroos from the family Macropodidae) are of particular interest because (1)  
106 they are abundant and often share habitats with Australian livestock; (2) they are a common  
107 source of game meat for both humans and companion animals (Hoffman & Cawthorn, 2012);  
108 and (3) a range of macropod species have been documented as possible reservoir hosts of the  
109 bacterium using serological or molecular evidence (Banazis, Bestall, Reid, & Fenwick, 2010;  
110 Alanna Cooper et al., 2012).

111 Patients diagnosed with Q fever in the state of Queensland, Australia are interviewed  
112 with a series of questions designed to document and investigate possible transmission  
113 pathways. A questionnaire is completed over the telephone (within five days of positive Q  
114 fever confirmation) and contains fields including onset date, demographics (age, gender,  
115 indigenous status), occupation and several text-based fields aimed at describing the possible  
116 pathways of exposure to livestock / game animals (see **Appendix S1** for the full  
117 questionnaire template). In 2012, an extended surveillance form was introduced to allow  
118 patients to more directly list all possible animal exposures, adding an additional layer of rich  
119 enhanced surveillance data.

120 Identifying potential exposure pathways in people with confirmed Q fever infection is  
121 a key step to reduce disease incidence and severity. In this study, we applied natural language  
122 processing to an 18-year dataset of Q fever notifications in Queensland, Australia to  
123 investigate whether patients belonging to different demographic groups commonly report  
124 different potential exposure pathways. We then used multivariate graphical models to explore  
125 associations among reported animal-based exposures. Finally, we used infection-related  
126 hospitalisation records as a proxy for disease severity to test whether patients with reports of  
127 particular types of animal exposures suffer from more severe acute Q fever infections.

## 128 129 **Methods**

### 130 Ethics statement

131 This research used data on Q fever notifications collected by the Queensland Department of  
132 Health in accordance with Section 284 of the *Public Health Act 2005* and was completed

133 under the ethical approval of The Children’s Health Queensland Human Research Ethics  
134 Committee (HREC08/QRCH/66AM03 8/05/2017).

135

136 A state-wide dataset of Q fever notifications from Queensland, Australia

137 The primary data for this study encompassed all available human cases of Q fever infection  
138 notified to the Queensland Department of Health from 1 July 1984 to 31 December 2017  
139 inclusive. According to national guidelines, human cases of Q fever must be confirmed using  
140 either laboratory definitive evidence or a combination of laboratory suggestive evidence and  
141 clinical evidence (Communicable Diseases Network Australia, 2018). Laboratory definitive  
142 evidence includes either (1) detection of *C. burnetii* by nucleic acid testing, (2)  
143 seroconversion or significant increase in antibody level to Phase II antigen in paired sera  
144 tested coupled with the absence of recent Q fever vaccination, or (3) detection of *C. burnetii*  
145 by culture. Laboratory suggestive evidence refers to detection of specific IgM in the absence  
146 of recent vaccination. The full dataset contained 7,495 Q fever notifications. We geocoded  
147 patient addresses to describe spatial variation in exposure reports. Based on available  
148 information, geocodes were taken from one of three hierarchical levels representing (from  
149 most to least precise) house number and street name of the patient’s address (n = 4,217),  
150 centroid of the street (n = 784) or centroid of the Statistical Local Area (SLA; n = 2,494).

151 We filtered the data to only include cases from the years 2001 – 2017, as follow-up  
152 questionnaires became mandatory in 2001. This reduced dataset contained 4,068 individual Q  
153 fever notifications. Patients in this dataset had a median age of 39 years (interquartile range  
154 [IQR]: 27 – 52 years). Males accounted for 74% of notifications. Because reported animal  
155 exposures were more data-rich following the rollout of the improved surveillance form in  
156 2012, we created a separate dataset containing only the 2012 – 2017 data (n = 979) for  
157 comparisons in co-exposure analyses (see “Identifying animal exposures and co-exposures  
158 using Markov Random Fields” below and see **Figure S1** in Supporting Information for a  
159 flowchart of observations used in each step of analysis).

160

161 Building an exposure dataset using text mining

162 We applied natural language processing to all open-ended (text-based) question fields to  
163 construct an exposure dataset whereby each patient’s responses were represented as a distinct  
164 text unit. A series of quality control steps were used to correct spelling errors and filter out  
165 uninformative terms. Briefly, we first removed numerics and filtered out stop words (i.e.

166 words that are very common and are consequently considered unimportant for search queries,  
167 such as ‘the’, ‘about’ or ‘said’; Fox, 1989). Next, we singularized words (i.e. by changing  
168 ‘kangaroos’ to ‘kangaroo’) and applied a fuzzy pattern matching spell check algorithm that  
169 suggests replacements for misspelled words using a United States English language  
170 dictionary. Finally, we removed words containing fewer than three letters. Words reported by  
171 a total of 2,044 individual patients were included after these filtering steps (**Figure S1**). Text  
172 processing was carried out in R version 3.3.3 (R Core Team, 2018) and primarily used  
173 functions from the packages *tidytext* (Silge & Robinson, 2016), *hunspell* (Ooms, 2017) and  
174 *tidyverse* (Wickham, 2017).

175

176 Latent Dirichlet Allocation to identify discriminatory response ‘topics’

177 We applied a topic model algorithm, also known as Latent Dirichlet Allocation, to our  
178 exposure dataset to ask whether the composition of words in a patient’s responses could  
179 provide information about their demographic features. Topic models are a class of generative,  
180 unsupervised machine learning methods designed to identify latent ‘topics’ containing similar  
181 term compositions and frequencies within a given collection of texts (Blei, 2012; Hornik &  
182 Grün, 2011). This is accomplished with a mixture model whereby word frequencies in each  
183 latent topic are drawn from an unknown Dirichlet distribution (Blei, Ng, & Jordan, 2003).  
184 We pooled text from individual cases into eight demographic groups representing different  
185 sex and age classes (**Table 1**). Age categories were chosen to represent school-age children  
186 (ages 0 – 18), working age young adults (ages 19 – 34), working age older adults (ages 35 –  
187 64) and retirees (ages 65 – 100), considering differing potential exposure risks. We did not  
188 have a sufficient sample size to analyse data from children under five years of age separately  
189 (only 16 pre-school age individuals were in the notification data). Words that were  
190 represented fewer than five times were removed to ensure we focused only on terms likely to  
191 be useful for discriminating between demographic groups.

192 We fit topic models to the resulting term matrix, which contained 14,338 observations  
193 for 397 unique terms. Because the number of word topics ( $k$ ) must be specified prior to fitting  
194 the model, we used a data-driven approach to identify the optimal number. We tested six  
195 topic models ( $k = 2 - 7$ ) and compared each model’s geometric mean per-word likelihood  
196 (also known as perplexity; Hornik & Grün, 2011). The model that minimized inverse  
197 perplexity while containing the fewest number of topics was considered the most  
198 parsimonious (see **Figure S2** in Supporting Information for perplexity scores from each  
199 tested model). From the best-fitting model, we calculated the relative contribution of each

200 topic to each demographic group's response document to assess whether patients from  
201 different demographics provided different sets of responses. Topic models were fit using  
202 functions in the *topicmodels* R package (Hornik & Grün, 2011).

203

204 Identifying potential animal exposures and co-exposures using Markov Random Fields

205 We next determined whether free-text fields included information that might represent

206 potential exposure to particular livestock species (cattle, sheep, goats or pigs) or macropods.

207 This involved searching through patient answers for key terms associated with each of these

208 target host species (e.g. 'cattle', 'heifer' or 'beef' for potential exposure to cattle; see **Table**

209 **S1** in Supporting Information for a full list of search terms for each target host group).

210 Mentions of these target species were recorded as binary indicator variables. A total of 1,380

211 cases mentioned exposure to at least one of the five target host species (**Figure S1**). Note that

212 multiple binary fields exist for detecting sheep exposure, such as 'work with wool' or 'work

213 in a shearing shed' (**Table S1**), and so detections of sheep exposure may have higher

214 accuracy. Excluding these binary survey questions resulted in a total of 313 sheep exposures

215 detected, compared to 384 with the binary fields included. We chose to use the full dataset

216 for all analyses, though we recognise this slight potential for bias towards sheep exposure

217 detection.

