THE LANCET
Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Bertran M, D'Aeth JC, Abdullahi F, et al. Invasive pneumococcal disease 3 years after introduction of a reduced 1 + 1 infant 13-valent pneumococcal conjugate vaccine immunisation schedule in England: a prospective national observational surveillance study. *Lancet Infect Dis* 2024; published online Feb 1. https://doi.org/10.1016/S1473-3099(23)00706-5.

Supplementary files

Invasive pneumococcal disease three years after introduction of a reduced 1+1 infant 13-valent pneumococcal conjugate vaccine (PCV13) immunisation schedule in England: prospective national observational surveillance, 2017/18-2022/23

Marta Bertran MSc[1], Joshua C D'Aeth PhD[2], Fariyo Abdullahi MSc[1], Seyi Eletu PhD[2], Nick J Andrews PhD[1], Prof Mary E Ramsay FFPH [1,3], David J Litt PhD[1,2], Prof Shamez N Ladhani PhD [1,4]

1 Immunisation and Vaccine Preventable Diseases Division, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, United Kingdom
2 Respiratory and Vaccine Preventable Bacteria Reference Unit, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, United Kingdom
3 Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

4 Paediatric Infectious Diseases Research Group, St. George's University of London, Cranmer Terrace, London SW17 0RE, United Kingdom

Corresponding author: Shamez Ladhani; shamez.ladhani@ukhsa.gov.uk

Immunisation and Vaccine Preventable Diseases Division, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, United Kingdom

S1 Methods: Additional Laboratory methods, whole Genomic Sequencing, assembly and MLST typing methods and genomic analyses

Genomic DNA was extracted from pure *S. pnuemoniae* cultures following overnight growth on Columbia blood agar plates. Extraction was performed using the Qiagen QIAsymphonySP platform and the QIAsymphony DSP DNA Mini Kit. Extracted DNA was then sequenced by the UKHSA's Central Sequencing laboratory using the Illumina HiSeq2500 and NextSeq1000 platforms.

Reads were de-multiplexed and trimmed of regions with a PHRED score below 30. The sequence reads were used to infer species with KmerID, the multi-locus sequence type (MLST) of isolates with MOST[1] and the publicly available MLST profiles from PubMLST[2], and the serotype of isolates with PneumoCAT[3]. If the results of PneumoCAT were inconclusive for an isolate, phenotypic serotyping with slide agglutination was used. Reads were assembled using Shovill v0.9.0 [https://github.com/tseemann/shovill], with contigs <500bp in length removed and quality control (QC) performed using QUAST[4] and CheckM[5].

For serotype 3, we further investigated GPSC12 isolates by assessing changes in different Clades, as described by Kwun *et al* 2022[6]. To assign GPSC12 isolates to these Clades, a phylogeny of the 2528 GPSC12 isolates in this study and the 890 GPSC12 isolates in Kwun *et al* 2022 was inferred. The assemblies of the 890 GPSC12 isolates were provided by the authors.[6] There were 92 isolates present in both collections, of which only the Kwun *et al* 2022 assemblies were used, giving a total dataset of 3,326 isolates. Isolates were mapped to the complete reference genome of the *S. pneumoniae* OXC141 isolate (accession code FQ312027) using SKA2 v0.3.1 [https://github.com/bacpop/ska.rust] Gubbins (v3.3.0) was then run with a RapidNJ[7] starting phylogeny and RAxML (v8.2.12)[8] to infer a phylogeny from the non-recombining regions of the genome. The phylogeny was visually inspected in microreact[9] [https://microreact.org/project/ihJawx3YtqS14bhcGmrFCp-ukhsa-gpsc12-ska2-2528] and used to assign the Clade nomenclature to the 2528 UKHSA GPSC12 isolates. The phylogeny was also visualised using iTOL[10].

