Supplementary material

Recurrent Invasive Pneumococcal Disease in Children: a retrospective cohort study, England, 2006/07-2017/18

Marta Bertran MSc[a], Fariyo Abdullahi MSc[a], Joshua C. D'Aeth PhD[b], Zahin Amin-Chowdhury MSc[a], Nick J Andrews PhD[a], Seyi Eletu PhD[b], David Litt PhD[b], Prof Mary E Ramsay FFPH[a], Godwin Oligbu PhD[a,c], Prof Shamez N Ladhani PhD[a,c]

a) Immunisation and Vaccine Preventable Diseases Division, UK Health Security Agency, 61 Colindale Avenue, London, NW9 5EQ, United Kingdom
b) Respiratory Vaccine-Preventable Bacteria Reference Unit (RVPBRU), UK Health Security Agency, 61 Colindale Avenue, London, NW9 5EQ, United Kingdom
c) Centre for Neonatal and Paediatric Infections, St George's University of London, Cranmer Terrace, London SW17 0RE, United Kingdom

Methods for genomic analysis in recurrent cases due to first serotype in the first and second IPD episode

In the UK, whole-genome sequencing was not routinely conducted for IPD samples during the study period, but stored samples were analysed on an ad-hoc basis for this study. For recurrent cases, we identified cases infected with the same serotype in the first and second episodes and prepared these samples for whole genome sequence analysis. The aim was to understand whether two consequent IPD episodes caused by the same serotype could be an ongoing infection.

Genomic DNA was extracted and sequenced as described in Kapatai et al 2016 (15). Briefly, we extracted DNA from pure *S. pneumoniae* cultures using the QIAGEN QIAsymphony DSP DNA Mini kit and sequenced the extracted DNA on Illumina HiSeq 2500 platforms. The deplexed fastq files produced were then trimmed using the Trimmomatic tool (16). Reads were then assembled using shovill v0.9.0 (17).

We determined the multi-locus sequence type (MLST) of these isolates using the MOST tool (18), taking the *S. pneumoniae* typing datasets from the species-specific PubMLST page (19). We also used PubMLST to determine the core-genome MLST (cgMLST). We used PopPUNK v2.6.0 (20) to assign Global pneumococcal sequencing clusters (GPSCs) to isolates with the database from the Global Pneumococcal Sequencing Project (21). We used MLST, cgMLST and GPSC profiles, along with the distance between isolates calculated through Mash v2.3 (22), to determine whether isolates from two subsequent episodes in a recurrent IPD case were likely to be caused by the same or different strains.

Enhanced surveillance and questionnaire return

Enhance surveillance changed throughout this period, with questionnaires initially sent only to children <5 years, and then extended to all ages. Single-episode cases were more likely to have a returned questionnaire than recurrent episode cases [75.9% (3,740/4,928) vs 67.4% (155/230), respectively; p=0.003] (Supp. Table 1), but this difference disappeared once surveillance was enhanced for all ages in 2010, when questionnaire return rate was 94.5% (1,617/1,712) and >90% for both groups (p=0.282) (Supp. Table 1). Furthermore, there was weak evidence for a lower return rate among recurrent cases (67.6%) compared to single-episode cases (75.9%, p=0.0504) when only comparing first episodes (Table 1).

Table S1: Number and proportion of questionnaires returned by recurrence group for all IPD episodes.

Data is reported for the whole time period and broken down by the periods in which surveillance was enhanced only for under 5-year-olds (<2010/11) and for all ages.

Questionnaire returned	Single-episode cases	Recurrent cases	Total	p-value
Overall	4,928 (100%)	230 (100%)	5,158 (100%)	0.003*
Yes	3,740 (75.9%)	155 (67.4%)	3,895 (75.5%)	
No	1,188 (24.1%)	75 (32.6%)	1,263 (24.5%)	
Surveillance enhanced only for <5s	3,302 (100%)	144 (100%)	3,446 (100%)	<0.0001*
Yes	2,202 (66.7%)	76 (52.8%)	2,278 (66.1%)	
No	1,100 (33.3%)	68 (47.2%)	1,168 (33.9%)	
Surveillance enhanced for all ages	1,626 (100%)	86 (100%)	1,712 (100%)	0.282¥
Yes	1,538 (94.6%)	79 (91.9%)	1,617 (94.5%)	
No	88 (5.4%)	7 (8.1%)	95 (5.5%)	
* chi-square; ¥ fisher exact test				