218 We fit a Markov Random Fields (MRF) model to our matrix of binary animal

219 exposures (N. J. Clark, Wells, & Lindberg, 2018a). This framework, commonly used in

220 multivariate classification problems (Fountain-Jones et al., 2019; Harris, 2016), is well-suited

221 to our exploration of exposure pathways as it allows us to ask whether pairs of animal

222 exposures were more or less likely to be jointly reported (co-exposure) after accounting for

223 all other types of animal exposure (e.g. are pairs of exposures conditionally associated after

224 accounting for all other exposures in the graph?). In our model, each type of animal exposure

225 was included as a node in the undirected network, with edges between nodes representing the

226 marginal relationships between pairs of reported exposures adjusting for all other

227 relationships present (N. J. Clark, Wells, & Lindberg, 2018b). We also included an additional

228 binary node representing whether or not the patient was hospitalised due to Q fever infection,

229 allowing us to ask whether certain animal exposures were more or less likely to be

230 statistically associated with hospitalisation. Conditional relationships were estimated using a

231 regularized node-wise regression approach through functions in the *MRFcov* R package (N. J.

232 Clark et al., 2018a). We fit a separate MRF using only the 2012 – 2017 data ( $n = 979$ ) to

233 investigate whether the extended surveillance questionnaire led to different estimates of co-  
234 exposure relationships.  
235  
236 Regression models to identify associations with hospitalisation probability  
237 The above analyses explore patient exposure reports and how exposures may be related to  
238 one another. We supplemented these models by fitting a series of supervised machine  
239 learning regressions to identify important exposure correlates of a Q fever patient's  
240 probability of being hospitalised. Tested covariates were: sheep exposure, cattle exposure,  
241 macropod exposure, goat exposure, pig exposure, shooting / hunting participation,  
242 Indigenous status, age, sex, whether the patient was previously vaccinated and whether the  
243 patient reported that they previously assisted in animal births. All exposure / activity  
244 questions related to the month prior to the onset of illness. In addition, we constructed a  
245 binary variable to distinguish between pre-2012 and 2012-present observations to account for  
246 possible differences between the two time periods. We first fit two regularized spatial logistic  
247 regression models that applied a coordinated gradient descent LASSO regularization  
248 algorithm to select important predictor variables (Friedman, Hastie, & Tibshirani, 2010): the  
249 first included latitude and longitude as covariates; the second expanded these coordinates into  
250 Gaussian process spatial regression splines (Kammann & Wand, 2003) to account for non-  
251 linear spatial patterns. In addition, we accounted for possible non-linearity in predictor-  
252 outcome relationships by fitting a generalised additive logistic model that also included  
253 spatial regression splines (Wood, 2003). For each of these three competing models, we  
254 calculated their predictive performance using a cross-validation process that involved fitting  
255 models to a random subset of the data containing ~80% of observations (~1104 individuals)  
256 and calculating prediction accuracy (e.g. proportion of observations correctly predicted) for  
257 the remaining ~ 20% of observations. This cross-validation process was repeated 100 times to  
258 quantify uncertainty in predictive performances. Regressions were fit using functions in the  
259 *glmnet* and *mgcv* R packages (Friedman et al., 2010; Wood, 2003). Due to the sensitivity of  
260 the notification data, raw data is not publicly available. However, word strings for each  
261 demographic group and R scripts to replicate the topic model and posthoc analyses are  
262 available in the Supporting Information.

263

## 264 **Results**

265 Topic model analysis

266 Our text-mining dataset contained 14,337 words reported by 2,044 patients from eight  
267 demographic groups (**Table 1**). The richest sources of words came from working-age patients  
268 (19 – 34 and 35 – 64 years old), with both sexes well represented in the dataset (**Table 1**).  
269 The best-fitting topic model identified five word ‘topics’ that could successfully classify Q  
270 fever patients into categories based on their age group and sex (**Figure 1**). While a total of  
271 397 words were included in the analysis, we identified some key influential terms that  
272 achieved high discriminatory power. In other words, the presence of these terms in a patient’s  
273 answers likely represented important differences in the lifestyles and/or exposure pathways  
274 exhibited by demographic groups. Differences in answers between the two sexes were  
275 apparent, as two of the five identified topics were almost entirely associated with males. We  
276 describe each of the topic groups and some of their key discriminatory terms in detail below.

277 Topic 1: the most easily distinguishable demographic group, which belonged almost  
278 entirely to this topic, consisted of working age males (ages 19 – 34 years). This group  
279 consistently mentioned informative occupational terms associated with the livestock trade,  
280 such as ‘export’, ‘abattoir’, ‘feedlot’, ‘beef’, ‘kill’ and ‘process’, that were not commonly  
281 mentioned by other demographic groups (**Figure 1**).

282 Topic 2: children from both sexes (ages 0 – 18 years) belonged almost entirely to this  
283 topic, as did young females (ages 19 – 34 years). Discriminatory terms included words  
284 associated with education, including ‘child’, ‘college’, ‘primary’, ‘school’, and ‘student’, as  
285 well as the name ‘Bollon’ (a town in the inland shire of Balonne, which is a region that  
286 consistently has high rates of Q fever notifications in Queensland).

287 Topic 3: this topic was strongly associated with working age and retired females (35  
288 years and older) as well as some males of retirement age (ages 65 – 100 years).  
289 Discriminatory terms included ‘wife’, ‘housewife’, ‘husband’, ‘vet’, ‘nurse’ and ‘cook’.

290 Topic 4: this topic was entirely composed of working age and retired males (ages 35  
291 years and older) and was distinguished by terms commonly associated with factory workers  
292 and tradesman, including ‘drain’, ‘factory’, ‘weld’, ‘milk’ and ‘handyman’.

293 Topic 5: this group included working age males and females (ages 35 – 64) and some  
294 younger working females (ages 19 – 34). Discriminatory terms reflected possible indirect  
295 exposure routes, including ‘manure’, ‘observe’, ‘post’, and ‘office’, as well as some terms  
296 that may reflect direct exposure including ‘bull’ and ‘meatworker’.

297 To assess spatial patterns in the distributions of topics, we predicted the most  
298 probable topic for each of the 2,044 patients using their individual responses. After adjusting  
299 for resident population sizes of the surrounding Local Government Area (LGA), we found

300 that all topics were generally more common in central areas of the state where Q fever  
301 notifications have traditionally been high (**Figure 2**; S. J. Tozer, 2015). However, some key  
302 spatial differences across topics were evident. Topics associated with working age males  
303 (ages 19 – 34 years) and children (males and females ages 0 – 18 years) primarily occurred in  
304 the central and central-south areas of the state (Topics 1 and 2; **Figure 2**). In contrast, topics  
305 associated primarily with patients aged 35+ years (Topics 3 and 4) were both generally more  
306 common in central-north areas of the state (**Figure 2**). Topic 5, which contained a mix of  
307 words suggesting non-occupational exposure, was more evenly distributed across northern  
308 and southern areas in central Queensland (**Figure 2**).

309

310 Animal exposures reported by Q fever patients

311 From 2001 – 2017, a total of 1,380 individual Q fever cases reported potential exposure to at  
312 least one of the five target animal groups. Of these, 890 (64%) reported exposure to cattle,  
313 638 (36%) to macropods, 384 (28%) to sheep, 347 (25%) to pigs and 237 (17%) to goats  
314 (note some individuals reported exposure to multiple animal groups; **Figure 3**). Notifications  
315 with reported exposures occurred across much of the state, though the majority were again  
316 concentrated in the central and central-south areas for each of the five target animal groups  
317 (**Figure 3**). This same pattern also held for those individuals that did not report exposure to  
318 any of the five target animal groups (**Figure S3**). Across years, cattle exposure was the most  
319 common animal exposure pathway, though reports of non-cattle exposure (particularly for  
320 macropods) were noticeably more common following the rollout of the expanded  
321 surveillance in 2012 (**Figure 4a**). Across all 1,380 individuals, a total of 187 cases (14%)  
322 were identified that reported exposure to macropods but did not report exposure to any  
323 livestock species. In contrast, 396 cases reported only cattle exposure (27%), 99 (7%)  
324 reported only sheep exposure and 30 (2%) reported only pig exposure.

325 Proportions of reported exposures attributed to each animal showed some noticeable  
326 variation across topic groups (**Figure 4b**). Most notably, patients belonging to Topic 3  
327 (working age and retired females and some retired males, primarily located in the central-  
328 north areas of the state) and Topic 4 (working age and retired males associated with factories  
329 or trades, also commonly found in the central-north) more often reported macropod exposure  
330 than any other group, with 35.6% and 29.3% of exposures attributed to macropods,  
331 respectively (**Figure 4b**). Patients from these groups also tended to report fewer exposures to  
332 sheep and goats than patients from other groups. In contrast, groups more commonly found in  
333 south-central areas of the state, including patients from Topics 1 and 2 (working age males in

334 the livestock trade and children / young females associated with the education industry),  
335 reported more even exposures across the five animal reservoir species (**Figure 4b**).

