

Whole-genome sequencing to determine the extent of *Clostridium*  
*difficile* transmission in a high incidence setting in North Wales in  
2015

\*§David W EYRE<sup>2</sup>, §Robert SHAW<sup>1</sup>, Helen ADAMS<sup>3</sup>, Tracey COOPER<sup>4</sup>, Derrick W CROOK<sup>1</sup>,  
Rhonda-Marie GRIFFIN<sup>3</sup>, Phil MANNION<sup>5</sup>, Mari MORGAN<sup>6</sup>, Trefor MORRIS<sup>7</sup>, Michael PERRY<sup>7</sup>,  
Sophie JONES<sup>7</sup>, Tim EA PETO<sup>1</sup>, Jonathan SUTTON<sup>3</sup>, A Sarah WALKER<sup>1</sup>, Dafydd WILLIAMS<sup>3</sup>,  
Noel CRAINE<sup>8</sup>

<sup>1</sup>Nuffield Department of Medicine, University of Oxford, England, UK

<sup>2</sup>Big Data Institute, University of Oxford, England, UK

<sup>3</sup>Betsi Cadwaladr University Health Board, Ysbyty Gwynedd, Bangor, Wales, UK

<sup>4</sup>Betsi Cadwaladr University Health Board, Ysbyty Glan Clwyd, Bangor, Wales, UK

<sup>5</sup>Public Health Wales, Microbiology, Ysbyty Glan Clwyd, Rhyl, Wales, UK

<sup>6</sup>Public Health Wales, Health Protection, Capital Quarter, Cardiff, Wales, UK

<sup>7</sup>Public Health Wales, Microbiology, Heath Hospital, Cardiff, Wales, UK

<sup>8</sup>Public Health Wales, Health Protection, Ysbyty Gwynedd, Bangor, Wales, UK (NC)

§These authors contributed equally.

\*Corresponding author: David Eyre, [david.eyre@ndm.ox.ac.uk](mailto:david.eyre@ndm.ox.ac.uk), +44 1865 220855

23

24 **Running title:** *C. difficile* infection in North Wales

25

## Abstract

### Objectives

Rates of *C. difficile* infection (CDI) are higher in North Wales than elsewhere in the UK. We used whole-genome sequencing to investigate if this is due to increased healthcare-associated transmission from other cases.

### Methods

Healthcare and community *C. difficile* isolates from patients across North Wales (February-July-2015) from glutamate dehydrogenase (GDH)-positive faecal samples underwent WGS. Data from patient records, hospital management systems, and national antimicrobial use surveillance were used.

### Results

338/499(68%) GDH-positive samples were sequenced, and 299 distinct infections/colonisations identified, 229/299(77%) with toxin genes. Only 39/229(17%) toxigenic isolates were related within  $\leq 2$  SNPs to  $\geq 1$  infection/colonisation from a previously sampled patient, i.e. demonstrated evidence of possible transmission. Independent predictors of possible transmission included healthcare exposure in the last 12 weeks ( $p=0.002$ , with varying rates by hospital), infection with multilocus sequence types ST-1 (ribotype-027) and ST-11 (predominantly ribotype-078) compared to all other toxigenic STs ( $p<0.001$ ), and cephalosporin exposure in the potential transmission recipient ( $p=0.02$ ). Adjusting for all these factors, there was no additional effect of ward workload ( $p=0.54$ ), or

48 failure to meet cleaning targets ( $p=0.25$ ). Use of antimicrobials is higher in North Wales  
49 compared to England and the rest of Wales.

50

## 51 **Conclusions**

52 Levels of transmission detected by WGS were comparable to previously described rates in  
53 endemic settings; other explanations, such as variations in antimicrobial use, are required to  
54 explain the high levels of CDI. Cephalosporins are a risk factor for infection with *C. difficile*  
55 by another infected or colonised case.

56

## Introduction

Using whole-genome sequencing in endemic settings only the minority of hospital and community *Clostridium difficile* infections, CDIs, are acquired from other symptomatic cases.<sup>1,2</sup> However, how acquisition from cases varies with increased *C. difficile* incidence is not known. Despite declines in CDI incidence over the last 15 years,<sup>3</sup> North Wales has some of the highest CDI incidence in the United Kingdom; in 2015/16 CDI incidence was 51.1 per 100000 population, compared to a Wales-wide rate<sup>3</sup> of 40.1, 25.8 in England<sup>4</sup> and 31.2 in Scotland<sup>5</sup> (calculated using total reported cases<sup>3-5</sup> and mid-2015 population estimates<sup>6</sup>). Surveillance methodologies in England<sup>4</sup> and Wales<sup>3</sup> are broadly similar. Reporting in Scotland<sup>5</sup> differs by including cases  $\geq 15$  years old, compared to  $\geq 2$  years in England and Wales.

To investigate the relatively high incidence of CDI in North Wales, a prospective WGS study was initiated to test the hypothesis that this was due to increased within-hospital *C. difficile* transmission. We also investigated whether risk factors for transmission could be identified, to suggest potential infection control and other preventative interventions.

## Methods

### Setting

Wrexham Maelor Hospital, Glan Clwyd Hospital, and Ysbyty Gwynedd are three district general hospitals providing secondary-level care to the entire region of North Wales. These hospitals serve a population of 694,473 (mid-year 2015 estimate), living in a mix of urban

and remote rural settings. All hospital and community samples submitted for *C. difficile* testing from these hospitals, smaller community hospitals in the same region, and general practitioners are processed by a single laboratory at Glan Clwyd Hospital. These hospitals are randomly identified as hospital A, B and C to anonymise study results. Hospital policy was to test inpatients >2 years with diarrhoea ( $\geq 3$  unformed stools in 24 hours), without another identified cause, for *C. difficile* infection. Community testing was advised when *C. difficile* was suspected, in particular with a documented history of antibiotic exposure within 6 weeks, in patients from residential or nursing homes or hospital exposure in the last 2 months.

#### Microbiology

Faecal samples submitted for *C. difficile* testing underwent glutamate dehydrogenase, GDH, testing using C. DIFF Chek 60 (TechLab, Blacksburg, VA, USA). Positive samples underwent C. DIFF QUIK CHEK COMPLETE (TechLab) to confirm the GDH result and detect the presence of *C. difficile* toxins A and B by enzyme immunoassay. Samples were saved, selectively cultured for *C. difficile* as described previously,<sup>7</sup> and isolates obtained underwent WGS. GDH-positive patients were considered infected or colonised, and those who were faecal toxin-positive patients to be infected (i.e. have CDI).<sup>8</sup> Cases were denoted healthcare facility-associated, community-associated or indeterminant using standard surveillance definitions.<sup>9</sup> Cases were assigned to a given hospital based on inpatient exposure in the last 12 weeks, excluding the 48 hours immediately prior to diagnosis.