Table S2: Number and percentage of cases with underlying conditions by age group in single-episode cases and first episode of recurrent IPD cases

	No (%) single cases N=4928	No (%) recurrent cases N=105	Total (n %) N=5,033	P-value‡
<5-years (2006/07-2017/2018)	3,378	70	3,448	
Healthy	2440 (72.2%)	15 (21.4%)	2455 (71.2%)	<0.001
Any comorbidity:	938 (27.8%)	55 (78.6%)	993 (28.8%)	<0.001
Immunosupression	220 (23.5%)	33 (60%)	253 (25.5%)	
Other comorbidities	718 (76.5%)	22 (40%)	740 (74.5%)	
≥5-years (2006/07-2017/2018)	379	8	387	
Healthy	159 (42.0%)	1 (12.5%)	160 (41.3%)	0.147*
Any comorbidity:	220 (58.0%)	7 (87.5%)	227 (58.7%)	
Immunosupression	74 (33.6%)	5 (71.4%)	79 (34.8%)	0.052
Other comorbidities	146 (66.4%)	2 (28.6%)	148 (65.2%)	
≥5-years* (2010/11 -2017/2018)	320	8	328	
Healthy	153 (47.8%)	1 (12.5%)	154 (47.0%)	0.071*
Any comorbidity:	167 (52.2%)	7 (87.5%)	174 (53.0%)	
Immunosupression	73 (43.7%)	5 (71.4%)	78 (44.8%)	0.245
Other comorbidities	94 (56.3%)	2 (28.6%)	96 (55.2%)	
‡ P-value obtained from chi-square (*) or fisher exact test (*) or Mann-Whitney test (†) *Period corresponding to enhanced surveillance for this age group				

Table S3: Prevalence of underlying conditions in single-episode and recurrent cases among children with known comorbidities

Underlying condition group	No (%) single cases N=1158	No (%) recurrent cases (N=62)	Total (n, %) N=1220	P-value‡
Congenital abnormalities	368 (31,8%)	19 (30,6%)	387 (31,7%)	0,852
Immunodeficiency	294 (25,4%)	38 (61,3%)	332 (27,2%)	<0,001
Respiratory disease	217 (18,7%)	9 (14,5%)	226 (18,5%)	0,503*
Chronic cardiac disease	182 (15,7%)	8 (12,9%)	190 (15,6%)	0,719*
Chronic liver disease	47 (4,1%)	6 (9,7%)	53 (4,3%)	0,047*
Chronic renal disease	76 (6,6%)	8 (12,9%)	84 (6,9%)	0,068*
Asplenia	62 (5,4%)	7 (11,3%)	69 (5,7%)	0,080*
Sickle cell disease or heamoglobinopathy	61 (5,3%)	3 (4,8%)	64 (5,2%)	1,000*
CSF leak	39 (3,4%)	6 (9,7%)	45 (3,7%)	0,023*
Diabetes	14 (1,2%)	1 (1,6%)	15 (1,2%)	0,545*
Cochlear implants	12 (1%)	2 (3,2%)	14 (1,1%)	0,156*
Coeliac disease	6 (0,5%)	2 (3,2%)	8 (0,7%)	0,058*
‡ P-value obtained from chi-square (no symbol) or fisher exact test (*) or Mann-Whitney test (†)				

Table S4: Clinical presentations in first episode of recurrent cases with same and different presentation in the first and second episode

Clinical presentation	Different presentation (n=17)	Same presentation (n=33)	p-value
Meningitis	1 (5,9%)	16 (94,1%)	0,007
Pneumonia	7 (46,7%)	8 (53,3%)	
Bacteremia	9 (50%)	9 (50%)	
Median interval (IQR, days)	336 (IQR 104-628)	124 (IQR 60-271)	0,232

Figure S1. Number of single-episode and recurrent cases for serotypes isolated in children with the same serotype in the first and second episode.

Shaded area represents the proportion of serotype-specific recurrent episodes compared to total episodes for the individual serotypes.