336

337 Animal co-exposures identified using Markov Random Fields

338 Our MRF model, built using the dataset of 1,380 patients from the years 2001 – 2017,  
339 identified a number of important conditional pairwise relationships between reported animal  
340 exposures (**Figure 5, top graph**). Patients that reported goat exposure were > 3 times more  
341 likely to also report sheep exposure after accounting for all other exposures (marginal Odds  
342 Ratio 95% credible interval [OR]: 2.88 – 3.71). A similarly strong positive relationship was  
343 found for macropod and pig exposures (OR: 2.81 – 3.59). In contrast, a strong negative  
344 relationship was identified between cattle and macropod exposures (OR: 0.29 – 0.36). In  
345 addition, we estimated that patients were approximately 50% more likely to be hospitalised if  
346 they reported macropod exposure than if they did not (OR: 1.32 – 1.65), while patients were  
347 20% less likely to be hospitalised if they reported pig exposure (OR: 0.75 – 0.95).

348

349 Associations with probability of hospitalisation

350 A total of 672 of the 1,380 patients included in the analysis dataset were admitted to hospital  
351 as a result of Q fever infection. Of the three logistic regressions we tested, the LASSO  
352 algorithm without spatial regression splines was the best-fitting and most parsimonious  
353 model (prediction accuracy range: 0.53 – 0.66, compared to ranges of 0.51 – 0.64 for the  
354 gaussian process LASSO and 0.51 – 0.63 for the spatial GAM). This model retained five  
355 important predictors: exposure to macropods, exposure to animal births, sex (male = 1), age  
356 and year of notification. Effect sizes revealed that all of these variables increased risk of  
357 hospitalisation apart from exposure to animal births (**Figure S4**). Being male increased risk  
358 by 48% (effect size 95% CI [ES]: 1.16 – 1.90), exposure to macropods increased risk by 34%  
359 (ES: 1.07 – 1.66) and each additional 5 years of age increased risk by 16% (ES: 0.15 – 0.17).  
360 Hospitalisation risk also increased by 10% each year from 2001 – 2017 (ES: 0.10 – 0.12)  
361 (**Figure S3**). Exposure to animal births decreased risk by 67% (ES: 0.52 – 0.85). There was  
362 no difference in numbers of patients that reported animal birth exposure between the sexes  
363 ( $\chi^2$  test:  $\chi^2 = 2.12$ ,  $p = 0.15$ ), though there was a moderate difference in ages. Specifically,  
364 patients that reported animal birth exposure were 0.5 – 4.5 years younger than those that did  
365 not (t test:  $t = -2.24$ ,  $p = 0.03$ ).

366

367 **Discussion**

368 Our study provides new insights into the complexities of Q fever epidemiology and  
369 showcases the utility of incorporating enhanced surveillance data into disease monitoring and  
370 research programmes. The undifferentiated nature of clinical presentations associated with Q  
371 fever and the lack of awareness of this disease as a potential diagnosis across geographical  
372 regions means that adequate treatment will often be delayed or missed (Dahlgren, Haberling,  
373 & McQuiston, 2015; Lindsay et al., 2018; Million & Raoult, 2017; Didier Raoult & Marrie,  
374 1995). A better understanding of exposure pathways is necessary to help design measures  
375 aimed at preventing exposure to *C. burnetii* (Angelakis & Raoult, 2011; Clutterbuck et al.,  
376 2018). The results of our text modelling approach demonstrate two clear patterns of reported  
377 exposures among Queensland's Q fever notifications: (1) responses to survey questions  
378 differed among demographic groups and (2) patients belonging to different exposure topics  
379 were often concentrated in different geographical areas. Moreover, we identify predictors of  
380 hospitalisation risk and show that the simplified exposure questionnaire performed similarly  
381 to the expanded questionnaire; these findings can help improve resource allocation to reduce  
382 the burden of Q fever infection. Collectively, our study indicates that follow-up surveillance  
383 combined with text modelling is useful for unravelling exposure pathways in the battle to  
384 reduce the incidence Q fever and other zoonotic diseases.

385 With one of the world's highest Q fever notification rates and a long history of  
386 livestock-based agriculture, Queensland is a focus area for research on Q fever epidemiology  
387 (Sivabalan et al., 2017; Sloan-Gardner et al., 2017; S. J. Tozer, 2015). Key among efforts to  
388 reduce Q fever incidence is Queensland Health's use of follow-up surveillance of notified  
389 cases. These crucial data, particularly following the 2012 rollout of an extended outbreak  
390 investigation form, are providing deeper insights into possible exposure pathways  
391 (Communicable Diseases Network Australia, 2018). However, making sense of text data that  
392 results from open-ended questions can be challenging and often requires model-based  
393 algorithms (Paul & Dredze, 2014; Roberts et al., 2014). By applying a topic model to an 18-  
394 year dataset of Queensland Q fever notifications, we show that patients from different  
395 demographic groups consistently reported distinct sets of exposure terms, suggesting  
396 demographic-specific transmission pathways. Moreover, our study expands on the well-  
397 known concentration of Q fever notification rates in rural Australia (Gidding et al., 2009) to  
398 demonstrate that patients associated with different exposure pathways showed different  
399 spatial patterns, with some concentrating more in the north and others in the south of the  
400 state. These findings provide an evidence-base for multifaceted and epidemiologically  
401 relevant health promotion campaigns that can act in tandem with ongoing Q fever

402 occupational vaccination programmes to increase Q fever awareness and decrease burdens of  
403 disease.

404 Our models provide strong evidence that open-response answers from younger  
405 working males (ages 19 – 34 years; Topic 1) were compositionally different to those from  
406 older working males (35 – 64 years; Topic 3). In general, younger males reported terms  
407 associated with the livestock industry while older males reported indirect exposure terms or  
408 terms associated with trades. With known discrepancies between occupations considered  
409 ‘high-risk’ by Australian health bodies and those thought of as ‘high-risk’ by rural  
410 practitioners (Lindsay et al., 2018), this finding that males likely encounter different  
411 occupational exposures between age groups provides useful information for designing  
412 education and vaccination programmes. In contrast, children and young females commonly  
413 reported terms associated with education (Topic 2), perhaps indicating they were less likely  
414 to directly participate in traditional high-risk activities. However, this maternal-child word  
415 topic strongly overlapped in space with areas that harboured relatively high densities of  
416 working males from the traditional occupational group (Topic 1). Moreover, patients from  
417 these two groups (young working age males and children / young working age females) were  
418 also very similar in terms of their animal exposure profiles, with both groups reporting  
419 moderate cattle exposure but more commonly reporting pig exposure compared to other  
420 groups. These results have public health implications due to the fact that (1) the current  
421 advice for Q fever vaccination is that it should not be administered to patients younger than  
422 15 years (Australian Technical Advisory Group on Immunisation (ATAGI), 2018) and (2)  
423 awareness programmes are not currently targeting family members, particularly children, of  
424 stockman (Armstrong et al., 2019; Gidding et al., 2009).

425 A prominent finding of our study is that older patients, particularly those residing in  
426 Queensland’s northern regional areas, represent a key and epidemiologically distinct at-risk  
427 group for Q fever infection. Patients aged 65 years and older were (1) more concentrated in  
428 the central-north of the state, (2) more likely to report macropod exposure but less likely to  
429 report goat or sheep exposure and (3) more likely to be hospitalised due to infection.  
430 Interestingly, patients in this group also commonly reported occupational exposure terms  
431 associated with the veterinary industry, including ‘vet’ and ‘nurse’. The recognition that older  
432 patients are exposed to *C. burnetti* through different pathways, and that risks of  
433 hospitalisation are higher, confirms previous findings from Australia (Karki et al., 2015) and  
434 elsewhere (Dupont et al., 1992). This has implications for the future distribution of public  
435 health resources. Population ageing resulting from accelerated expansion of older people is a

436 major phenomenon affecting many of the world's developed countries, and Australia is no  
437 exception (Ofori-Asenso, Zomer, Curtis, Zoungas, & Gambhir, 2018). From the year 1996 to  
438 2016, the proportion of Australia's population aged 65 years and over increased from 12.0 to  
439 15.3%, and such increases are expected to continue (Australian Bureau of Statistics (ABS),  
440 2016). By demonstrating a strong correlation between patient age and the probability of  
441 hospitalisation due to Q fever infection, our study contributes to growing evidence that aging  
442 populations are associated with increased demands for healthcare (Beard & Bloom, 2015).  
443 Understanding how much of this increasing demand is driven by changes in reporting,  
444 heightened awareness or the rollout of intervention programs (such as Q fever vaccination in  
445 Australia) should be a topic of future research. This is particularly true given our finding that  
446 probability of hospitalization increased with year of onset across our study's timeframe. It is  
447 unlikely that Q fever severity has increased over time. Rather, this pattern could reflect  
448 heightened awareness of the disease and its health impacts, or perhaps a shift from primarily  
449 acute cases in livestock workers to non-occupational cases that are more difficult to diagnose  
450 due to a lack of obvious exposure pathways. Indeed, the authors of a recent time-series  
451 analysis of Q fever notifications in Victoria, Australia found evidence for such a pattern and  
452 postulated that many mild cases likely remain undiagnosed, leading to a relatively high  
453 hospitalisation rate for those more severe cases that are confirmed as Q fever (Bond et al.,  
454 2018).

