



Collin, S. M. et al. (2021) Uncovering infant group B streptococcal (GBS) disease clusters in the UK and Ireland through genomic analysis: a population-based epidemiological study. *Clinical Infectious Diseases*, 72(9), e296-e302.

(doi: [10.1093/cid/ciaa1087](https://doi.org/10.1093/cid/ciaa1087))

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<https://eprints.gla.ac.uk/221630/>

Deposited on: 3 August 2020

# Uncovering infant group B streptococcal (GBS) disease clusters in the UK and Ireland through genomic analysis: a population-based epidemiological study

Simon M Collin,<sup>1</sup> Natalie Groves,<sup>2</sup> Catherine O’Sullivan,<sup>3</sup> Elita Jauneikaite,<sup>4,5</sup> Darshana Patel,<sup>2</sup> Robert Cunney,<sup>6,7</sup> Mary Meehan,<sup>7</sup> Arlene Reynolds,<sup>8</sup> Andrew Smith,<sup>9,10</sup> Diane Lindsay,<sup>10</sup> Lorraine Doherty,<sup>11</sup> Eleri Davies,<sup>12</sup> Victoria Chalker,<sup>2,4</sup> Peter Lamb<sup>1</sup>, Baharak Afshar,<sup>2</sup> Sooria Balasegaram,<sup>13</sup> Juliana Coelho,<sup>2</sup> Derren Ready,<sup>2</sup> Colin S Brown,<sup>1</sup> Androulla Efstratiou,<sup>4,14</sup> Kirsty Le Doare,<sup>3,15,16</sup> Shiranee Sriskandan,<sup>4,17</sup> Paul T Heath,<sup>3</sup> Theresa Lamagni<sup>1,4</sup>

1. Healthcare-Associated Infection and Antimicrobial Resistance (HCAI & AMR) Division, National Infection Service, Public Health England, London, UK
2. Respiratory and Vaccine Preventable Bacteria Reference Unit, National Infection Service, Public Health England, London, UK
3. Paediatric Infectious Diseases Research Group, St George’s University of London, London, UK
4. NIHR Health Protection Unit in Healthcare-associated Infections and Antimicrobial Resistance, Imperial College, London, UK
5. Department of Infectious Disease Epidemiology, School of Public Health, Imperial College, London, UK
6. Health Service Executive Health Protection Surveillance Centre, Dublin, Ireland
7. Irish Meningitis and Sepsis Reference Laboratory, Children’s Health Ireland at Temple Street, Dublin, Ireland
8. Health Protection Scotland, Glasgow, UK
9. College of Medical, Veterinary & Life Sciences, Glasgow Dental Hospital & School, University of Glasgow, Glasgow, UK

10. Scottish Microbiology Reference Laboratory, Glasgow, UK
11. Public Health Agency Northern Ireland, Belfast, UK
12. Public Health Wales, Cardiff, UK
13. Field Service, National Infection Service, Public Health England, London, UK
14. WHO Global Collaborating Centre for Reference and Research on Diphtheria and Streptococcal Infections, National Infection Service, Public Health England, London, UK
15. Medical Research Council/Uganda Virus Research Institute (MRC/UVRI) and London School of Hygiene and Tropical Medicine Uganda Research Unit, Entebbe, Uganda
16. Pathogen Immunity Group, Public Health England, Porton Down, UK
17. MRC Centre for Molecular Bacteriology & Infection, Department of Infectious Disease, Imperial College, London, UK

**Key words:** group B streptococcal disease; infant; cluster; healthcare-associated infection; whole genome sequencing

**Running title:** Infant group B streptococcal clusters

**Corresponding author:** Dr Simon Collin, Healthcare-Associated Infection and Antimicrobial Resistance (HCAI & AMR) Division, National Infection Service, Public Health England, 61 Colindale Avenue, London NW9 5EQ, United Kingdom.

simon.collin@phe.gov.uk; Telephone: +44 (0)7796 565002

**Alternate corresponding author:** Dr Theresa Lamagni, Healthcare-Associated Infection and Antimicrobial Resistance (HCAI & AMR) Division, National Infection Service, Public Health England, 61 Colindale Avenue, London NW9 5EQ, United Kingdom.

theresa.lamagni@phe.gov.uk; Telephone: +44 (0)208 3276628

**40-word summary:** We used whole genome sequencing data from invasive group B *Streptococcus* isolates from a 13-month period of national enhanced surveillance in the UK and Ireland to identify 7 clusters, 4 of which had not been detected when they first occurred.