## 102 Sequencing

103 DNA was extracted after subculture of a single colony and sequenced using Illumina  
104 HiSeq2500. Sequence data were processed as previously,<sup>1,10,11</sup> mapping sequenced reads to  
105 the *C. difficile* 630 reference genome. Sequences were compared using SNPs, obtaining  
106 differences between sequences from maximum-likelihood phylogenies,<sup>12</sup> corrected for  
107 recombination using ClonalFrameML.<sup>13</sup> Sequence reads were also assembled *de novo* with  
108 Velvet,<sup>14</sup> using VelvetOptimiser. Toxin genes, *tcdA* and *tcdB*, were identified from *de novo*  
109 assemblies using BLAST searches, on the basis of >1kb of sequence identity to each gene.  
110 Multi-locus sequence types were predicted from *de novo* assemblies.<sup>1</sup> Sequence data have  
111 been deposited under NCBI BioProject PRJNA412541.

112

## 113 Genomic analysis

114 Based on *C. difficile* evolutionary rates and within-host diversity<sup>1,15</sup> >95% of transmission  
115 pairs sampled ≤123 days apart are expected to have ≤2 SNPs between them and cases up to  
116 124-364 days apart ≤3 SNPs, but with 3 SNPs uncommon.<sup>1</sup> Therefore, during the 5.5 months  
117 of the study ≤2 SNPs were expected between the large majority of transmitted isolates.  
118 Where patients had multiple samples, subsequent isolates >10 SNPs different to previous  
119 isolates from the same patient were considered to represent a new acquisition of a distinct  
120 *C. difficile* strain. This higher SNP threshold almost completely excludes isolates being from  
121 the same infection within the study. We used a previously described correction factor<sup>11</sup> to  
122 adjust for sequencing only a subset of GDH-positive samples, assuming sequenced and non-  
123 sequenced cases transmit onwards at the same rate and cases are missing at random.

124

## 125 Risk factor analysis

126 Data from paper and electronic patient records was extracted into a Public Health Wales  
127 data warehouse. Data were available on admissions and ward movements for  
128 infected/colonised patients from the three district general hospitals and for smaller  
129 community hospitals and nursing homes. Additional data were available on prescribing,  
130 ward workload (ward admissions per day), and from mandatory audits of cleaning  
131 compliance within the hospital setting.

132

133 Multivariate logistic regression was used to identify independent predictors of a case being  
134 genetically-related to  $\geq 1$  previous case within  $\leq 2$  SNPs, selecting a final model from factors  
135 shown in Table 1 using backwards elimination with an exit p-value of  $>0.1$ . Multiple  
136 fractional polynomials were used to allow for non-linear effects of continuous factors.  
137 Following initial model selection, each excluded variable was added back to model one at a  
138 time and retained if its Wald p-value was  $<0.1$ . Interactions between main effects in the final  
139 model were retained where interaction  $p < 0.01$ . 1All analyses were performed using Stata  
140 14.1 (Stata Corp, College Station, TX, USA).

141

142 Publicly available demographic<sup>16,17</sup> and antimicrobial surveillance data<sup>18,19</sup> were used to  
143 investigate alternative explanations for variation in CDI incidence. Sex and age adjusted  
144 rates of primary care antibiotic use were compared using items prescribed per 1000 Specific  
145 Therapeutic group Age-sex Related Prescribing Units (STAR-PU)<sup>20</sup>.

146



## Ethics

Ethical approval was not required as the work formed part of the Betsi Cadwaladr University Health Board's response to *C. difficile* infection. Sequencing was carried out on *C. difficile* isolates following routine isolation.

## Results

Between 01-February-2015 and 16-July-2015, 499 *C. difficile* GDH-positive samples were obtained from 417 patients. 182 (36%) samples from 159 patients were faecal toxin-positive and considered to represent infections. One patient had evidence of a genetically distinct second infection. Of these 160 CDIs, 33 (21%) were community-associated, i.e. had no healthcare facility exposure for >12 weeks, representing a rate of 4.8 per 100000 population per year. 118 (74%) were healthcare-facility associated (healthcare exposure within 4 weeks) and 9 (6%) indeterminate (healthcare exposure 4-12 weeks ago), together representing a rate of 5.7 per 10000 bed-days. Monthly CDI incidence, with historic rates,<sup>3</sup> is shown in Figure 1.

Of 499 GDH-positive samples, 338 (68%) underwent WGS (144/182 [79%] faecal toxin-positive samples and 194/317 [61%] faecal toxin-negative samples). Rates of GDH-positive sample retrieval were similar by hospital, 95/136 (70%), 55/81 (68%), 92/134 (69%) at hospitals A, B and C respectively, 5/6 (83%) for patients exposed to both hospital A and C. 6/7 (86%) of samples from patients with only community hospital exposure were retrieved and 85/135 (63%) of samples from patients without recent hospital exposure.

Considering all GDH-positive samples, irrespective of faecal toxin status, the 338 sequenced samples contained 299 distinct infections/colonisations in 290 patients. Of these, 229/299 (77%) had detectable toxin genes on WGS, and within these potentially toxigenic isolates, 114/229 (50%) were from consistently faecal toxin-positive patients, 103/229 (45%) from consistently faecal toxin-negative patients, and 12/229 (5%) from patients with both faecal toxin-positive and negative results on different samples. Of the 70 distinct colonisations without detectable toxin genes on WGS, 65 (93%) were consistently faecal toxin-negative, 4 (6%) were faecal toxin-positive and 1 (1%) had both faecal toxin-positive and toxin-negative results on different samples.

#### Genetic and epidemiological links between samples

Of the 299 sequenced distinct infections/colonisations, 43 (14%) were within  $\leq 2$  SNPs of  $\geq 1$  infection/colonisation from a previously sampled patient, i.e. had evidence of possible transmission (Figure 2). 39/43 (91%) genetically-linked cases were toxigenic (i.e. had toxin genes), and 39/229 (17%) distinct toxigenic infections/colonisations were within  $\leq 2$  SNPs of  $\geq 1$  infection/colonisation. Figure 3 shows the relationship between donor and recipient faecal toxin status. Faecal toxin-positive cases were not more likely to have a faecal toxin-positive donor; instead faecal toxin-negative recipients had predominantly positive donors, and some faecal toxin-positive recipients had faecal toxin-negative donors ( $p=0.006$  versus no relationship between donor and recipient toxin status). Of the 43 potentially transmitted infections/colonisations, 26 (60%) had a single possible source, 9 (21%) had 2 possible sources, and 4 (9%), 3 (7%), and 1 (2%) had 3, 4 and 5 possible sources respectively. The

median (IQR) [range] time from the most recently sampled case within  $\leq 2$  SNPs of the potential recipient was 21 (7-47) [0–117] days.