455 Many species of wildlife have long shown evidence of exposure to *C. burnetti*, and  
456 some authors have made the suggestions that these species can pose greater zoonotic risks  
457 than livestock in particular environments (Enright et al., 1971; González-Barrio & Ruiz-Fons,  
458 2019; Koehler, Kloppert, Hamann, El-Sayed, & Zschöck, 2019). Multiple lines of evidence  
459 from our study confirm previous findings that macropods may be a primary reservoir host for  
460 *C. burnetii* (Banazis et al., 2010; Alanna Cooper et al., 2012; A Cooper et al., 2013). First,  
461 following the implementation in 2012 of the enhanced surveillance exposure questionnaire  
462 macropod exposure has become the second most common reported animal exposure among  
463 patients with confirmed Q fever infection. This pattern has been quite stable since 2012.  
464 Second, 14% of patients reported exposure to macropods without reporting exposure to any  
465 of the more frequently implicated livestock species such as cattle, sheep and goats. Finally,  
466 our study provides limited evidence that exposure to macropods may be an indicator of a Q  
467 fever patient's severity of disease. Reported macropod exposure correlated with an up to 66%  
468 increased risk of hospitalisation after accounting for other factors such as patient age, sex and  
469 the year of onset. We note however that increased rates of reported exposure to a particular

470 animal species does not imply it represents a prominent source of *C. burnetii*. Many people in  
471 Australia observe and encounter macropods on a regular basis without contracting Q fever,  
472 and we are unaware of any empirical evidence that pigs are a source of *C. burnetii* in  
473 Australia. Our findings should be used to motivate further empirical studies to identify  
474 transmission pathways among cohorts of individuals reporting different exposure profiles.  
475 Useful future studies can also address whether there is any difference between domestic vs  
476 feral animal exposure rates and can investigate other possible animal exposures for their  
477 associations with patient cohorts

478 While risk of hospitalisation due to Q fever may not necessarily be a robust proxy for  
479 severity, associations with hospitalisation risk can still uncover important patterns in the  
480 burden of disease. In addition to the risk factor of macropod exposure and consistent with  
481 previous studies in Australia, we found that working-age and older males were at higher risk  
482 of hospitalisation (Garner, Longbottom, Cannon, & Plant, 1997; Sloan-Gardner et al., 2017).  
483 An interesting association was the negative influence of exposure to animal births on  
484 hospitalisation risk. Traditionally, assisting in livestock births is considered one of the riskiest  
485 occupational activities for acquiring Q fever, particularly if this occurs during a *Coxiella*-  
486 induced abortion wave (Berri, Rousset, Champion, Russo, & Rodolakis, 2007; Boden,  
487 Brasche, Straube, & Bischof, 2014). Without an in-depth understanding of who attends  
488 animal births on each property, it is difficult to ascertain whether this finding is being  
489 confounded by other factors that were not captured by our exposure dataset.

490 Several limitations of our study should be considered when interpreting our results.  
491 First, frequent patterns of reported co-exposures make it challenging to pinpoint the exact  
492 source of infection. For example, sheep and goat exposures were very commonly co-reported,  
493 as were macropod and pig exposures. This is not surprising. Mixed-species farms are  
494 common in Queensland and both macropods and feral pigs are widespread across the state  
495 (Bastin, Smith, Watson, & Fisher, 2009; Gentle, Speed, & Marshall, 2015; Woodall, 1983).  
496 Reports of 'exposure' may in many cases simply relate to observations of a nearby animal,  
497 rather than any meaningful interaction that could represent a transmission pathway.  
498 Household investigations to improve estimates of source attribution are needed to tease these  
499 patterns apart. Second, a lack of data to distinguish between pre-school and school-aged  
500 children meant that we could not assess whether these groups may have different exposure  
501 profiles. Variation in the stringency of follow-up investigation across different treatment  
502 centres may lead to inconsistencies in the detail of exposure reports. And finally, our reliance

503 on notification data means that only a proportion of the total cases occurring in Queensland  
504 during the study period were included.

505 In conclusion, our study has demonstrated that Q fever epidemiology in Queensland is  
506 non-stationary in that exposure factors for Q fever notifications and risk of hospitalisation  
507 play different roles depending on location. Our findings suggest local investigations are  
508 necessary to uncover factors associated with exposure to infection in the high-risk areas and  
509 populations identified in this study.

510

### 511 **Acknowledgements**

512 This study was supported with funding from the Australian Government Department of  
513 Agriculture and Water Resources as part of its Rural R&D for Profit programme  
514 administered by AgriFutures Australia: ‘Taking the Q (query) out of Q Fever’ (Project  
515 RnD4Profit-15-02-008). Simon Firestone is supported by an Australian Research Council  
516 Discovery Early Career Researcher Award (project number DE160100477).

517

### 518 **Conflict of interest statement**

519 The authors declare that we have no conflict of interest

520

### 521 **Data availability statement**

522 Restrictions apply to the availability of these data, which were used under license for this  
523 study. Data are available upon justified request from Queensland Health’s Epidemiology &  
524 Research Unit.

525

### 526 **References**

- 527 Allan-Blitz, L.-T., Sakona, A., Wallace, W. D., & Klausner, J. D. (2018). *Coxiella burnetii*  
528 endocarditis and meningitis, California, USA, 2017. *Emerging Infectious Diseases*,  
529 24(8), 1555-1557. doi:10.3201/eid2408.180249
- 530 Angelakis, E., & Raoult, D. (2011). Emergence of Q fever. *Iranian Journal of Public Health*,  
531 40(3), 1.
- 532 Armstrong, M., Francis, J., Robson, J., Graves, S., Mills, D., Ferguson, J., & Nourse, C.  
533 (2019). Q fever vaccination of children in Australia: limited experience to date.  
534 *Journal of Paediatrics and Child Health*. doi:doi:10.1111/jpc.14364
- 535 Arricau-Bouvery, N., & Rodolakis, A. (2005). Is Q fever an emerging or re-emerging  
536 zoonosis? *Veterinary research*, 36(3), 327-349.

- 537 Australian Bureau of Statistics (ABS). (2016). *Australian historical population statistics,*  
538 *2016. ABS cat. no. 3105.0.65.001.* Retrieved from Canberra:
- 539 Australian Technical Advisory Group on Immunisation (ATAGI). (2018). *Australian*  
540 *Immunisation Handbook.* Retrieved from Canberra,  
541 [immunisationhandbook.health.gov.au](http://immunisationhandbook.health.gov.au): [immunisationhandbook.health.gov.au](http://immunisationhandbook.health.gov.au)
- 542 Banazis, M. J., Bestall, A. S., Reid, S. A., & Fenwick, S. G. (2010). A survey of Western  
543 Australian sheep, cattle and kangaroos to determine the prevalence of *Coxiella*  
544 *burnetii*. *Veterinary Microbiology, 143*(2-4), 337-345.
- 545 Bastin, G., Smith, D. S., Watson, I., & Fisher, A. (2009). The Australian Collaborative  
546 Rangelands Information System: preparing for a climate of change. *The Rangeland*  
547 *Journal, 31*(1), 111-125.
- 548 Beard, J. R., & Bloom, D. E. (2015). Towards a comprehensive public health response to  
549 population ageing. *Lancet (London, England), 385*(9968), 658-661.  
550 doi:10.1016/S0140-6736(14)61461-6
- 551 Berri, M., Rousset, E., Champion, J., Russo, P., & Rodolakis, A. (2007). Goats may  
552 experience reproductive failures and shed *Coxiella burnetii* at two successive  
553 parturitions after a Q fever infection. *Research in Veterinary Science, 83*(1), 47-52.
- 554 Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.
- 555 Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of*  
556 *Machine Learning Research, 3*(Jan), 993-1022.
- 557 Boden, K., Brasche, S., Straube, E., & Bischof, W. (2014). Specific risk factors for  
558 contracting Q fever: lessons from the outbreak Jena. *International Journal of Hygiene*  
559 *and Environmental Health, 217*(1), 110-115.
- 560 Bond, K., Franklin, L., Sutton, B., Stevenson, M., & Firestone, S. (2018). Review of 20 years  
561 of human acute Q fever notifications in Victoria, 1994–2013. *Australian Veterinary*  
562 *Journal, 96*(6), 223-230.
- 563 Clark, N., & Soares Magalhães, R. J. (2018). Airborne geographical dispersal of Q Fever  
564 from livestock holdings to human communities: a systematic review and critical  
565 appraisal of evidence. *BMC Infectious Diseases*, doi: 10.1186/s12879-12018-13135-  
566 12874. doi:10.1186/s12879-018-3135-4
- 567 Clark, N. J., Wells, K., & Lindberg, O. (2018a). MRFCov: Markov Random Fields with  
568 additional covariates. R package version 1.0.  
569 <https://github.com/nicholasjclark/MRFCov>: GitHub.