## Abstract

**Background:** The true frequency of hospital outbreaks of invasive group B streptococcal (iGBS; *Streptococcus agalactiae*) disease in infants is unknown. We used whole genome sequencing (WGS) of iGBS isolates collected during a period of enhanced surveillance of infant iGBS disease in the UK and Ireland to determine the number of clustered cases.

**Methods:** Potentially linked iGBS cases from infants with early (<7 days of life) or late-onset (7-89 days) disease were identified from WGS data (HiSeq 2500 platform, Illumina) from clinical sterile site isolates collected between 04/2014 and 04/2015. We assessed time and place of cases to determine a single-nucleotide polymorphism (SNP) difference threshold for clustered cases. Case details were augmented through linkage to national hospital admission data and hospital record review by local microbiologists.

**Results:** Analysis of sequences indicated a cut-off of  $\leq 5$  SNP differences to define iGBS clusters. Among 410 infant iGBS isolates, we identified 7 clusters (4 genetically-identical pairs with 0 SNP differences, one pair with 3 SNP differences, one cluster of 4 cases with  $\leq 1$  SNP differences) of which 4 clusters were uncovered for the first time. The clusters comprised 16 cases, of which 15 were late-onset (of 192 late-onset cases with sequenced isolates) and one an early-onset index case. Serial intervals between cases ranged from 0-59 (median 12) days.

**Conclusions:** Approximately one in 12 late-onset infant iGBS cases were part of a hospital cluster. Over half of clusters were previously undetected, emphasising the importance of routine submission of iGBS isolates to reference laboratories for cluster identification and genomic confirmation.

## Introduction

*Streptococcus agalactiae* (group B *Streptococcus*, GBS) is a leading cause of neonatal sepsis, currently classified as early-onset disease (EOD) if infection occurs during the first 6 days of life or late-onset disease (LOD) if it develops 7-89 days after birth. These classifications correspond to the mode of transmission, with EOD resulting from vertical transmission during or just before birth and LOD often being attributed to horizontal transmission [1, 2]. Our understanding of transmission after birth however, remains very limited, the presumption being that GBS is acquired nosocomially or from a community source.

It has long been known that GBS can spread within hospital nurseries [3]. A recent systematic review of invasive GBS (iGBS) outbreaks in healthcare settings showed that iGBS clusters may occur more frequently than previously thought because intervals between consecutive cases can be long (up to 50 days), with clusters not being detected if reliant on clinicians and microbiologists detecting unusual patterns [4]. The advent of genomic analysis allows us to identify and understand transmission events [5], paving the way for control measures to be designed.

Our aim was to investigate clustering of cases of iGBS disease in infants <90 days old in hospitals using whole genome sequencing data obtained from isolates submitted to Public Health England (PHE) from hospital and reference laboratories across the UK and Republic of Ireland during a period of enhanced national surveillance of infant iGBS disease [6].

## Methods

### Data sources

We used existing whole genome sequence (WGS) data from clinical sterile site isolates (blood, CSF) from infants with invasive GBS (iGBS) disease received by the Public Health England (PHE) Respiratory and Vaccine Preventable Bacteria Reference Unit during a period of enhanced surveillance for cases of early (<7 days of life) and late-onset (7-89 days) infant disease in the UK and Republic of Ireland (01/04/2014 to 30/04/2015) [6].

Identifying data for isolates from infant iGBS cases included patient name, date of birth, date of specimen collection, and name of referring hospital. Additional data (gestational age, birth weight, singleton/multiplet) were obtained from British Paediatric Surveillance Unit (BPSU) paediatrician reports submitted during the period of enhanced surveillance [6]. For clusters in England, patient identifiable data (date of birth, name) from sequenced isolates were sent to NHS Digital's Demographic Batch Service (DBS) to obtain/confirm patient unique patient identifiers (NHS number) and obtain dates of death [7]. NHS numbers were then used to extract corresponding hospital admission records from Hospital Episode Statistics (HES; NHS Digital. Copyright 2018) to create timelines for clusters. HES data provide clinical and organisational information, including diagnoses and procedures, dates of admission and discharge, admission method, and identity of care provider (hospital) [8]. Medical microbiologists at hospitals in England were contacted to determine whether there had been any suspicion of cross-infection between clustered cases at the time they occurred and whether cases had been in the same unit/on the same ward or in adjacent cots. Further information on clustered cases in Scotland, Wales and Republic of Ireland was sought from enhanced surveillance study collaborators in each of these countries. Additional iGBS cases ( $\pm 60$  days of cluster cases) which either did not have isolates available for analysis or were

separated from cluster cases by SNP differences above the cluster threshold were identified by reviewing enhanced surveillance data from hospitals with clusters.