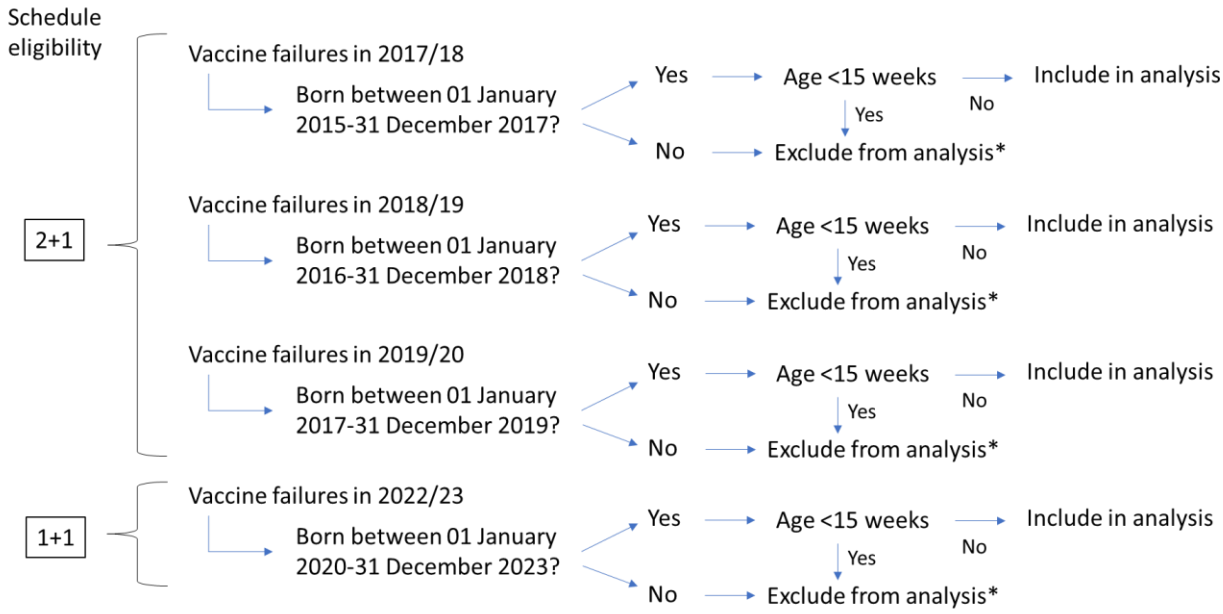
For creation of the serotype 4 GPSC162 phylogeny, isolates were first mapped to the reference ST801 sequence used in Gladstone *et al* 2022 (GCA_90129732) [11]. This was prepared in the same way, namely contigs of < 500 bp in length were removed, and the remaining contigs were reordered with ABACAS v1.3.1[12] against the completed reference genome (ATCC700669), to give a final length of 2,051,095 bp. The mapping was performed with SKA2 v0.3.1 [GitHub - bacpop/ska.rust: Split k-mer analysis – version 2], with a kmer size of 31. The resultant alignment was then input into Gubbins v3.3.0[13], run with a RapidNJ[7] starting phylogeny and a RAxML v8.2.12[8], to produce a phylogeny from the non-recombining regions of the genome. This phylogeny was then visualised with iTOL[10] and uploaded to Microreact for interactive visualisation: https://microreact.org/project/qgAHix5srw1sEYXze3ArAa-gpsc162-ukhsa [9].

Routine WGS was implemented for all samples in October 2017, with some samples between July and October 2017 also processed through WGS. Overall, WGS results were available for 22,185/24195 (91.7%) of confirmed and serotyped cases, and 85.9% (22,185/25829) of all cases.

S2 Vaccine failures and breakthrough infections rates:

Flowchart representing the inclusion/exclusion criteria for vaccine failures and breakthrough infections for comparability between the 1+1 and 2+1 schedules

For assessment of vaccine failures in 2022/23 in children eligible for the 1+1 schedule, cases had to be born between 1 January 2020 and 31 December 2022. To ensure comparability, we used the same interval to select birth cohorts for children eligible for the 2+1 schedule in the pre-pandemic years: 1 January 2017 to 31 December 2019 for cases occurring in 2019/20, 1 January 2016 to 31 December 2018 for 2017/18 cases and 1 January 2015 to 31 December 2017 for 2017/18 cases. Breakthrough infections in children aged <15 weeks were excluded to ensure comparability of schedule.



Note (*): To calculate the rates, counts are calculated individually for each year and eligible birth cohort. Therefore, a failure/breakthrough infection may be excluded from the 2017/18 count but included in the 2018/19 count.

Figure S3. Percentage of IPD cases by financial year and region, England, All ages