Healthcare exposure in the 12 weeks prior to diagnosis was an important predictor of genetic linkage to a previous case; 40/217 (18%) patients with healthcare exposure were genetically linked to a previous case, compared to 3/82 (4%) without ( $p=0.001$ , Figure 4A). Rates of genetic linkage to previous cases varied at the 3 hospitals: 9/80 (11%), 11/51 (22%), 20/75 (27%) at hospitals A, B and C respectively ( $p=0.04$ , Figure 4A). Transfers between hospitals were uncommon; five patients were exposed to both hospital A and C, and six patients only to smaller community hospitals; none of these 11 patients were genetically linked to a previous case. Genetic linkage did not correspond to the overall rates of healthcare-associated/indeterminant GDH-positive *C. difficile* colonisation/infection at hospitals A, B and C, which were 16.8, 11.3, 12.2 per 10000 bed-days respectively or to faecal toxin-positive CDI, occurring at 7.0, 5.2, 5.2 per 10000 bed-days, respectively.

Of the 43 genetically linked cases, 11 (26%) shared time and space on the same hospital ward with their potential donor between the dates of their diagnoses, 9 in a district general hospital and 2 in a community hospital (Figure 2). A further 2/43 (5%) patients shared time and space on the same district general hospital ward before either were diagnosed. Another 8/43 (19%) patients shared the same ward location at different times within the 28 days prior to diagnosis, 5 in a district general hospital, 2 in a community hospital and 1 in a nursing home. Finally, 2/43 (5%) patients without any other link shared time in the same district general hospital between the dates of their diagnoses, but not specific wards. Thus 20/43 (47%) potential recipients had no recent or concurrent shared healthcare exposure

with any previous case within  $\leq 2$  SNPs even at the broadest level of the hospital, and accounting for smaller community hospitals and nursing homes.

The most commonly occurring toxigenic STs were: ST-6 (30/229 toxigenic infections/colonisations, 13%); ST-2 (27, 12%); ST-8 (21, 9%); and ST-44 (18, 8%) all from *C. difficile* clade 1<sup>21</sup>; and ST-11 (19, 8%, equivalent most commonly to ribotype-078) and ST-1 (18, 8%, ribotype-027). Rates of genetic linkage were higher in ST-1 and ST-11 than the combined group of all other toxigenic STs (Figure 4B,  $p < 0.001$ ). Rates of genetic linkage were lower for non-toxigenic *C. difficile* despite all tested patients having diarrhoea.

Similar percentages of sequenced infections/colonisations after the first 3 months of the study were within  $\leq 2$  SNPs of an earlier sequenced case (20/123 [16%] versus 23/176 [13%] before,  $p = 0.27$ ) even though cases earlier in the study may have been less likely to have had their source sampled. We applied a previously published correction<sup>11</sup> to adjust for having only sequenced 68% of *C. difficile*-positive samples. This provided a corrected estimate for the percentage of cases after the first 3 months of the study that were genetically-linked to a prior case of 24% (i.e.  $20/123 * 1/0.68$ ). Restricting only to potentially toxigenic cases, this figure was 30% ( $16/87 * 1/0.68$ ).

#### Risk factors for transmission

Independent risk factors for genetic linkage within  $\leq 2$  SNPs to a previous case (Table 1) included healthcare exposure in the last 12 weeks, in hospital A (OR 3.15 [95%CI 0.77-12.9]), in hospital B (5.63 [1.40-22.7]), and in hospital C (10.1 [2.75-37.4]), compared to no

healthcare exposure ( $p=0.002$ ). *C. difficile* genotype was also associated with genetic linkage ( $p<0.001$ ); compared to all other toxigenic STs, ST-1 cases were independently more likely to be linked to a previous ST1 case (OR 7.61 [95%CI 2.50-23.2]), and there was some evidence for similar associations for ST-11 (2.27 [0.67-7.68]). Older patients were somewhat more likely to be genetically linked to a previous case ( $p=0.06$ ). Second/third-generation cephalosporin exposure in the last 90 days in the potential transmission recipient increased the risk of genetic linkage (OR 6.03 [95%CI 1.42-25.5,  $p=0.02$ ]), however only 5/43 (12%) of cases and 6/256 (2%) of controls were exposed. Adjusting for all these factors, within the limits of the power of the study, we found no evidence for any additional effects on transmission of ward workload ( $p=0.54$ ), or failure to meet cleaning audit targets ( $p=0.25$ ).

#### Population risk factors for CDI

We considered explanations other than increased transmission for rates of CDI in North Wales. The majority of antibiotics in the UK are prescribed in primary care by general practitioners. Rates of community antibiotic use (in the second quarter of 2015) were higher in North Wales (296.7 per 1000 STAR-PUs) and Wales overall (296.9 per 1000 STAR-PUs), compared to England (243 items per 1000 STAR-PUs) (Figure 5).<sup>18</sup> Comparing acute hospital total antibiotic use in defined daily doses per 1000 bed-days in 2015, for the 17 acute hospitals in Wales, Ysbyty Gwynedd had the 2<sup>nd</sup> highest rate, Ysbyty Glan Clwyd the 6<sup>th</sup> highest and Wrexham Maelor the 12<sup>th</sup>.<sup>19</sup>

Similarly, age is another CDI risk factor. The population in North Wales is older than Wales as a whole, 22.6% of the population are >65 years old,<sup>16</sup> compared to Wales 20.4% and England 17.9% (mid-2016 data).<sup>17</sup>

## Discussion

Despite high CDI incidence in North Wales, based on WGS only 39/229 (17%) of toxigenic infections/colonisations could have been plausibly acquired from another case. Adjusting for only sequencing 68% of isolates, this proportion was still only 30%. This is higher than in a study of six English hospitals, where rates of genetic linkage to previous cases were between 7% and 24% by hospital and 20% overall.<sup>11</sup> However these differences are insufficient to explain CDI incidence being nearly double in North Wales compared to England.<sup>3,4</sup> Therefore, higher incidence is likely to be driven predominantly by factors other than lapses in infection control.