### **SNP difference thresholds to define iGBS clusters**

To determine an appropriate SNP-difference threshold for clustered cases we used existing WGS data from clinical sterile site isolates from patients of all ages received by the PHE Respiratory and Vaccine Preventable Bacteria Reference Unit between 01/2010 and 05/2017. We reviewed distributions of SNP differences within clonal complexes from all iGBS sterile site isolates and initially applied a threshold of  $\leq 10$  SNP differences to identify potential clusters, with further exploration up to 20 SNPs. An initial 10-SNP threshold was chosen as it exceeds the expected mutation rate (1-2 SNPs per genome per year) [9, 10] and the SNP differences ( $\leq 5$ ) in published clusters [5]. Information (specimen date, referring hospital, date of birth, name) accompanying specimens from infants (<90 days old) was examined to assess plausibility of cases being linked and to identify twin and mother-baby (early-onset neonatal) dyads. Combining case-cluster information with SNP difference distributions allowed us to identify a suitably discriminatory threshold with which to define probable clusters for further epidemiological investigation.

### **Whole genome sequencing**

All isolates had been sequenced using a HiSeq 2500 platform (Illumina; San Diego, USA). Sequence types were determined using MOST and clonal complexes were determined from these profiles [11]. Reference genomes used for analysis were based on kmer-based distance estimates from MASH to identify representative complete genomes (see **Supplementary File eTable 1**) [12]. To remove SNPs from suspected regions of recombination, regions of high SNP density were removed from the analyses using GUBBINS [13]. SNP matrices for each clonal complex were produced using PHENix (<https://github.com/phe->

[bioinformatics/PHEnix](#)). Isolate sequences from this study are available via ENA (see **Supplementary File eTable 2** for accession numbers). The age and sex distribution for the isolates is shown in **Supplementary File eTable 3**.

### **Ethics**

All data were collected within statutory approvals granted to the respective public health authorities in the UK and Ireland for infectious disease surveillance and control. Information was held securely and in accordance with the Data Protection Act 2018 and Caldicott guidelines.

### **Role of the funding source**

This study had no specific funding.

## Results

### SNP difference thresholds to define iGBS clusters

Whole genome sequence data were available for 754 unique (non-duplicate) iGBS isolates received by the Reference Unit between 2010 and 2017 including 410 infant iGBS isolates from the 13-month period of enhanced surveillance between April 2014 and April 2015. The distributions of SNP differences across each GBS clonal complex (CC) are shown in **Figure 1** (see also **Supplementary eFigure S1 & eFigure S2**). These show clear separation within each CC between pairs of isolates that are genetically identical (0 SNP differences) or similar ( $\leq 5$  SNPs) and all other pairs of isolates. Just two pairs fell between 5 and 10 SNP differences, each with an 8 SNP difference; these are described below.

In total there were 17 potential clusters of isolates involving an infant with  $<10$  SNP differences between isolates (**Figure 2**). These potential clusters included four mother-baby (early-onset) pairs and 4 twin pairs (two early-onset pairs, two late-onset pairs - see **Supplementary File eTable 4**). The 9 remaining potential clusters comprised two pairs separated by 8 SNPs between cases, one pair with 3 SNP differences, 3 identical pairs (0 SNP difference), and 3 serial clusters (all  $\leq 1$  SNP difference). Raising the initial threshold to  $\leq 20$  SNPs did not reveal any additional clusters involving infants. Of the two pairs with 8 SNP differences: one was a pair of early-onset cases (at 0 and 1 day of life), therefore not plausibly attributable to horizontal transmission (samples were 114 days apart); the other comprised an early-onset case followed 6 months (193 days) later by a late-onset case in the same hospital. No clusters were identified under the initial  $<10$  SNP-difference threshold involving children older than 90 days or adults (other than the mother-infant pairs).