570 Clark, N. J., Wells, K., & Lindberg, O. (2018b). Unravelling changing interspecific  
571 interactions across environmental gradients using Markov random fields. *Ecology*,  
572 doi: 10.1002/ecy.2221.

573 Clutterbuck, H., Eastwood, K., Massey, P. D., Hope, K., & Mor, S. M. (2018). Surveillance  
574 system enhancements for Q fever in NSW, 2005-2015. *Communicable Diseases*  
575 *Intelligence*, 42, PII:S2209-6051(2218)00012-00010.

576 Communicable Diseases Network Australia. (2018). *Q fever*. CDNA Series of National  
577 Guidelines (SoNG) for Public Health Units: accessed at  
578 ([http://www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF7](http://www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF71CA2583520001F02F/$File/Q-fever-SoNG2018.pdf)  
579 [1CA2583520001F02F/\\$File/Q-fever-SoNG2018.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF71CA2583520001F02F/$File/Q-fever-SoNG2018.pdf)) Retrieved from  
580 [www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF7](http://www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF71CA2583520001F02F/$File/Q-fever-SoNG2018.pdf)  
581 [1CA2583520001F02F/\\$File/Q-fever-SoNG2018.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/56DFBAB23468BF71CA2583520001F02F/$File/Q-fever-SoNG2018.pdf).

581 Cooper, A., Barnes, T., Potter, A., Ketheesan, N., & Govan, B. (2012). Determination of  
582 *Coxiella burnetii* seroprevalence in macropods in Australia. *Veterinary Microbiology*,  
583 155(2-4), 317-323.

584 Cooper, A., Stephens, J., Ketheesan, N., & Govan, B. (2013). Detection of *Coxiella burnetii*  
585 DNA in wildlife and ticks in northern Queensland, Australia. *Vector-Borne and*  
586 *Zoonotic Diseases*, 13(1), 12-16.

587 Dahlgren, F. S., Haberling, D. L., & McQuiston, J. H. (2015). Q fever is underestimated in  
588 the United States: a comparison of fatal Q fever cases from two national reporting  
589 systems. *The American Journal of Tropical Medicine and Hygiene*, 92(2), 244-246.

590 Dupont, H., Raoult, D., Brouqui, P., Janbon, F., Peyramond, D., Weiller, P.-J., . . . Poirier, R.  
591 (1992). Epidemiologic features and clinical presentation of acute Q fever in  
592 hospitalized patients: 323 French cases. *The American journal of medicine*, 93(4),  
593 427-434.

594 Enright, J. B., Franti, C., Behymer, D., Longhurst, W., Dutson, V., & WRIGHT, M. E.  
595 (1971). *Coxiella burnetii* in a wildlife-livestock environment: distribution of Q fever in  
596 wild mammals. *American Journal of Epidemiology*, 94(1), 79-90.

597 Fenollar, F., Fournier, P.-E., Carrieri, M. P., Habib, G., Messana, T., & Raoult, D. (2001).  
598 Risks factors and prevention of Q fever endocarditis. *Clinical Infectious Diseases*,  
599 33(3), 312-316.

600 Fitzpatrick, K. A., Kersh, G. J., & Massung, R. F. (2010). Practical method for extraction of  
601 PCR-quality DNA from environmental soil samples. *Applied and Environmental*  
602 *Microbiology*, 76(13), 4571-4573.

603 Flint, J., Dalton, C. B., Merritt, T. D., Graves, S., Ferguson, J. K., Osbourn, M., . . .  
604 Durrheim, D. N. (2016). Q fever and contact with kangaroos in New South Wales.  
605 *Communicable diseases intelligence quarterly report*, 40(2), E202.

606 Fountain-Jones, N. M., Clark, N., Kinsley, A., Carstensen, M., Forrester, J., Johnson, T. J., . . .  
607 . Craft, M. (2019). Microbial associations and spatial proximity predict North  
608 American moose (*Alces alces*) gastrointestinal community composition. *bioRxiv*,  
609 514604.

610 Fox, C. (1989). A stop list for general text. *ACM SIGIR Forum*, 24(1-2), 19-21.

611 Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear  
612 models via coordinate descent. *Journal of Statistical Software*, 33, 1-22.

613 Garner, M. G., Longbottom, H. M., Cannon, R. M., & Plant, A. J. (1997). A review of Q  
614 fever in Australia 1991–1994. *Australian and New Zealand journal of public health*,  
615 21(7), 722-730.

616 Gentle, M., Speed, J., & Marshall, D. (2015). Consumption of crops by feral pigs (*Sus scrofa*)  
617 in a fragmented agricultural landscape. *Australian Mammalogy*, 37(2), 194-200.

618 Gidding, H. F., Wallace, C., Lawrence, G. L., & McIntyre, P. B. (2009). Australia's national  
619 Q fever vaccination program. *Vaccine*, 27(14), 2037-2041.

620 González-Barrio, D., & Ruiz-Fons, F. (2019). *Coxiella burnetii* in wild mammals: a  
621 systematic review. *Transboundary and emerging diseases*, 66(2), 662-671.

622 Graves, S. R., & Islam, A. (2016). Endemic Q fever in New South Wales, Australia: a case  
623 series (2005–2013). *The American Journal of Tropical Medicine and Hygiene*, 95(1),  
624 55-59.

625 Guatteo, R., Beaudeau, F., Berri, M., Rodolakis, A., Joly, A., & Seegers, H. (2006). Shedding  
626 routes of *Coxiella burnetii* in dairy cows: implications for detection and control.  
627 *Veterinary research*, 37(6), 827-833.

628 Gyuranecz, M., Sulyok, K. M., Balla, E., Mag, T., Balazs, A., Simor, Z., . . . Hornstra, H.  
629 (2014). Q fever epidemic in Hungary, April to July 2013. *Eurosurveillance*, 19(30), 5.

630 Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov  
631 networks. *Ecology*, 97(12), 3308-3314.

632 Hoffman, L. C., & Cawthorn, D.-M. (2012). What is the role and contribution of meat from  
633 wildlife in providing high quality protein for consumption? *Animal frontiers*, 2(4), 40-  
634 53.