Antimicrobial exposure is an important CDI risk factor.<sup>22</sup> Rates of antibiotic use in primary care are higher in Wales than in England, but similar in North Wales to Wales overall, potentially explaining some of the differences between North Wales and England, but not between North Wales and elsewhere in Wales. Additionally, two of the three hospitals in North Wales are among the highest users of antibiotics of all the acute hospitals in Wales. Similarly, increasing age is another important risk factor for CDI,<sup>22</sup> and the population in North Wales is older than Wales as a whole and England. Other factors may also be important; the area of the country served by the three hospitals contains extensive areas of livestock farming. Disease causing *C. difficile* strains have been isolated from livestock,<sup>23</sup>

with overlap seen between isolates from CDI cases, healthy humans and livestock on WGS.<sup>24</sup> However a large scale environmental survey 20 years ago in South Wales identified relatively little *C. difficile* in livestock.<sup>25</sup> Asymptomatic patients are another potential source of *C. difficile* infection, however it is not known if rates of *C. difficile* colonization differ across the UK.

Recent healthcare exposure was an important risk factor for potential acquisition from a previous case; 40/43 (93%) of genetically-related cases were in hospital in the 12 weeks prior to their diagnosis. The median (IQR) time between genetically-related cases was 21 (7-47) days. However shared space and time on the same hospital ward could only explain the minority of genetically-related cases, and nearly half such cases had no healthcare contact including allowing for shared time in hospital resulting in overlap outside of wards, e.g. diagnostic areas. Additionally, although our study is only moderately powered, we found no signal that failure to meet cleaning audit targets or high levels of patient turnover were associated with more transmission. However, the proportion of cases linked to a previous case varied between 11% and 27% at the three main hospitals suggesting potential for reductions in overall incidence. Supporting the previously described role in transmission of GDH-positive patients without detectable faecal toxin,<sup>26</sup> 7/39 (18%) of toxigenic *C. difficile* acquisitions could only be linked to consistently toxin-negative sources. Therefore, all patients with toxigenic *C. difficile* should be a focus of infection control efforts, not just those with detectable faecal toxin. ST-1 (ribotype-027) and ST-11 (ribotype-078) were associated with higher rates of genetic linkage replicating previous findings from England<sup>27</sup> and for ST-1 from Canada.<sup>28</sup> The underlying reasons for this may be multifactorial including more severe disease<sup>29</sup> leading to greater environmental contamination, enhanced

environmental persistence, and also a greater likelihood of clinically detectable disease in transmission recipients.

Antimicrobials are risk factors for CDI.<sup>22</sup> We investigated more specifically the effect of recent antimicrobial exposure on acquisition of *C. difficile* from another case. Second/third generation cephalosporin exposure, but not antibiotics in general or any other specific antibiotic class, increased the risk of being a transmission recipient. The effect of cephalosporins may reflect intrinsic resistance in *C. difficile*,<sup>30</sup> and more variable susceptibility to other antibiotics in the population studied.

The main limitation of this study is that only 68% of samples tested were available for sequencing; this was due to a failure by the research team to ensure all samples taken for diagnostic purposes were successfully processed prior to sequencing within the study. This will have reduced the observed rates of linkage to previous cases, as demonstrated in previous simulations.<sup>11</sup> However, by applying a correction factor for missing data we were able to estimate the true proportion of cases linked. As rates of sample retrieval for sequencing were similar between the three hospitals, the differences observed in linkage rates are unlikely to have been differentially affected by sample retrieval rates at each site. The small number of samples, 5/299 (2%) infections/colonisations, that were faecal toxin positive, but yielded isolates that lacked toxin genes on sequencing may have arisen as a result of mixed infections, laboratory error or a false-positive faecal toxin assay. Mixed infections are a potential additional source of underestimates of the extent of transmission from other cases, but previous work suggests this is uncommon in *C. difficile*.<sup>31</sup>



331 This study was based on prospective storage of samples, culture of isolates and sequencing  
332 in response to a period of high CDI incidence. An alternative approach which may allow  
333 similar methods to be applied more widely is the storage of *C. difficile* GDH-positive faecal  
334 samples, e.g. on a rolling annual basis. These can then be cultured and sequenced  
335 retrospectively if increased incidence is noted, as demonstrated recently in six English  
336 hospitals.<sup>11</sup> The development of surveillance systems that interpret CDI incidence and  
337 sequencing data and present it back to clinicians in a timely manner is essential to guide  
338 local and national infection prevention and control responses.

339

340 In summary, despite relatively high CDI incidence in North Wales, levels of transmission  
341 detected by WGS were comparable to previously described rates in endemic settings; other  
342 explanations, including variations in antimicrobial use, are required to understand the  
343 reasons for the high levels of CDI.

344

## 345 [Acknowledgements](#)

346 We would like to thank Dai Griffiths, Kim Bowden, Helen Booth, Gary Porter Jones, Melissa  
347 Van Der Bijl and Sarah Davies for their support.

348

## 349 [Funding](#)

350 The study was funded by the North Wales Awyr Las charitable funds and by NHS (Wales)  
351 Research and Development support costs. DWE is a NIHR Clinical Lecturer and a Robertson  
352 Foundation Fellow. TEAP is an NIHR Senior Investigator.

353

## 354 [Transparency declaration](#)

355 No author has any conflict of interest to declare.

356

## References

1. Eyre DW, Cule ML, Wilson DJ, *et al.* Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 2013;**369**:1195–205.
2. Kumar N, Miyajima F, He M, *et al.* Genome-Based Infection Tracking Reveals Dynamics of *Clostridium difficile* Transmission and Disease Recurrence. *Clin Infect Dis* 2016;**62**:746–52.
3. Public Health Wales. Welsh Healthcare Associated Infection Programme (WHAIP) - *Clostridium difficile* and *Staphylococcus aureus* bacteraemia Surveillance Update. <http://www.wales.nhs.uk/sites3/page.cfm?orgid=379&pid=67899>
4. Public Health England. Annual Epidemiological Commentary. Mandatory MRSA, MSSA and *E. coli* bacteraemia and *C. difficile* infection data 2016/17. Available from: <http://webarchive.nationalarchives.gov.uk/20180410202808/https://www.gov.uk/government/statistics/mrsa-mssa-and-e-coli-bacteraemia-and-c-difficile-infection-annual-epidemiological-commentary>
5. Health Protection Scotland. Healthcare Associated Infection Annual Report 2015. <http://www.hps.scot.nhs.uk/haic/sshaip/resourcedetail.aspx?id=1717>
6. Office for National Statistics. Population estimates and components of population change. Detailed time series 2001 to 2015. <https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/populationandmi>

377       gration/populationestimates/datasets/populationestimatesforukenglandandwalessco  
378       tlandandnorthernireland/mid2015/ukmye2015.zip

379    7.     Griffiths D, Fawley W, Kachrimanidou M, *et al.* Multilocus sequence typing of  
380       *Clostridium difficile*. J Clin Microbiol 2010;**48**:770–8.

381    8.     Planche TD, Davies KA, Coen PG, *et al.* Differences in outcome according to  
382       *Clostridium difficile* testing method: a prospective multicentre diagnostic validation  
383       study of *C difficile* infection. Lancet Infect Dis 2013;**13**:936–45.