### **Infant iGBS clusters**

Using a cut-off of  $\leq 5$  SNP differences to define an iGBS cluster, 7 clusters (16 cases in total) were identified in the UK and Ireland during the enhanced surveillance period (**Table 1**). Six of the 7 clusters involved pairs of cases, with the remaining cluster comprising 4 cases. Serial intervals ranged from 0-25 days for pairs with  $\leq 1$  SNP difference; the cases in the pair with 3 SNP differences occurred 59 days apart. Of the 7 clusters, 4 had not been identified by the hospitals as potential clusters. The other 3 were known serial clusters that occurred in a NICU comprising: an initial cluster of 4 serotype V late-onset cases over a 4-week period; a second cluster, 9 months later, of two serotype III late-onset cases 13 days apart; a third cluster, 10 months after the first cluster, comprising twins with late-onset disease (12 days apart) preceded 9 days earlier by an infant with a positive rectal swab which differed from the twin cases by 5 SNPs. Our analysis replicated exactly the previously-reported SNP differences [5].

### **Previously undetected clusters**

One cluster was identified in Ireland, which comprising a pair of pre-term LOD cases 9 days apart with 0 SNP differences and linked to the same hospital (**Table 1**). No further timeline information was available for this cluster. The remaining clusters occurred in hospitals in England as described below, with timelines shown in **Figure 3**.

#### *Cluster A*

This cluster (serotype III, CC17, ST17, 3 SNP differences) comprised two late-onset cases occurring 59 days apart. Both infants were born at full term in the same hospital (A) 66 days apart and both were discharged home 2 days after birth. Case #1 was admitted to hospital A via A&E at 19 days of life, and GBS was isolated from a blood sample taken on the day of admission; the infant was discharged home 16 days later. Forty-three days after case #1 was discharged, case #2 was admitted for 'septic shock' to a different (but geographically close)

hospital (B) *via* A&E at 12 days of life. Medical records at the hospital (A) did not indicate any suspicion of a link between the cases at the time they occurred. Case #1 had been re-admitted to hospital A *via* A&E 30 days later for ‘gastro-oesophageal reflux disease without oesophagitis’, staying for 2 days before being discharged home. The second day of this admission coincided with the birth of case #2 in hospital A, but case #1’s re-admission was to a general paediatric ward one floor above the Delivery Suite where case #2 was born. The post-natal and general paediatric wards were usually staffed by different teams and no evidence was found to suggest that case #2 was transferred from the post-natal ward to the general paediatric ward. No other cases reported during the enhanced surveillance period could be linked to the cases in cluster A based on time ( $\pm 60$  days from either case) or SNP threshold; the closest case in time was a late-onset case 95 days after case #2 with 115 SNP differences.

#### *Cluster B*

This cluster (serotype III, CC17, ST17, 0 SNP difference) comprised an early-onset case (Case #1) followed 12 days later by a late-onset case (Case #2), both born at the same hospital. Case #2 was born preterm (at 29 weeks) 24 days before the birth of case #1 and was still hospitalised at the time of case #1’s birth. Case #1 (full term) had GBS at 0 days of life and was discharged home after 9 days. Case #2 developed late-onset GBS at 36 days of life whilst still in hospital, was discharged home on day 57, but readmitted 2 days later *via* A&E with ‘streptococcal meningitis’ (GBS was isolated from a second blood sample taken on the day of re-admission, no WGS data available), leading to a further hospital stay of 15 days’ duration before being discharged home. Bed management system records at the hospital showed that there was a period of 4 days when both cases were in the same bay on the special care baby unit. No other cases reported during the enhanced surveillance period could be linked to the cases in cluster B; the closest cases in time occurred 58 days and 70 days after

case #2, the first a late-onset case with no isolate, the second a late-onset case with 373 SNP differences.

### *Cluster C*

This cluster (serotype Ia, CC23, ST23, 0 SNP difference) comprised two late-onset cases occurring 25 days apart born at the same hospital. Case #1 was preterm (29 weeks) and hospitalised from birth (for 73 days), with GBS occurring at 26 days of life. Case #2 was one of two preterm twins (30 weeks) and was born during the hospital stay of case #1; late-onset GBS developed at 17 days of life. Hospital records showed that both cases were nursed on the same ward. No late-onset cases with the same serotype were reported within 6 months of the two cases in cluster C.