- 635 Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. *Journal*  
636 *of Statistical Software*, 40(13), 1-30.
- 637 Kammann, E., & Wand, M. P. (2003). Geoaddivitive models. *Journal of the Royal Statistical*  
638 *Society: Series C (Applied Statistics)*, 52(1), 1-18.
- 639 Karagiannis, I., Schimmer, B., Van Lier, A., Timen, A., Schneeberger, P., Van Rotterdam,  
640 B., . . . Van Duynhoven, Y. (2009). Investigation of a Q fever outbreak in a rural area  
641 of The Netherlands. *Epidemiology and Infection*, 137(9), 1283-1294.  
642 doi:10.1017/S0950268808001908
- 643 Karki, S., Gidding, H. F., Newall, A. T., McIntyre, P. B., & Liu, B. C. (2015). Risk factors  
644 and burden of acute Q fever in older adults in New South Wales: a prospective cohort  
645 study. *Medical Journal of Australia*, 203(11), 438-438.
- 646 Koehler, L. M., Kloppert, B., Hamann, H.-P., El-Sayed, A., & Zschöck, M. (2019).  
647 Comprehensive literature review of the sources of infection and transmission routes of  
648 *Coxiella burnetii*, with particular regard to the criteria of “evidence-based medicine”.  
649 *Comparative immunology, microbiology and infectious diseases*, 64, 67-72.  
650 doi:<https://doi.org/10.1016/j.cimid.2019.02.004>
- 651 Lindsay, P. J., Rohailla, S., & Miyakis, S. (2018). Q fever in rural Australia: education versus  
652 vaccination. *Vector-Borne and Zoonotic Diseases*, 0(0), null.  
653 doi:10.1089/vbz.2018.2307
- 654 Million, M., & Raoult, D. (2017). No such thing as chronic Q fever. *Emerging Infectious*  
655 *Diseases*, 23(5), 856.
- 656 Mori, M., & Roest, H.-J. (2018). Farming, Q fever and public health: agricultural practices  
657 and beyond. *Archives of Public Health*, 76(1), 2.
- 658 Ofori-Asenso, R., Zomer, E., Curtis, A. J., Zoungas, S., & Gambhir, M. (2018). Measures of  
659 population ageing in Australia from 1950 to 2050. *Journal of Population Ageing*,  
660 11(4), 367-385. doi:10.1007/s12062-017-9203-5
- 661 Ooms, J. (2017). hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker.  
662 <https://CRAN.R-project.org/package=hunspell>: R package version 2.9. Retrieved  
663 from <https://CRAN.R-project.org/package=hunspell>
- 664 Paul, M. J., & Dredze, M. (2014). Discovering health topics in social media using topic  
665 models. *Plos One*, 9(8), e103408.
- 666 R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna,  
667 Austria. Retrieved from <https://www.R-project.org>
- 668 Raoult, D., & Marrie, T. (1995). Q fever. *Clinical Infectious Diseases*, 489-495.

- 669 Raoult, D., Marrie, T., & Mege, J. (2005). Natural history and pathophysiology of Q fever.  
670 *The Lancet Infectious Diseases*, 5(4), 219-226.
- 671 Reedijk, M., Van Leuken, J., & Van Der Hoek, W. (2013). Particulate matter strongly  
672 associated with human Q fever in The Netherlands: an ecological study.  
673 *Epidemiology & Infection*, 141(12), 2623-2633.
- 674 Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . .  
675 Rand, D. G. (2014). Structural topic models for open-ended survey responses.  
676 *American Journal of Political Science*, 58(4), 1064-1082.
- 677 Roest, H., Tilburg, J., Van der Hoek, W., Vellema, P., Van Zijderveld, F., Klaassen, C., &  
678 Raoult, D. (2011). The Q fever epidemic in The Netherlands: history, onset, response  
679 and reflection. *Epidemiology & Infection*, 139(1), 1-12.
- 680 Sellens, E., Bosward, K., Willis, S., Heller, J., Cobbold, R., Comeau, J., . . . Wood, N. (2018).  
681 Frequency of adverse events following Q fever immunisation in young adults.  
682 *Vaccines*, 6(4), 83.
- 683 Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data  
684 Principles in R. *The Open Journal*, 1(3), 10.21105/joss.00037.
- 685 Sivabalan, P., Saboo, A., Yew, J., & Norton, R. (2017). Q fever in an endemic region of  
686 North Queensland, Australia: A 10year review. *One Health*, 3, 51-55.  
687 doi:<https://doi.org/10.1016/j.onehlt.2017.03.002>
- 688 Sloan-Gardner, T., Massey, P., Hutchinson, P., Knope, K., & Fearnley, E. (2017). Trends and  
689 risk factors for human Q fever in Australia, 1991–2014. *Epidemiology & Infection*,  
690 145(4), 787-795.
- 691 Tozer, S., Lambert, S., Sloots, T., & Nissen, M. (2011). Q fever seroprevalence in  
692 metropolitan samples is similar to rural/remote samples in Queensland, Australia.  
693 *European journal of clinical microbiology & infectious diseases*, 30(10), 1287.
- 694 Tozer, S. J. (2015). *Epidemiology, Diagnosis and Prevention of Q fever in Queensland*. The  
695 University of Queensland, Brisbane, Australia.
- 696 Van der Hoek, W., Dijkstra, F., Schimmer, B., Schneeberger, P., Vellema, P., Wijkmans, C., .  
697 . . . Van Duynhoven, Y. (2010). Q fever in the Netherlands: an update on the  
698 epidemiology and control measures. *Eurosurveillance*, 15(12), 19520.
- 699 Webster, J., Lloyd, G., & Macdonald, D. (1995). Q fever (*Coxiella burnetii*) reservoir in wild  
700 brown rat (*Rattus norvegicus*) populations in the UK. *Parasitology*, 110(1), 31-35.

701 Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse': R package version  
 702 1.2.1. Retrieved from <https://CRAN.R-project.org/package=tidyverse>  
 703 Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society:*  
 704 *Series B (Statistical Methodology)*, 65(1), 95-114.  
 705 Woodall, P. F. (1983). Distribution and population dynamics of dingoes (*Canis familiaris*)  
 706 and feral pigs (*Sus scrofa*) in Queensland, 1945-1976. *Journal of Applied Ecology*,  
 707 85-95.

708  
 709  
 710 **Table 1:** Summary statistics for the eight demographic groups used in topic modelling  
 711 analysis.

Demographic group	# individuals	# total words (mean per individual)	# unique words (mean per individual)
<i>Females</i>			
0 - 18 years	56	361 (6.45)	94 (1.68)
19 – 34 years	69	489 (7.09)	144 (2.09)
35 - 64 years	346	2,222 (6.42)	285 (0.82)
65 - 100 years	63	437 (6.94)	115 (1.83)
<i>Males</i>			
0 - 18 years	125	1,223 (9.78)	195 (1.56)
19 – 34 years	283	2,268 (8.01)	305 (1.08)
35 - 64 years	905	6,163 (6.81)	381 (0.42)
65 - 100 years	197	1,319 (6.70)	216 (1.10)

712 Individual patients were grouped by age and sex categories and their responses to open-ended exposure questions were pooled to form a  
 713 document term matrix. Stop words, numerics and words recorded fewer than five times overall were removed prior to analysis.

714

715

716 **FIGURE LEGENDS**

717 **Figure 1:** Results from topic model analysis of Q fever patient responses to open-ended  
718 exposure questions. (Left) relative contributions of each of the five latent response word  
719 groups (i.e. topics) to each demographic group's total word composition; (Right) wordclouds  
720 depicting words that had the highest discriminatory power for each of the five latent response  
721 word topics. Colours of wordclouds correspond to colours of word topics. Sizes of words are  
722 proportional to their discriminatory power (larger size indicates a word is more strongly  
723 associated with that particular word topic). Bold text indicates key discriminatory words  
724 indicative of possible exposure pathways. Note that italic text refers to names of towns in  
725 rural Queensland: Chinchilla, Boonah, Winton, Isisford, Mitchell, Dirranbandi, Bollon,  
726 Charleville, Minnel, Morven, Emerald and Sarina.

727  
728 **Figure 2:** Distributions of human Q fever notifications assigned to each word topic across  
729 Local Government Areas (LGAs) in Queensland, Australia from the years 2001 – 2017.  
730 Topics were identified by applying a topic model analysis of Q fever patient responses to  
731 open-ended exposure questions. Numbers of notifications are adjusted for the resident human  
732 population size in each LGA to present a notification rate. This figure was generated in R  
733 version 3.3.3 using a shapefile of Queensland LGAs, available from [data.qld.gov.au](http://data.qld.gov.au).

734  
735 **Figure 3:** Locations of human Q fever notifications across Local Government Areas (LGAs)  
736 in Queensland, Australia with reported exposure to target animal groups from the years 2001  
737 – 2017. Numbers of notifications are adjusted for the resident human population size in each  
738 LGA to present a notification rate. Note, the final dataset included 1,380 patients but some of  
739 these reported exposure to two or more animal groups. This figure was generated in R  
740 version 3.3.3 using a shapefile of Queensland LGAs, available from [data.qld.gov.au](http://data.qld.gov.au).

741  
742 **Figure 4:** Proportions of Q fever notifications in Queensland, Australia reporting potential  
743 exposure to target animal groups, by year (a) and word topic identified by Latent Dirichlet  
744 Allocation modelling (b). Colours of stacked bar charts represent the proportions of  
745 notifications with reported exposure for each of the five animal groups. Note that an  
746 expanded surveillance form was rolled out in Queensland from 2012 to specifically prompt  
747 patients to report animal exposures.