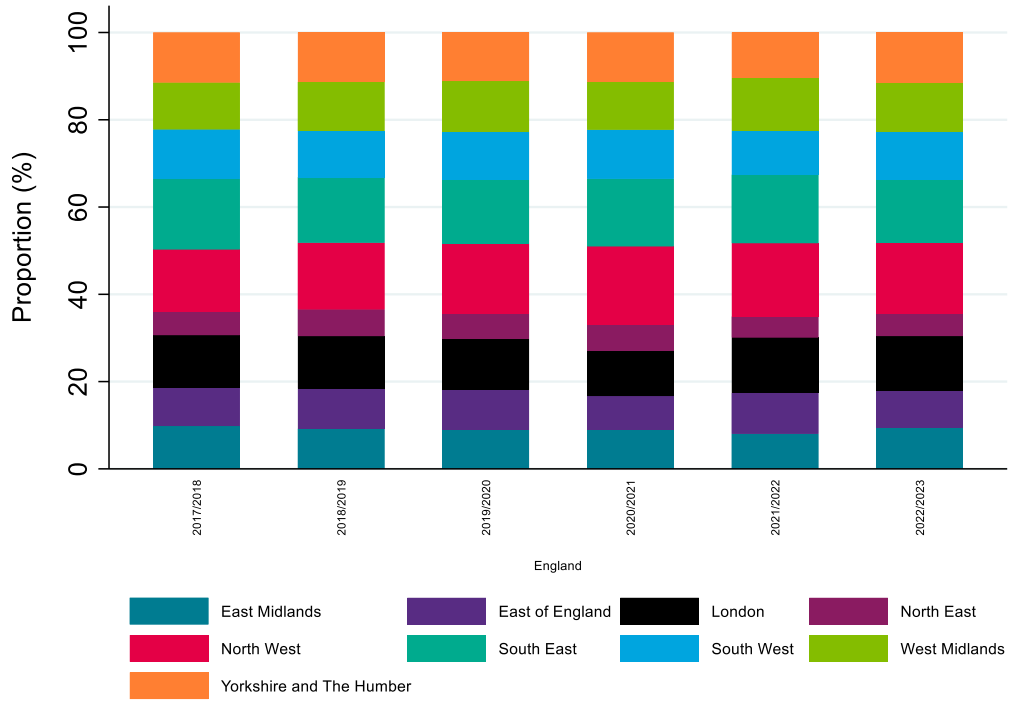
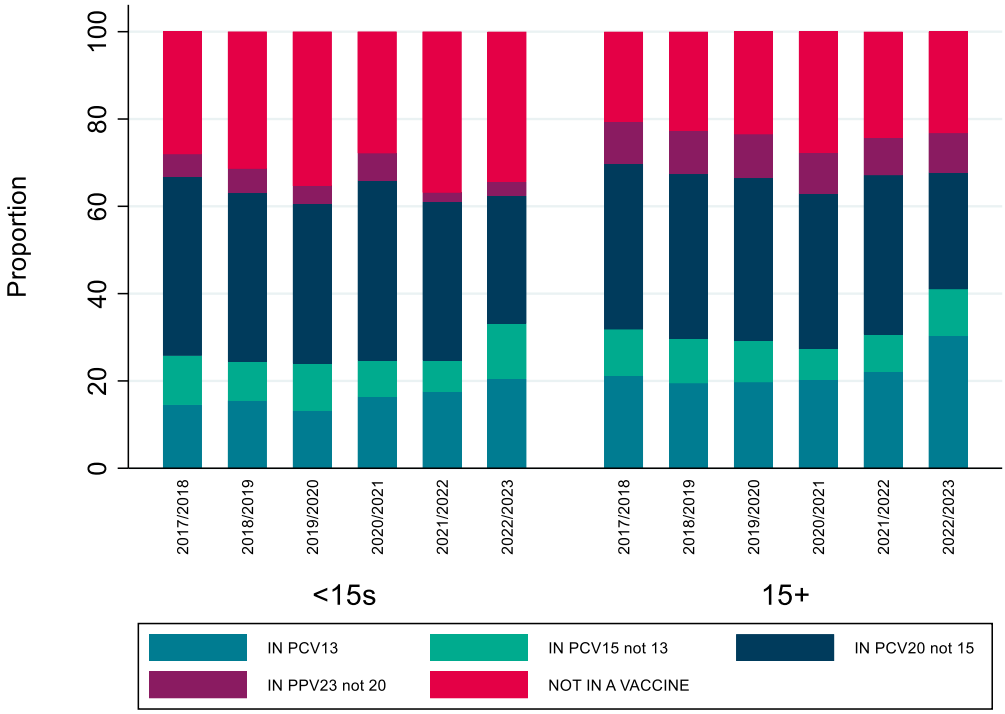


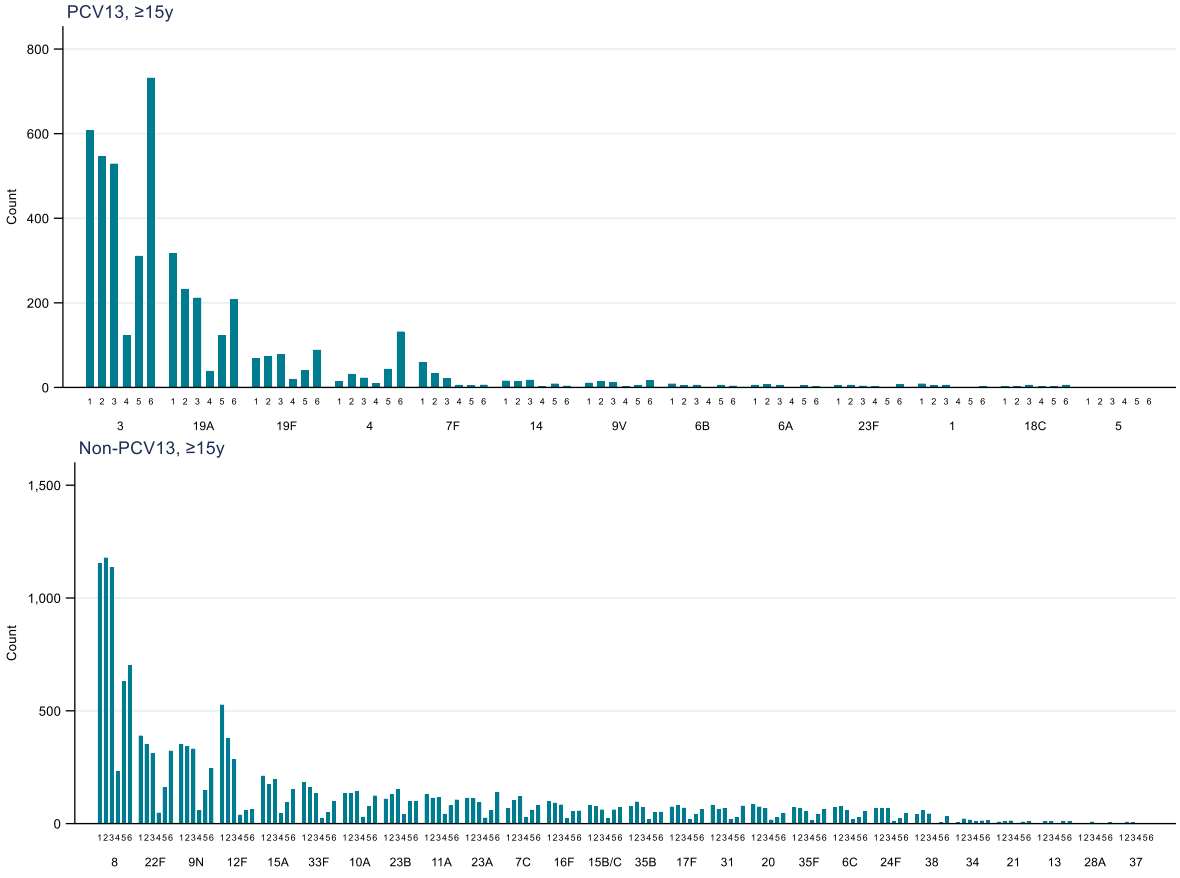
Figure S4. Percentage of IPD cases by vaccine serotype group and financial year, broken down by age group, England