384    9.     McDonald LC, Coignard B, Dubberke E, *et al.* Recommendations for surveillance of  
385       *Clostridium difficile*-associated disease. Infect Control Hosp Epidemiol 2007;**28**:140–5.

386    10.    De Silva D, Peters J, Cole K, *et al.* Whole-genome sequencing to determine  
387       transmission of *Neisseria gonorrhoeae*: an observational study. Lancet Infect Dis  
388       2016;**16**:1295–303.

389    11.    Eyre DW, Fawley WN, Rajgopal A, *et al.* Comparison of Control of *Clostridium difficile*  
390       Infection in Six English Hospitals Using Whole-Genome Sequencing. Clin Infect Dis  
391       2017;**65**:433–41.

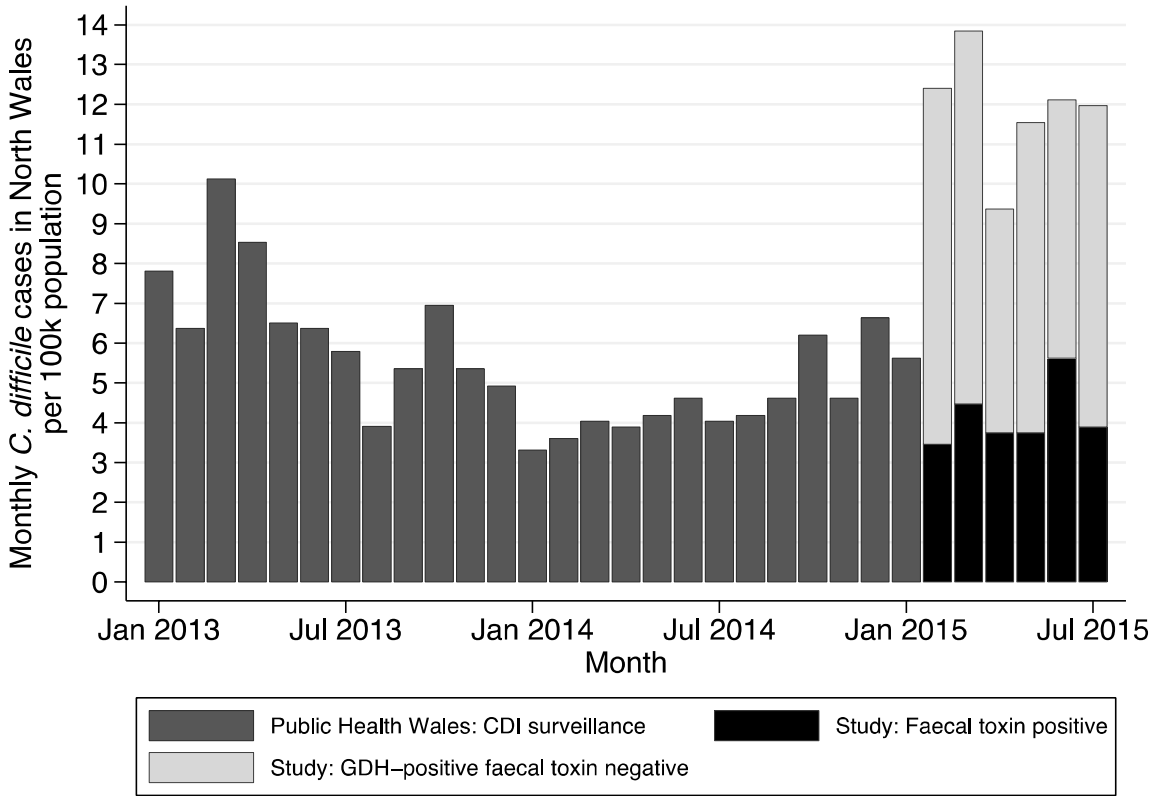
392    12.    Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large  
393       phylogenies by maximum likelihood. Syst Biol 2003;**52**:696–704.

394    13.    Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole  
395       bacterial genomes. PLoS Comput Biol 2015;**11**:e1004041.

- 396 14. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de  
397 Bruijn graphs. *Genome Res* 2008;**18**:821–9.
- 398 15. Eyre DW, Fawley WN, Best EL, *et al.* Comparison of multilocus variable-number  
399 tandem-repeat analysis and whole-genome sequencing for investigation of  
400 *Clostridium difficile* transmission. *J Clin Microbiol* 2013;**51**:4141–9.
- 401 16. Stats Wales. National level population estimates by year, age and UK country.  
402 [https://statswales.gov.wales/Catalogue/Population-and-](https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-ukcountry)  
403 [Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-](https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-ukcountry)  
404 [ukcountry](https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-ukcountry)
- 405 17. Office for National Statistics. Population Estimates for UK, England and Wales,  
406 Scotland and Northern Ireland: mid-2016.  
407 [https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/p](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2016)  
408 [opulationestimates/bulletins/annualmidyearpopulationestimates/mid2016](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2016)
- 409 18. NHS Business Services Authority. Information Services Portal.  
410 <https://apps.nhsbsa.nhs.uk/infosystems/welcome>
- 411 19. Public Health Wales. Antimicrobial Usage in Secondary Care in Wales 2006-2015.  
412 [https://www.wales.nhs.uk/sitesplus/documents/888/Antibacterial%20Usage%20in%](https://www.wales.nhs.uk/sitesplus/documents/888/Antibacterial%20Usage%20in%20Secondary%20Care%20in%20Wales%202006-2015.pdf)  
413 [20Secondary%20Care%20in%20Wales%202006-2015.pdf](https://www.wales.nhs.uk/sitesplus/documents/888/Antibacterial%20Usage%20in%20Secondary%20Care%20in%20Wales%202006-2015.pdf)
- 414 20. Lloyd DCEF, Harris CM, Roberts DJ. Specific therapeutic group age-sex related  
415 prescribing units (STAR-PU): weightings for analysing general practices' prescribing in  
416 England. *BMJ* 1995;**311**:991–4.