## Discussion

This is the first population-wide study to make a systematic assessment of infant iGBS clusters in hospitals and to determine a genomic threshold for identifying such clusters. We found over half of genomic clusters were previously undetected. All unreported clusters were epidemiologically plausible. Of the 16 cases involved in these clusters, 3 died, including 1 secondary case. Whilst we cannot determine the preventability of these infections, our findings should direct further public health research to identify opportunities for prevention.

To date, 30 hospital GBS clusters have been documented across 11 countries worldwide [4]. Among them were 4 clusters which occurred in a neonatal unit in England, 3 of which fell within the period of enhanced surveillance in the UK and Ireland. These 3 clusters were independently identified in our study through our assessment of genomic variation, validating the methodology (SNP thresholds) developed during this study [5].

Our study confirms that links between infant GBS cases will escape detection without use of genomic analysis. We have also provided further evidence that serial intervals between linked cases can be quite long ( $\geq 2$  weeks), a factor which potentially hinders detection of infant GBS clusters in healthcare settings and therefore needs to be brought to the attention of healthcare professionals [4]. We also note that the second case in our cluster A was admitted with iGBS to a different hospital, an additional obstacle to detection. Taken together, the implication of these factors is that our understanding of late-onset infant GBS attributable to nosocomial transmission remains very limited and in need of prospective genomic studies.

The two previously unreported clusters in our study where the second case was born during a prolonged period of hospitalisation of a preterm index case (clusters B and C) comprised case pairs with genetically identical isolates, and serial intervals of 12 and 25 days. In the other

previously unreported cluster (A), the two cases were term infants who were discharged home prior to late-onset of GBS, and the 3-SNP difference between the isolates is consistent with the longer interval between the two cases (59 days). In this cluster, horizontal transmission could have occurred through environmental contamination which persisted from the time of birth of the index case or when the index case was re-admitted for gastro-oesophageal reflux. This readmission coincided with the birth of the second case, whose late-onset GBS developed 12 days later. Horizontal transmission could have occurred if the index case was still colonised with the same GBS strain (data not available) and if there was a transmission route via staff and/or equipment; a third possibility is that both cases arose from persistent carriage of GBS by a healthcare worker. We did not have sufficient information to determine whether the re-admitted infant stayed in proximity to the subsequent case or was cared for by the same personnel.

The next most closely related cases in our study occurred in two clusters where the case pairs within each cluster differed by 8 SNPs. The epidemiological plausibility of these two clusters was very weak: one comprised two early-onset cases 114 days apart; the other two late-onset cases 193 days apart. We did not find any clusters with 4-7 or 9 SNP differences to investigate epidemiological plausibility, but our results suggest that  $\leq 5$  SNP differences is a reasonable threshold for potentially linking infant GBS cases.

The main strength of our study is that it used data and isolates collected during a period of prospective, active national surveillance of invasive GBS disease in infants [6]. During this time, the British Paediatric Surveillance Unit (BPSU) facilitated reporting of iGBS cases by paediatricians, whilst PHE triangulated case data with routine microbiology and reference laboratory data. The main limitations of our study were that isolates were submitted for only half the number of reported cases (410/856), therefore some clusters may have escaped detection. Also, whilst the BPSU-supported study used multiple sources to maximise case

ascertainment, this does not guarantee that all cases were reported, and infants with probable (culture-negative) invasive GBS were not included.

Epidemiological investigation in our study was limited to contacting consultant microbiologists to obtain information retrospectively. Detailed epidemiological data might have allowed us to identify mechanisms of transmission or sources of infection. However, we note that two of the three clusters reported previously by Jauneikaite *et al.* were identified as part of a prospective within-hospital study which indicated deficient breast pump hygiene as a possible route of infection but was otherwise unable to identify mechanisms or sources of cross-infection. This difficulty was reflected in our systematic review of iGBS outbreaks in healthcare settings, in which half of the 17 included studies suggested lapses in infection control, including inadequate disinfection of equipment and surfaces and close proximity of cots compounded by factors such as crowding and high patient-to-nurse ratio, but none were unable to provide microbiological evidence of sources or routes of cross-infection or whether outbreaks were propagated rather than being spread from a point source [4].