748  
749 **Figure 5:** Conditional associations among exposure (recorded as 'reported exposure' or 'no  
750 reported exposure') and hospitalisation (recorded as 'hospitalised' or 'not hospitalised')

751 variables estimated from Markov Random Fields network models for the full dataset (from  
752 the years 2001 – 2017; n = 1,380; top graph) and a reduced dataset that followed the rollout  
753 of an extended surveillance form (years 2012 – 2017; n = 979; bottom graph). Numbers on  
754 the diagonals indicate the total number of Q fever notifications in Queensland, Australia in  
755 which a single exposure was recorded (i.e. the variable in the specified row was recorded as a  
756 ‘1’ while all other variables were recorded as ‘0’). Numbers in the off-diagonals represent  
757 numbers of co-exposures. Darker reds indicate that a variable pair’s exposure probabilities  
758 are positively associated after accounting for all other variables in the graph, while darker  
759 blues indicate negative associations among pairs of variables.

760

761

762 **APPENDIX 1**

Case name: ..... *First name* ..... *Surname* ..... DOB ...../...../..... Notification ID: .....

## Q Fever Case Report Form



Public Health Unit ..... Outbreak ID: .....  
Completed by: ..... Date sent to NOCS: ...../...../.....  
Telephone: ..... Fax: .....

### NOTIFICATION:

Date PHU notified: ...../...../..... Date initial response: ...../...../.....  
Notifier: ..... Organisation: .....  
Telephone: ..... Fax: ..... Email: .....  
Treating Dr: .....  
Telephone: ..... Fax: ..... Email: .....

### CASE DETAILS:

### UR No: .....

Name: ..... *First name* ..... *Surname* .....  
Date of birth: ...../...../..... Age: ..... Years ..... Months Sex:  Male  Female .....  
Name of parent/carer: .....  
 Aboriginal  Torres Strait Islander  Aboriginal & Torres Strait Islander  Non-Indigenous  Unknown  
English preferred language:  Yes  No – *specify* ..... Ethnicity – *specify* .....  
Permanent address: ..... Postcode: .....  
Home tel: ..... Mob: ..... Email: .....  
Occupation: ..... Work telephone: .....  
Temporary address in Queensland (*if different from permanent address*): ..... Postcode: .....  
Telephone: ..... Mob: ..... Email: .....  
General Practitioner: Dr .....  
Address: ..... Postcode: .....  
Telephone: ..... Fax: ..... Email: .....

### CLINICAL DETAILS:

Date of onset of symptoms: ...../...../..... Date of first consultation: ...../...../.....  
 Fever  Sweats  Chills  Headache  Fatigue  Loss of appetite  
 Abdominal pain  Nausea  Vomiting  Diarrhoea  Jaundice  Eye pain  
 Cough  Pneumonia  Shortness of breath  Chest pain  Sore throat  Any heart problems  
 Joint Pains  Muscle aches  Memory difficulties  Mood changes  Weight loss  
Other symptoms: .....  
Was the patient hospitalised?  Yes  No  Unknown Days hospitalised: ..... Days off work: .....  
Complications:  Yes – *specify* .....  No  Unknown  
Antibiotics:  Yes – *specify* .....  No  Unknown

Case name: .....  
First name Surname

DOB .....

Notification ID: .....

**LABORATORY CRITERIA:**

Laboratory: ..... First collection date: .....

Has there been any previous Q Fever Testing?  Yes  No  Unknown

Lab: ..... Date: ..... Result: .....

Lab: ..... Date: ..... Result: .....

**VACCINATION DETAILS:**

Previous screening:  Yes  No  Unknown Date: ..... Specify: .....

Previous vaccination:  Yes  No  Unknown Date: ..... Specify: .....

Did patient think they were at risk of Q Fever?  Yes  No  Unknown

Was patient aware of the Q Fever vaccination?  Yes  No  Unknown

**EXPOSURE PERIOD:** All questions in this section relate to the month prior to illness onset.

Date: ..... to Date: .....  
(Onset of symptoms - 1 month) (Date of onset of symptoms)

**Abattoir exposure:**

Worked in an abattoir:  Yes  No  Unknown *If Yes, go to next question. If No, go to 'Animal exposure'.*

Duties:  Slaughter floor  Boning  Rendering plant  Producing meat meal or blood and bone

Packer  Cleaner  Maintenance  Other - specify .....

Animals slaughtered:  Cattle  Sheep  Goats  Kangaroo  Other .....

Worked in the grounds of the abattoir:  Yes - list duties .....  No  Unknown

Contract worker at an abattoir:  Yes - list duties .....  No  Unknown

Visitor to an abattoir:  Yes - list duties .....  No  Unknown

**Animal exposure:**

Contact with any of the following animals/insects:  Cattle  Sheep  Domestic goats  Feral goats

Domestic pigs  Feral pigs  Dogs  Cats

Kangaroos  Small marsupials (bandicoots)  Ticks

Other - specify .....

Assisted or observed an animal birth:  Yes - what animal/s .....  No

Involvement in slaughtering, skinning, or meat processing:  Yes - what animal/s .....  No

Any involvement in shooting/hunting:  Yes - what animal/s .....  No

What area hunting in: .....

Worked with wool:  Yes  No Shearing shed  Yes  No Wool processing  Yes  No

Worked with straw or animal bedding:  Yes  No

Worked with animal manure/animal fertiliser e.g. in the garden:  Yes  No

Attended a saleyard or animal show:  Yes - where .....  No

Case name: ..... DOB ...../...../..... Notification ID: .....  
First name Surname

**Environmental exposure:**

- Live on a farm:  Yes  No Visited a farm:  Yes  No
- Exposure to dust from paddocks or animal yards:  Yes  No  Unknown
- Live/work within 1km of an abattoir/animal grazing area/saleyards:  Yes  No  Unknown
- Exposure to trucks for transporting sheep, cattle or goats:  Yes  No
- Laundered clothes from someone who works with animals:  Yes  No
- Had household contact with a Q Fever infected person:  Yes  No  Unknown
- Consumed unpasteurised milk or milk products:  Yes  No  Unknown
- Had contact with untreated water (dams, irrigation sprays):  Yes  No  Unknown
- Details: .....
- Live/work within 300m of a bush/scrub/forest area:  Yes  No  Unknown

Outcome:  Survived  Died Date of death: ...../...../.....  Died of condition  Unknown

**PLACE ACQUIRED:**

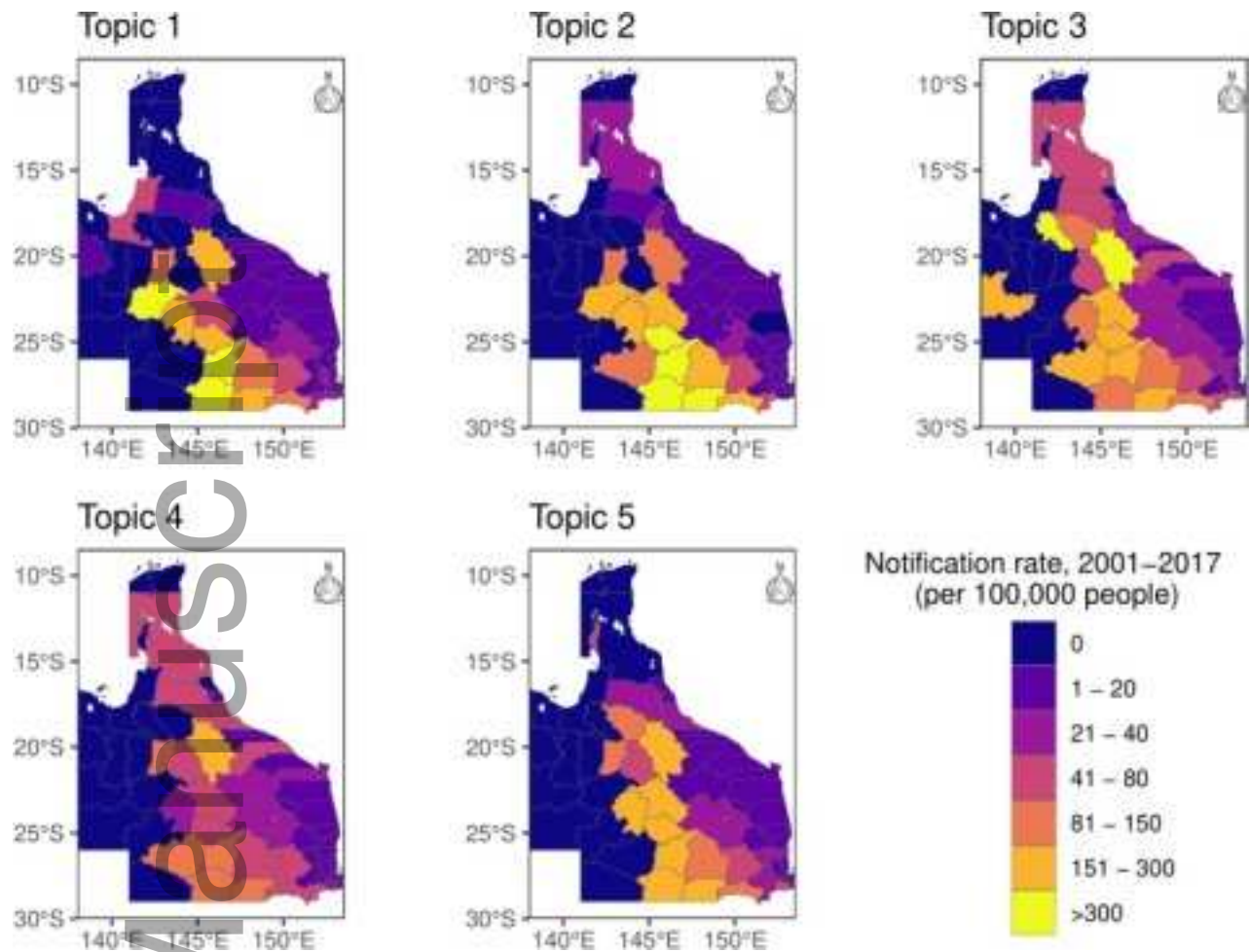
- Queensland  Other Australian state/territory - *specify* .....
- Unknown  Other country - *specify* .....