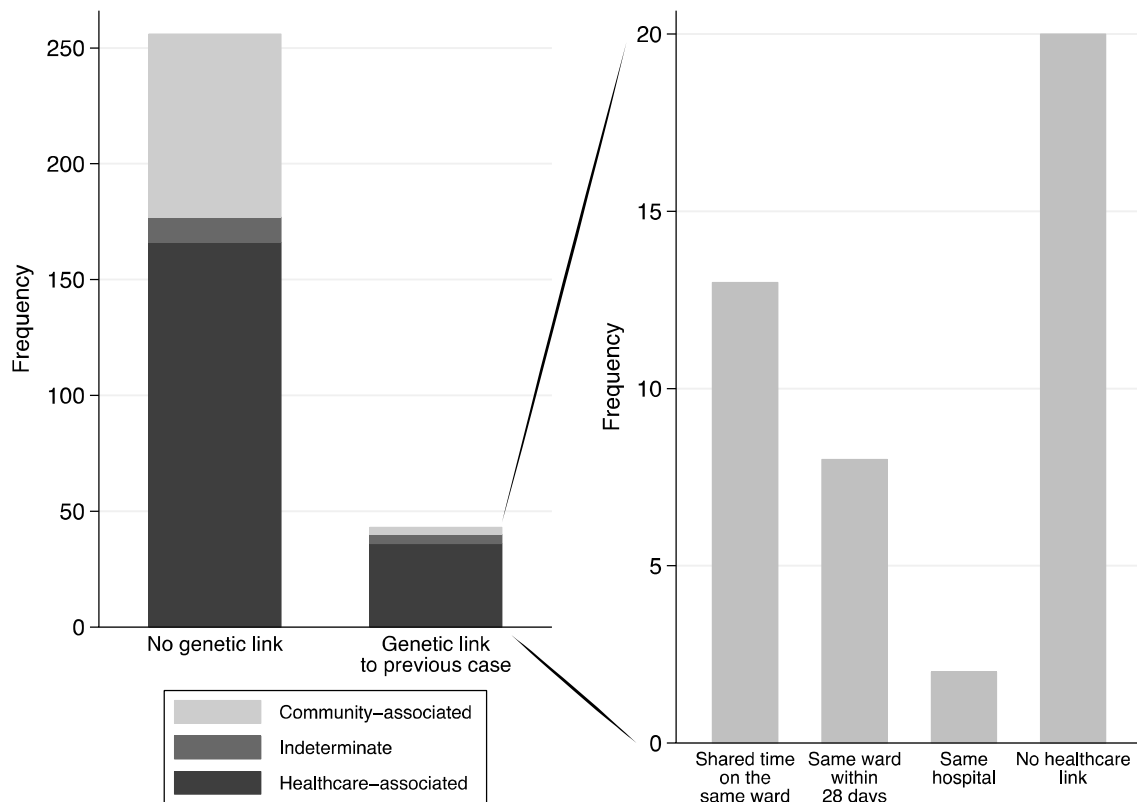
- 417 21. Dingle KE, Elliott B, Robinson E, *et al.* Evolutionary History of the *Clostridium difficile*  
418 Pathogenicity Locus. *Genome Biol Evol* 2014;**6**:36–52.
- 419 22. Loo VG, Bourgault A-M, Poirier L, *et al.* Host and pathogen factors for *Clostridium*  
420 *difficile* infection and colonization. *N Engl J Med* 2011;**365**:1693–703.
- 421 23. Hensgens MPM, Keessen EC, Squire MM, *et al.* *Clostridium difficile* infection in the  
422 community: a zoonotic disease? *Clin Microbiol Infect* 2012;**18**:635–45.
- 423 24. Knetsch CW, Connor TR, Mutreja A, *et al.* Whole genome sequencing reveals  
424 potential spread of *Clostridium difficile* between humans and farm animals in the  
425 Netherlands, 2002 to 2011. *Euro Surveill* 2014;**19**:20954–262.
- 426 25. Saif al N, Brazier JS. The distribution of *Clostridium difficile* in the environment of  
427 South Wales. *J Med Microbiol* 1996;**45**:133–7.
- 428 26. Mawer DPC, Eyre DW, Griffiths D, *et al.* Contribution to *Clostridium difficile*  
429 Transmission of Symptomatic Patients With Toxigenic Strains Who Are Fecal Toxin  
430 Negative. *Clin Infect Dis* 2017;**64**:1163–70.
- 431 27. Martin JSH, Eyre DW, Fawley WN, *et al.* Patient and strain characteristics associated  
432 with *Clostridium difficile* transmission and adverse outcomes. *Clin Infect Dis* 2018;  
433 **67**:1379-1387
- 434 28. Kong LY, Eyre DW, Corbeil J, *et al.* *Clostridium difficile*: Investigating Transmission  
435 Patterns between Infected and Colonized Patients using whole Genome Sequencing.  
436 *Clin Infect Dis* 2018 [epub ahead of print].

- 437 29. Walker AS, Eyre DW, Wyllie DH, *et al.* Relationship Between Bacterial Strain Type,  
438 Host Biomarkers, and Mortality in *Clostridium difficile* Infection. Clin Infect Dis  
439 2013;**56**:1589–600.
- 440 30. Freeman J, Wilcox MH. Antibiotic activity against genotypically distinct and  
441 indistinguishable *Clostridium difficile* isolates. J Antimicrob Chemother 2001;**47**:244–  
442 6.
- 443 31. Eyre DW, Cule ML, Griffiths D, *et al.* Detection of Mixed Infection from Bacterial  
444 Whole Genome Sequence Data Allows Assessment of Its Role in *Clostridium difficile*  
445 Transmission. PLoS Comput Biol 2013; **9**: e1003059.
- 446
- 447



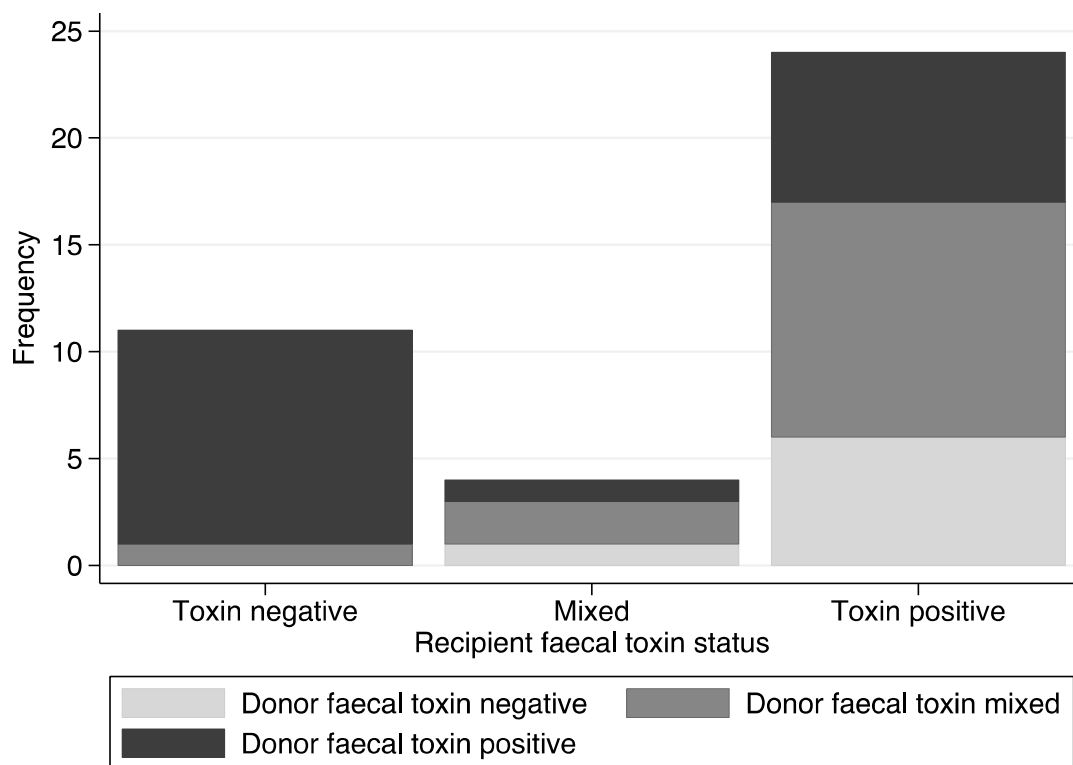
**Figure 1. *C. difficile* incidence in North Wales 2013-2015.** Public Health Wales surveillance data are for faecal toxin positive CDI cases.