Cases by vaccine serotype group and age group

Pre-pandemic (2017/18-2019/20), four of the five most prevalent serotypes causing IPD were non-PCV13 serotypes: 8 (22%), 12F (8%), 22F (7%) and 9N (7%), in addition to PCV13 serotype 3 (11%). In 2022/23, serotype 3 became the most prevalent (18%) followed by serotype 8 (17%), 22F (8%) and 9N (6%), while serotype 12F was responsible for <2% of cases (19 in ranking) and 19A becoming the fifth (from sixth) most prevalent serotype (5%). The same pattern was observed in adults (figure S5a).

Figure S5a. IPD cases due to serotypes included in PCV13 and not included in PCV13 (non-PCV13) in adults ≥15-years. Note: For non-PCV13 serotypes, bars have been supressed if the total count for the serotype was <20 (adults). To facilitate readability, bar labels 1-6 are ordered by financial year from 2017/18 (1) to 2022/23 (6). Axis on both graphs differ and the order of the serotypes do not align between different graphs.



In children, the five main serotypes responsible for pre-pandemic IPD (2017/18-2019/20) were 12F (10%), 10A (9 %), 8 (9%), 23B (9%), 15B/C (8%). Among PCV13 serotypes, serotype 3 (59/898, 6.6%; sixth), 19A (42/898, 4.7%; eighth) and 19F (24/898, 2.7%; eleventh) predominated in children in the pre-pandemic years. In 2022/2023, there were only two 12F cases (0.6%), and the top five ranking changed to 10A (n=40/341; 12.0%), 23B (n=33; 10.0%, 3 (n=28; 8.0%), 15B/C (n=27, 8.0%) and 22F (n=25, 7.0%). The proportion of 19A (n=24, 7.0%, sixth) and 19F (n=15; 4.4% eleventh) also increased (figure S5b).

Figure S5b. IPD cases due to serotypes included in PCV13 and not included in PCV13 (non-PCV13) in <15-year-olds. Note: For non-PCV13 serotypes, bars have been supressed if the total count for the serotype was <5 (children). To facilitate readability, bar labels 1-6 are ordered by financial year from 2017/18 (1) to 2022/23 (6). Axis on both graphs differ and the order of the serotypes do not align between different graphs.