Our linked data analysis only allowed us to link to the infant's birth record within the hospital episode database (mother's and baby's records are not routinely linked for data protection reasons), and the enhanced surveillance system did not record maternal data. We were therefore unable to enrich the reported data with information about the mother of each case, including maternal GBS status and receipt of intrapartum antibiotic prophylaxis.

During the period of enhanced surveillance covered by our study the predominant serotypes were III (60%) and Ia (17%) [6]. Five serotypes (Ia, Ib, II, III, V) accounted for 94% of all serotyped isolates and most serotype III (74%) and Ia isolates (74%) belonged to sequence types ST17 and ST23, respectively. Our results reflect these common serotypes and, although we cannot be certain that our analysis did not miss iGBS clusters of other serotypes, these would need to have markedly different SNP-difference distributions

Of the 13 late-onset cases that we identified in England (8 in 3 clusters reported by Jauneikaite *et al.* plus 5 in 3 previously unreported clusters), 7 are plausibly attributable to horizontal transmission from a preceding early or late-onset case and 1 (the first case in the twin case pair reported by Jauneikaite *et al.*) from a colonised infant. Isolates from these 8 cases were among 130 sequenced isolates with known time of disease onset collected during the period of enhanced surveillance. Therefore, we can estimate that as many as 8/130 (6%) of late-onset cases arose through nosocomial transmission. This denominator includes cases admitted from home – the proportion of cases occurring within an intensive or special care unit that are attributable to nosocomial transmission will probably be higher.

Given the global burden and long-term sequelae of infant iGBS disease and the absence of measures to prevent late-onset disease [14-16], further research is warranted to better understand transmission routes for late-onset iGBS to inform strategies for prevention, including the likely clinical and cost-effectiveness of GBS vaccination of mothers and the wider population [17-19]. Consideration needs to be given to whether additional infection control precautions are warranted, although evidence suggests that cross-infection occurs when existing measures fail [4]. In the meantime, clinicians should maintain a high index of suspicion for potential GBS clusters and ensure routine submission of isolates to a reference laboratory to facilitate future investigation and confirmation through genomic analysis.

## **Declarations**

**Funding:** This study received no specific funding.

**Conflict of interest:** None of the authors have any conflicting interests to declare.

**Acknowledgements:** We would like to thank the clinical microbiologists who provided further information about the previously unreported clusters. EJ is a Rosetrees/Stoneygate 2017 Imperial College Research Fellow (number M683). EJ, VC, AE, SS, and TL acknowledge the support of the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare-Associated Infections and Antimicrobial Resistance at Imperial College in partnership with Public Health England (PHE), in collaboration with The Sanger Institute, the University of Cambridge Veterinary School and Imperial College Health Partners. SS is based at Imperial College and acknowledges the NIHR Imperial Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

## Figure legends

Figure 1: Distributions of SNP differences between any pair of isolates below 51 SNPs across each invasive group B streptococcal (iGBS) clonal complex in isolates received by the Public Health England (PHE) Respiratory and Vaccine Preventable Bacteria Reference Unit, 2010-2017 (N=754)

Figure 2: Flowchart showing data sources and linkage used to determine SNP difference thresholds for invasive group B streptococcal (iGBS) clusters and to identify iGBS clusters during a period of enhanced surveillance for cases of early (<7 days of life) onset disease (EOD) and late-onset (7-89 days) disease (LOD) in the UK and Ireland, 2014/2015

Figure 3: Time lines for previously undetected infant invasive group B streptococcal (iGBS) clusters showing dates of birth, iGBS onset and hospital admission/discharge