**Figure 2. Proportion of cases genetically linked within  $\leq 2$  SNPs, classified by surveillance definitions and epidemiological relationships between linked cases.** Cases sharing the same ward or hospital did so with their potential donor between the dates of their diagnoses, or prior to the diagnosis of either case. For cases sharing the same ward within 28 days, the potential recipient spent time on the same hospital ward after the discharge of an already diagnosed donor.

462



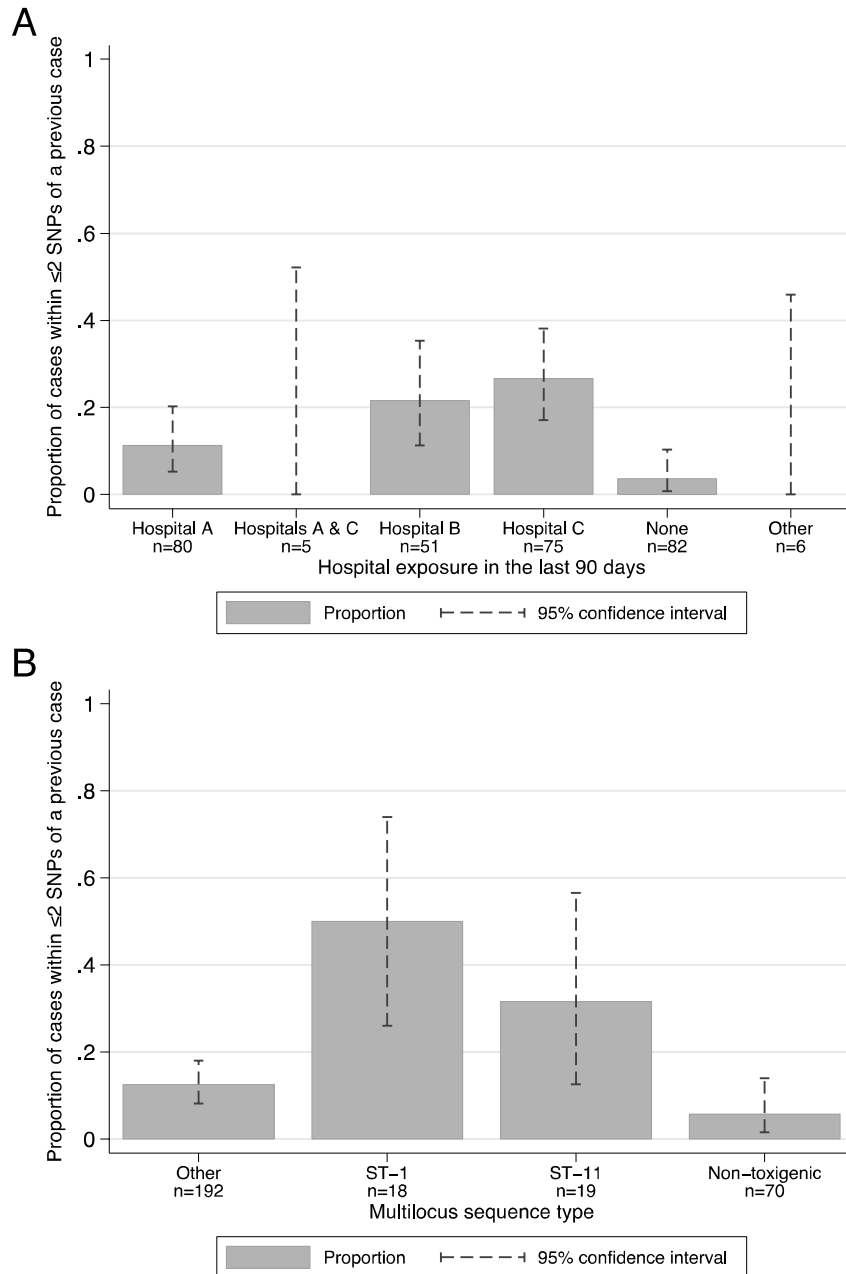
463

464 **Figure 3. Relationship between potential transmission donor and recipient faecal toxin**

465 **status.** A mixed toxin status patient had  $\geq 1$  faecal toxin positive and  $\geq 1$  faecal toxin negative