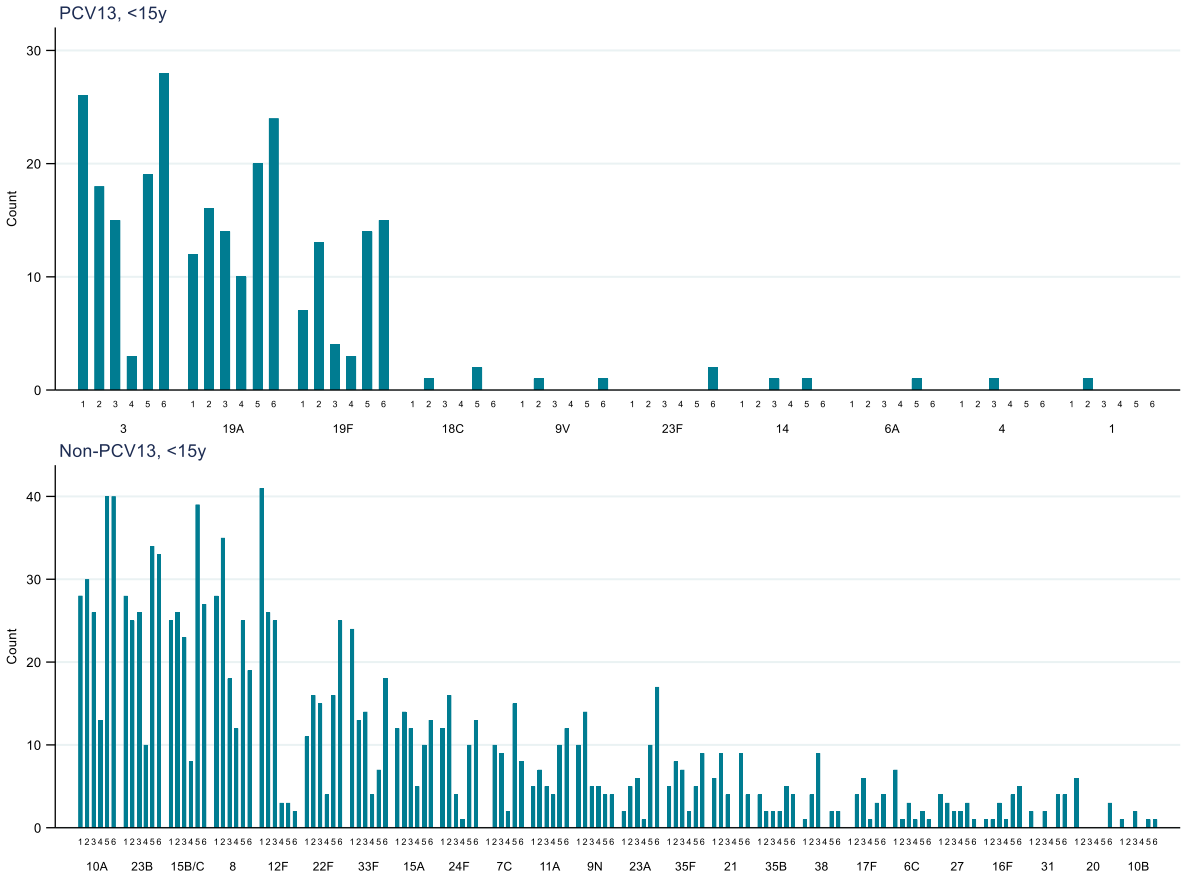


Table S6 Raw and adjusted number of IPD cases, incidence per 100,000 and incidence ratio comparing PCV13-IPD and non-PCV13-IPD in 2022/23 to 2019/20 by age group assuming serotype 3 is a non-vaccine type

	2019/20 n (adjusted n*)	2019/20 Incidence per 100,000	2022/23 n (adjusted n*)	2022/23 Incidence per 100,000	Incidence rate ratio 2022/23 vs 2019/20 (95%CI)**	p-value
PCV-13 IPD (minus serotype 3)						
<1 years	6 (7)	1.07	17 (22)	3.25	3.03 (0.79-11.66)	0.11
1to4 years	11 (11)	0.44	14 (16)	0.61	1.38 (0.44-4.32)	0.58
5to14 years	3 (3)	0.05	11 (11)	0.19	3.74 (0.59-23.77)	0.16
15to44 years	57 (62)	0.29	105 (110)	0.51	1.81 (1.14-2.89)	0.010
45to64 years	117 (114)	0.87	158 (149)	1.14	1.33 (0.94-1.89)	0.10
65to79 years	114 (99)	1.63	135 (114)	1.88	1.16 (0.81-1.67)	0.41
≥80 years	105 (93)	3.96	83 (73)	3.11	0.8 (0.52-1.21)	0.28
All ages	413 (412)	0.79	523 (513)	0.98	1.26 (1.05-1.52)	0.020
Non-PCV-13 IPD (plus serotype 3)						
<1 years	73 (86)	12.99	66 (84)	12.63	0.97 (0.6-1.56)	0.89
1to4 years	118 (120)	4.71	172 (192)	7.54	1.58 (1.12-2.21)	0.010
5to14 years	55 (56)	0.92	61 (64)	1.05	1.13 (0.67-1.92)	0.65
15to44 years	521 (563)	2.62	457 (480)	2.24	0.86 (0.72-1.04)	0.11
45to64 years	1231 (1200)	9.16	976 (920)	7.03	0.78 (0.69-0.88)	<0.001
65to79 years	1343 (1163)	19.25	1154 (974)	16.11	0.84 (0.75-0.95)	0.004
≥80 years	1193 (1060)	44.97	917 (811)	34.41	0.77 (0.68-0.88)	<0.001
All ages	4534 (4518)	8.66	3803 (3732)	7.15	0.84 (0.78-0.89)	<0.001

**IPD cases adjusted as if they had the same population as in 2009/10 in both years, for comparison purposes. There were no cases with missing age in this period and the adjustment accounting for non-serotyped cases does not have an impact in overall IPD cases.*

*** Incidence rate ratio calculated using a Poison regression model with overdispersion*

Figure S7. Simpson diversity index (SDI) of GPSCs within serotypes 3, 19F, 19A and 4 by financial year

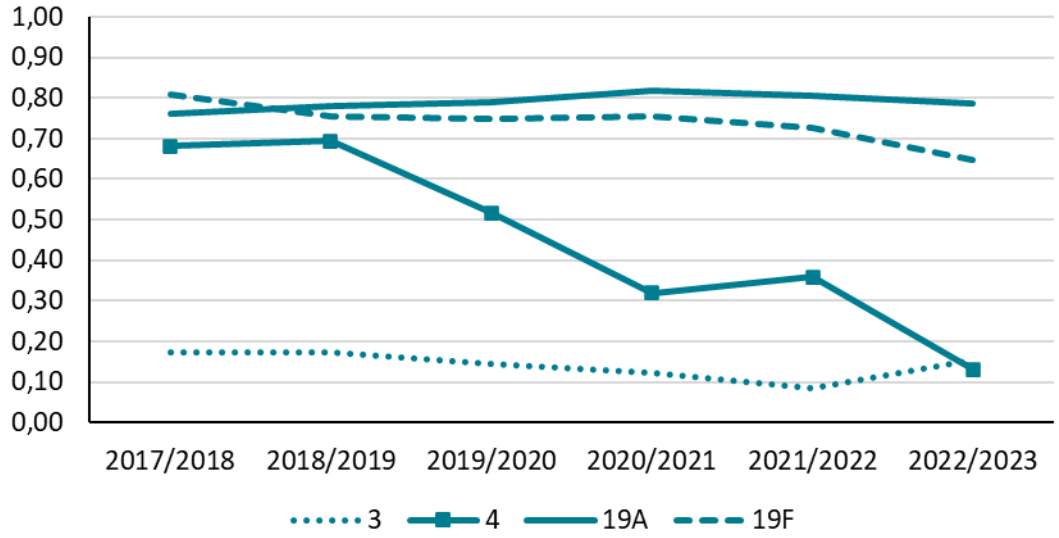


Figure S8. GPSCs by financial year for serotype 3 isolates: number and proportion