**NOTIFICATION DECISION:** (see notification criteria)

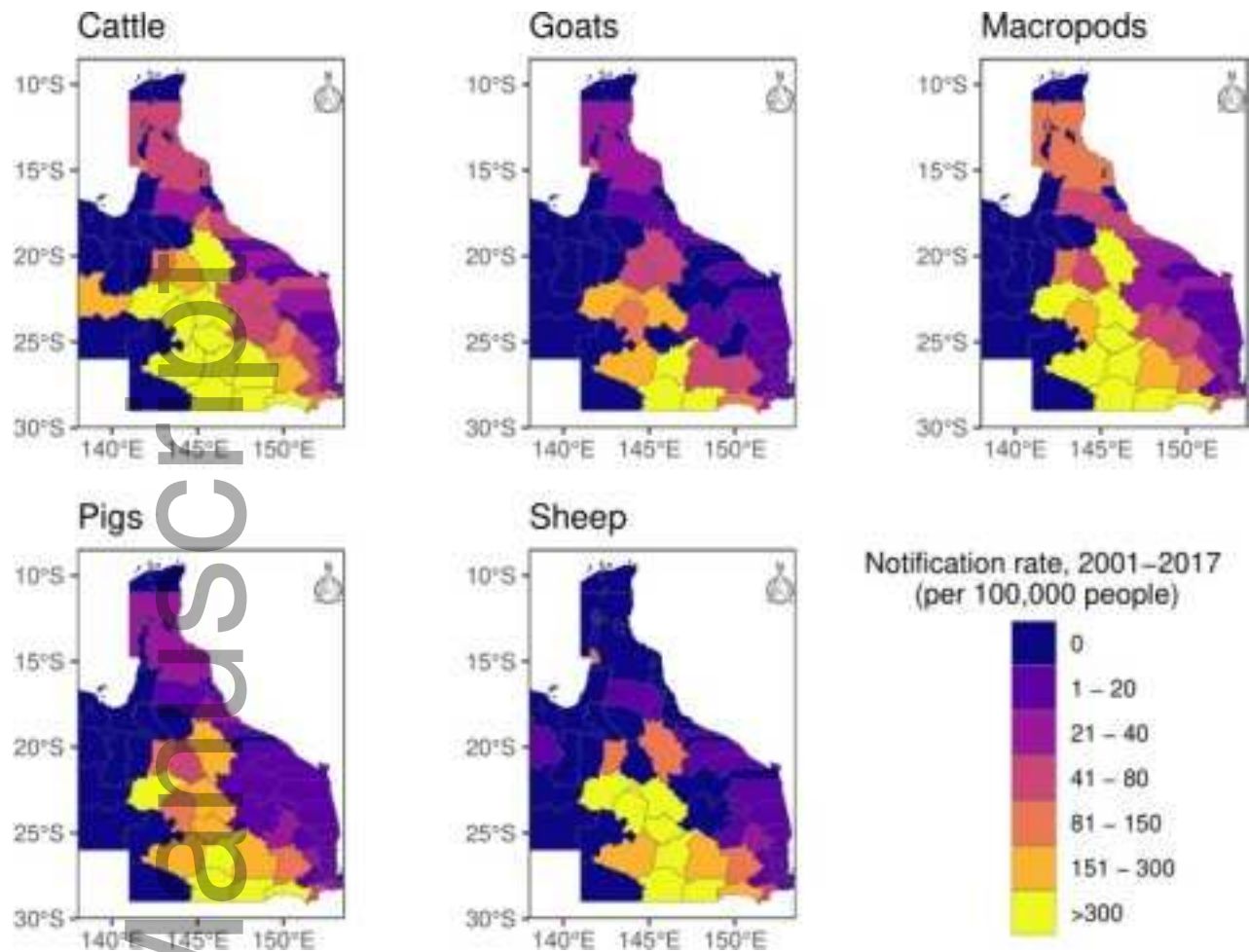
- Confirmed Acute Q Fever  Confirmed Chronic Q Fever  Unlikely to be Q Fever  Q Fever results pending

**COMMENTS:**

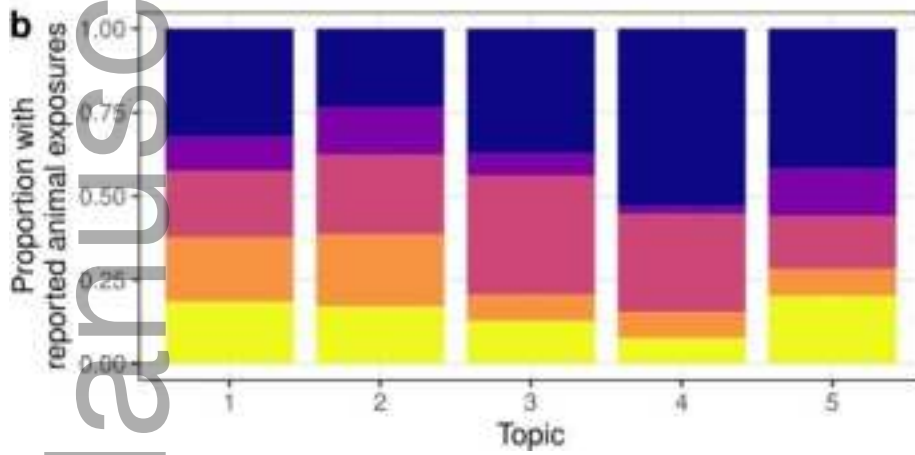
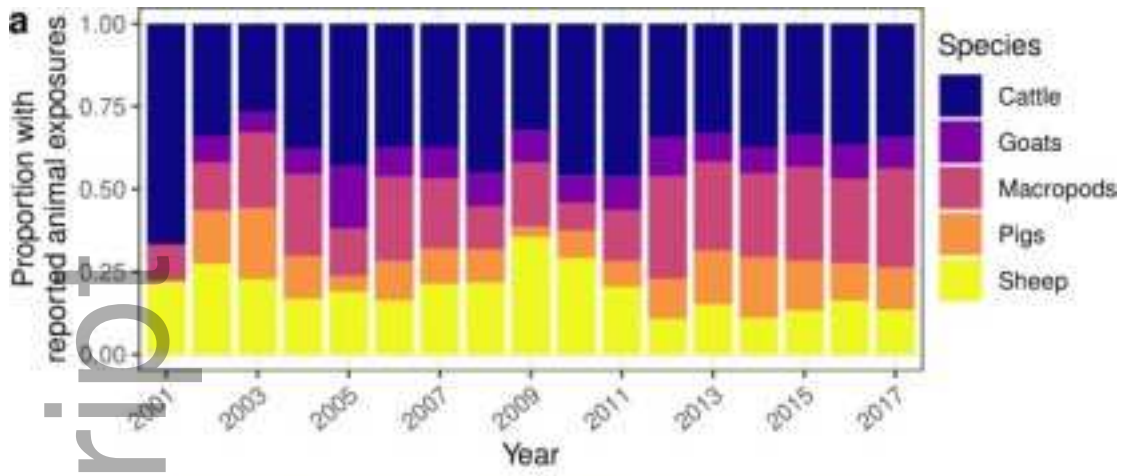




tbed\_13565\_f2.tiff



tbed\_13565\_f3.tiff



tbed\_13565\_f4.tiff