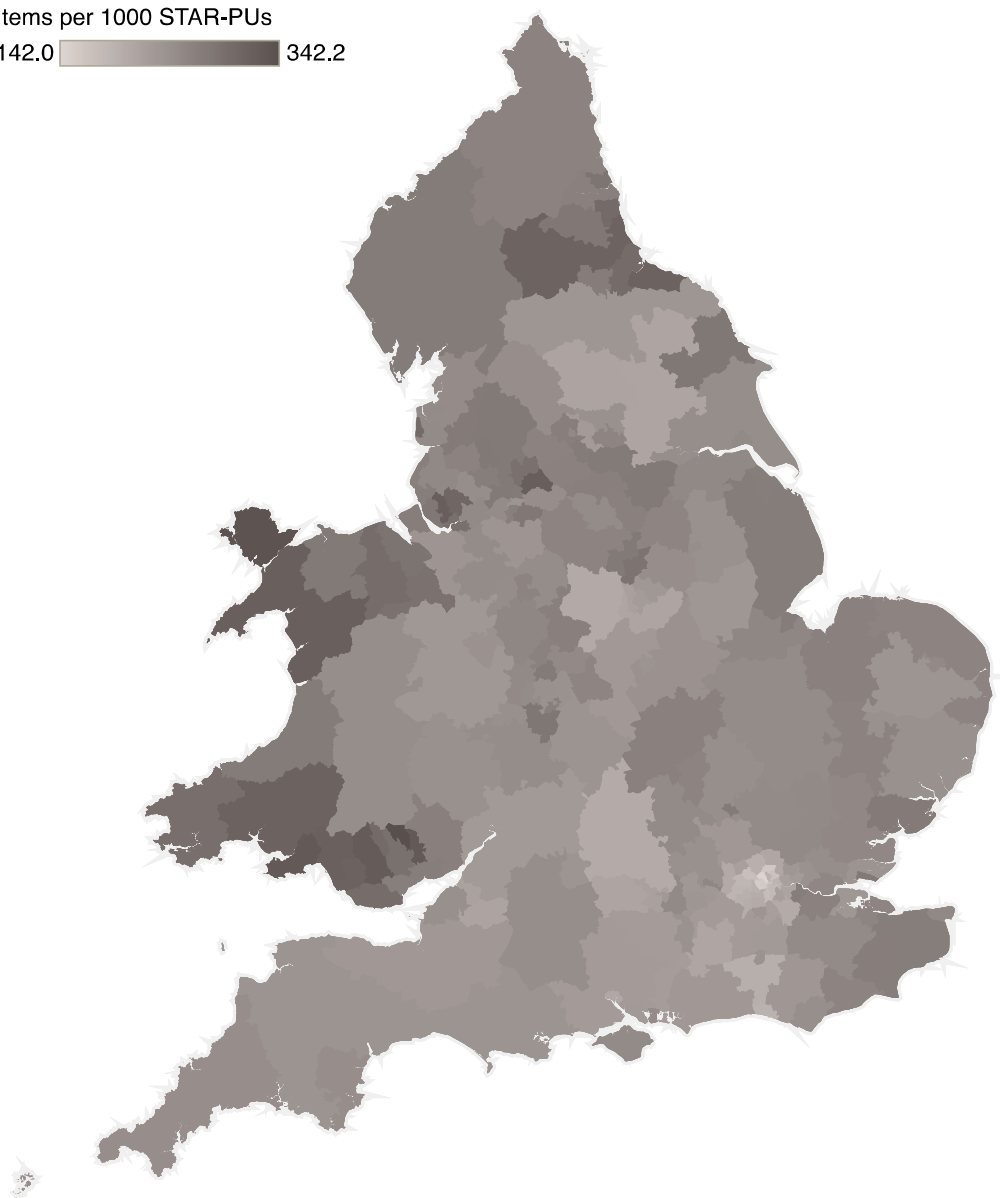
466 sample. Overall p value = 0.006.

467



**Figure 4. Unadjusted proportion of cases with a previous case within  $\leq 2$  SNPs, by hospital exposure in the last 12 weeks (panel A) and multilocus sequence type (panel B).**

Items per 1000 STAR-PUs  
142.0 342.2



473

474 **Figure 5. Antibiotic prescribing in primary care in England and Wales, items prescribed per**

475 **1000 Specific Therapeutic group Age-sex Related Prescribing Units (STAR-PUs).** Data are

476 presented for July – September 2015. Areas shaded are Welsh Unitary Authorities and

477 English Clinical Commissioning Groups. Source: <sup>18</sup>.

478

	Genetically-unlinked (N=256)		Genetically-linked (N=43)		Univariate			Multivariate		
	n / median	% / IQR	n / median	% / IQR	Odds ratio	95% Confidence interval	p value	Odds ratio	95% Confidence interval	p value
<b>Any hospital exposure in last 12 weeks</b>										
- None	79	31%	3	7%	1.00	Baseline	0.001	1.00	Baseline	0.002
- in hospital A	71	28%	9	21%	3.34	(0.87, 12.82)		3.15	(0.77, 12.87)	
- in hospital B	40	16%	11	26%	7.24	(1.91, 27.44)		5.63	(1.40, 22.68)	
- in hospital C	55	21%	20	47%	9.58	(2.71, 33.80)		10.13	(2.75, 37.39)	
- in both hospitals A and C	5	2%	0	0%	*					
- in community hospital only	6	2%	0	0%	*					
<b>Multilocus sequence type</b>										
- Other	168	66%	24	56%	1.00	Baseline	<0.001	1.00	Baseline	<0.001
- ST-1	9	4%	9	21%	7.00	(2.53, 19.38)		7.61	(2.50, 23.16)	
- ST-11	13	5%	6	14%	3.23	(1.12, 9.30)		2.27	(0.67, 7.68)	
- Non-toxigenic	66	26%	4	9%	0.42	(0.14, 1.27)		0.36	(0.11, 1.17)	
Sex, female	166	68%	28	65%	0.91	(0.46, 1.80)				
Age, years	79	69 - 86	82	71 - 88	1.02	(1.00, 1.05)	0.06	1.03	(1.00, 1.05)	0.06
Recipient faecal toxin positive	101	39%	28	65%	2.86	(1.46, 5.63)	0.002			
Inpatient days in last 90 days	12	3.5 - 25	17.5	8 - 41	1.02	(1.00, 1.04)	0.03			
Any antibiotic	142	58%	28	65%	1.36	(0.69, 2.67)	0.37			
Fluoroquinolone	21	9%	5	12%	1.47	(0.52, 4.14)	0.46			
Cephalosporin, 2nd/3rd generation	6	2%	5	12%	5.48	(1.59, 18.85)	0.007	6.03	(1.42, 25.50)	0.02
Beta-lactam/beta-lactamase inhibitor	61	25%	15	35%	1.54	(0.78, 3.06)	0.22			
Meropenem	10	4%	4	9%	2.52	(0.75, 8.44)	0.13			
Proton pump inhibitor	35	14%	6	14%	1.02	(0.40, 2.60)	0.96			
Laxative	18	7%	3	7%	0.99	(0.28, 3.52)	0.99			
Cleaning audit, per day below target	10	2 - 24.5	16.5	7 - 36	1.01	(1.00, 1.03)	0.06			
Admissions, per admission exposed to	56	21.5 - 110	85	37 - 141	1.00	(1.00, 1.01)	0.07			

**Table 1. Risk factors for genetic linkage ( $\leq 2$  SNPs) with a previous case.** Antibiotic and proton pump exposures are ever receiving the relevant agent in the 90 days prior to diagnosis, and laxative exposure in the 30 days prior to diagnosis. Cleaning audit exposure is the total number of days in the 90 days prior to diagnosis spent on a ward that had failed to meet the audit standard at the last available audit. Ward workload was

judged by the total number of other patient admissions that occurred during all inpatient days in the 90 days prior to diagnosis. \*These hospital exposure categories had no genetic links and so an odds ratio cannot be calculated.