As described in the main text, whole genome sequencing identified the same circulating strains among serotype 3 IPD isolates before and after pandemic restrictions, with most cases belonging to GPSC12 (mainly ST180 within this). We did not identify any new strains within serotype 3 which might be responsible for the recent increase, but amongst GPSC 12 clade IV is now predominating. However, although in small numbers, serotype 3 GPSC1 isolates appears to be increasing with 11 cases in 2022/23 compared to only three previous cases in the whole period (one in 2018/19 and two in 2021/22).

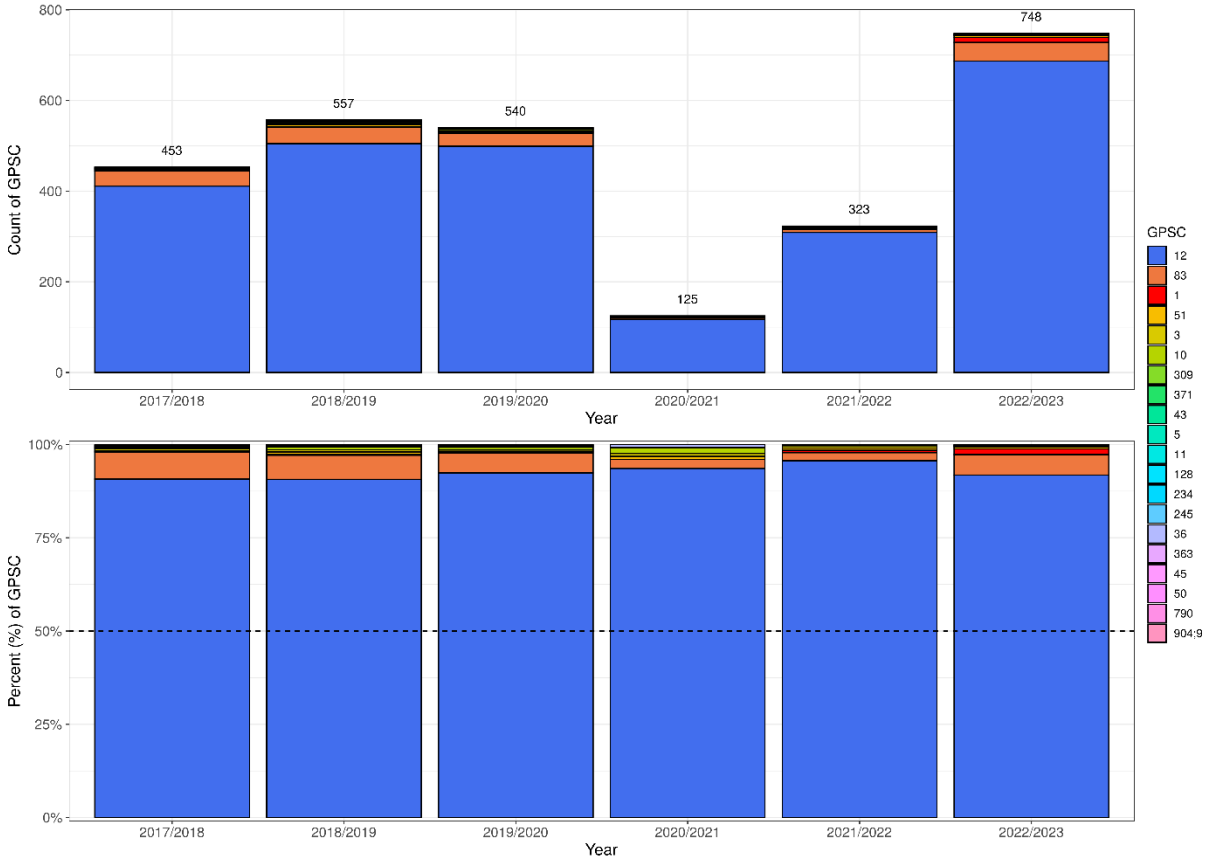


Figure S9. Maximum likelihood phylogeny formed from the non-recombinant regions of the 2568 isolate GPSC12 alignment for serotype 3 isolates.
Note: The inner annotation ring represent the clade nomenclature (Kwun et al 2022). Clades are ordered by overall frequency in the legend from most to least. The outer ring represents the year of isolation for the isolate. The branch scale is in SNPS.