## References

1. Le Doare K, Heath PT. An overview of global GBS epidemiology. *Vaccine* **2013**; 31 Suppl 4: D7-12.
2. Nanduri SA, Petit S, Smelser C, et al. Epidemiology of Invasive Early-Onset and Late-Onset Group B Streptococcal Disease in the United States, 2006 to 2015: Multistate Laboratory and Population-Based Surveillance. *JAMA Pediatr* **2019**; 173(3): 224-33.
3. Aber RC, Allen N, Howell JT, Wilkenson HW, Facklam RR. Nosocomial transmission of group B streptococci. *Pediatrics* **1976**; 58(3): 346-53.
4. Collin SM, Lamb P, Jauneikaite E, et al. Hospital clusters of invasive Group B Streptococcal disease: A systematic review. *J Infect* **2019**; 79(6): 521-7.
5. Jauneikaite E, Kapatai G, Davies F, et al. Serial Clustering of Late-Onset Group B Streptococcal Infections in the Neonatal Unit: A Genomic Re-evaluation of Causality. *Clin Infect Dis* **2018**; 67(6): 854-60.
6. O'Sullivan CP, Lamagni T, Patel D, et al. Group B streptococcal disease in UK and Irish infants younger than 90 days, 2014-15: a prospective surveillance study. *Lancet Infect Dis* **2019**; 19(1): 83-90.
7. NHS Digital. Demographics Batch Service Bureau (DBSB): NHS Number Batch Tracing. London, UK: Health and Social Care Information Centre, **2017**.
8. NHS Digital. Hospital Episode Statistics (HES). Available at: [digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics](https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics). Accessed 13/02/2020.
9. Da Cunha V, Davies MR, Douarre PE, et al. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nat Commun* **2014**; 5(1): 4544.
10. Jamrozy D, de Goffau MC, Bijlsma MW, et al. Temporal population structure of invasive Group B *Streptococcus* during a period of rising disease incidence shows expansion of a CC17 clone. *bioRxiv* **2018**: 447037.
11. Tewolde R, Dallman T, Schaefer U, et al. MOST: a modified MLST typing tool based on short read sequencing. *PeerJ* **2016**; 4: e2308.
12. Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **2016**; 17(1): 132.
13. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **2014**; 43(3): e15-e.
14. Kohli-Lynch M, Russell NJ, Seale AC, et al. Neurodevelopmental Impairment in Children After Group B Streptococcal Disease Worldwide: Systematic Review and Meta-analyses. *Clin Infect Dis* **2017**; 65(suppl\_2): S190-S9.
15. Tann CJ, Martinello KA, Sadoo S, et al. Neonatal Encephalopathy With Group B Streptococcal Disease Worldwide: Systematic Review, Investigator Group Datasets, and Meta-analysis. *Clin Infect Dis* **2017**; 65(suppl\_2): S173-S89.

16. Mukhopadhyay S, Puopolo KM. Preventing Neonatal Group B *Streptococcus* Disease: The Limits of Success. *JAMA Pediatr* **2019**; 173(3): 219-20.
17. Edwards MS, Baker CJ. Group B Streptococcal Disease: Interim Prevention at 50 Years and Counting. *Clin Infect Dis* **2019**; (ciz738).
18. Dangor Z, Khan M, Kwatra G, et al. The Association Between Breast Milk Group B Streptococcal Capsular Antibody Levels and Late-onset Disease in Young Infants. *Clin Infect Dis* **2020**; 70(6): 1110-4.
19. Barshak MB. Group B *Streptococcus*, an A-List Pathogen in Nonpregnant Adults. *JAMA Intern Med* **2019**; 179(4): 488-9.

**Table 1: Characteristics of cases in infant invasive group B streptococcal (iGBS) disease clusters in UK and Ireland identified by genomic analysis from isolates collected during a period of enhanced surveillance (04/2014 – 04/2015)**

Country	Serotype (sequence type, clonal complex)	Case (1= index)	SNP difference	Gestational age (weeks)	Singleton or multiplet	Age at onset of iGBS (days)	Serial interval (days)
Ireland	Ib (ST8, CC10)	1	-	25	singleton	22	-
		2	0 SNPs	24	singleton	74	9
England (cluster A)	III (ST17, CC17)	1	-	40	singleton	19	-
		2	3 SNPs	40	singleton	12	59
England (cluster B)	III (ST17, CC17)	1	-	40	singleton	0	-
		2	0 SNPs	29	singleton	36	12
England (cluster C)	Ia (ST23, CC23)	1	-	29	singleton	26	-
		2	0 SNPs	30	twin	17	25
England <sup>a</sup>	V (ST1, CC1)	1	-	27	twin	12	-
		2	≤1 SNP	24	singleton	9	11
		3	≤1 SNP	26	singleton	44	17
		4	≤1 SNP	29	singleton	51	0
England <sup>a</sup>	III (ST17, CC17)	1	-	39	singleton	39	-
		2	0 SNPs	32	singleton	32	13
England <sup>a</sup>	Ib (ST139, CC1)	1	-	25	twins	46	-
		2	0 SNPs	25		58	12

<sup>a</sup> Previously reported in Jauneikaite *et al.* [5]