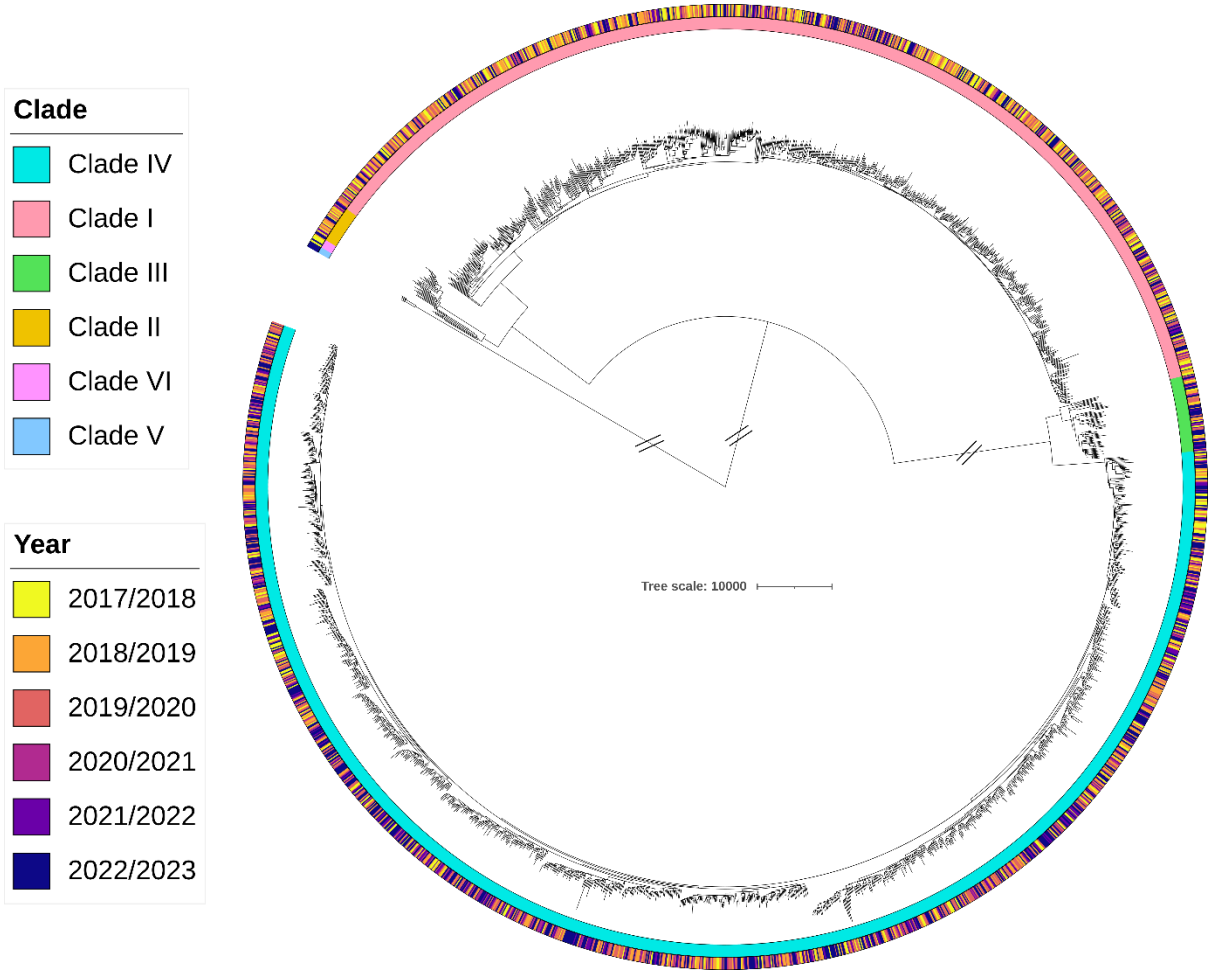


Figure S10. GPSCs by financial year for serotype 19F isolates: number and proportion

In contrast to serotype 3, there is more diversity in the strains expressing serotype 19F. However, its diversity has reduced in this period (**Figure S7**), mainly driven by the expansion of GPSC119, alongside a reduction of GPSC6 isolates.

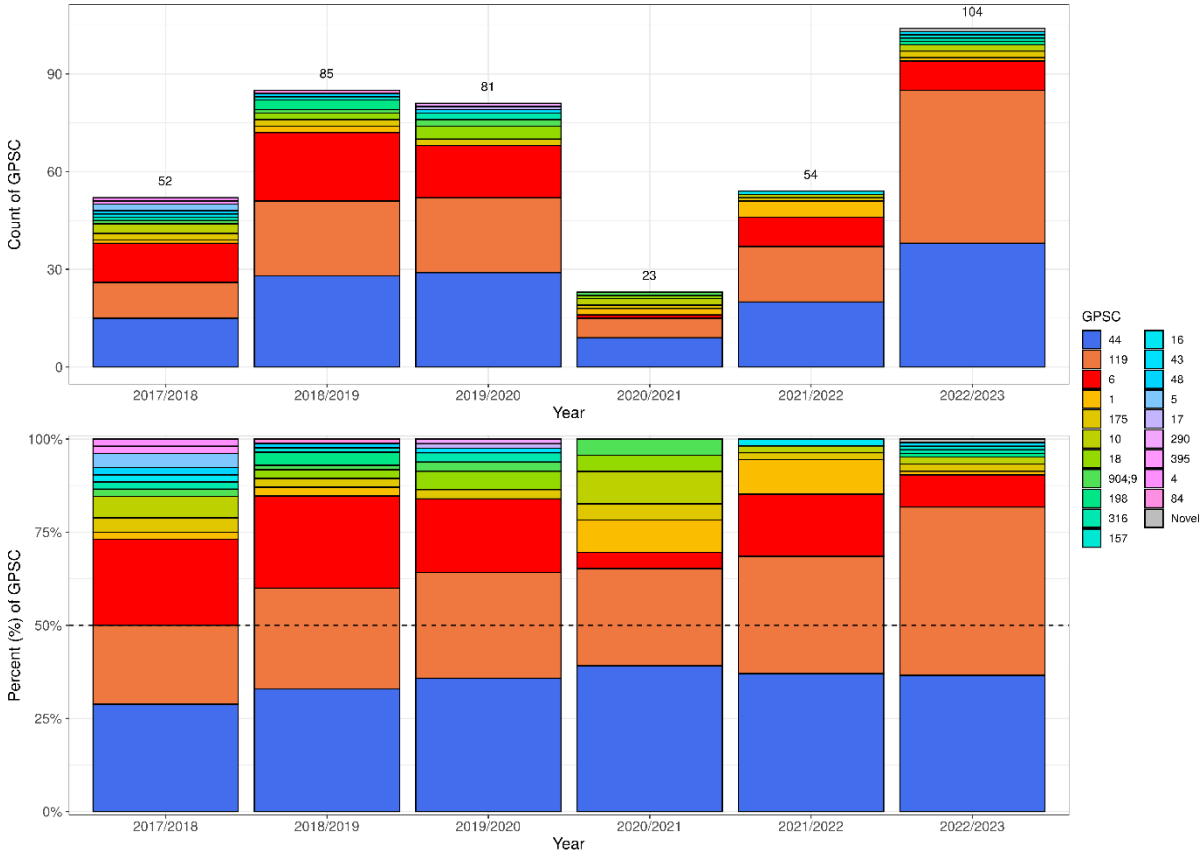


Figure S11 GPSCs by financial year for serotype 19A isolates: number and proportion

Diversity has remained relatively stable despite some changes. We observed a downward trend in the contribution of GPSC4 alongside an upward trend in the contribution of GPSC17. Within GPSC17, this spread is entirely caused by ST2062, which makes up 100% of the 60 isolates in 2022/23 (n=60) and 99% (224/225) of the total GPSC17 isolates.

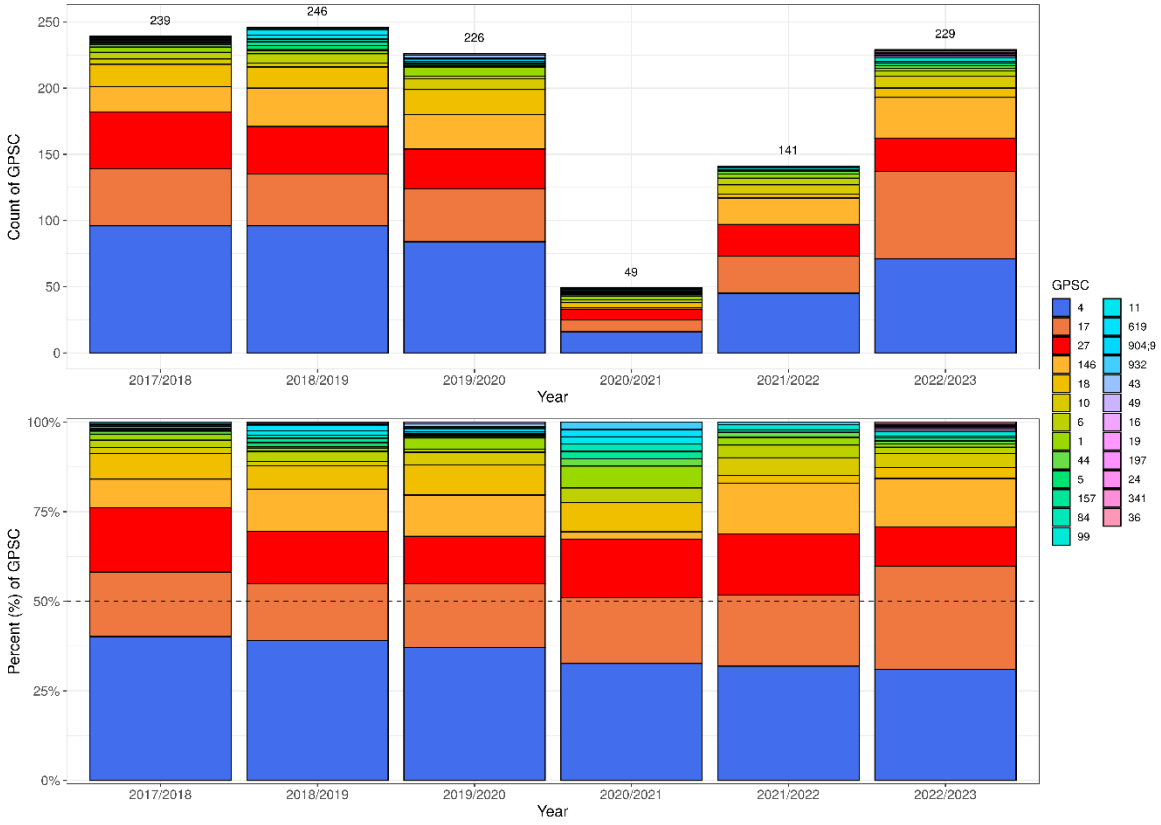
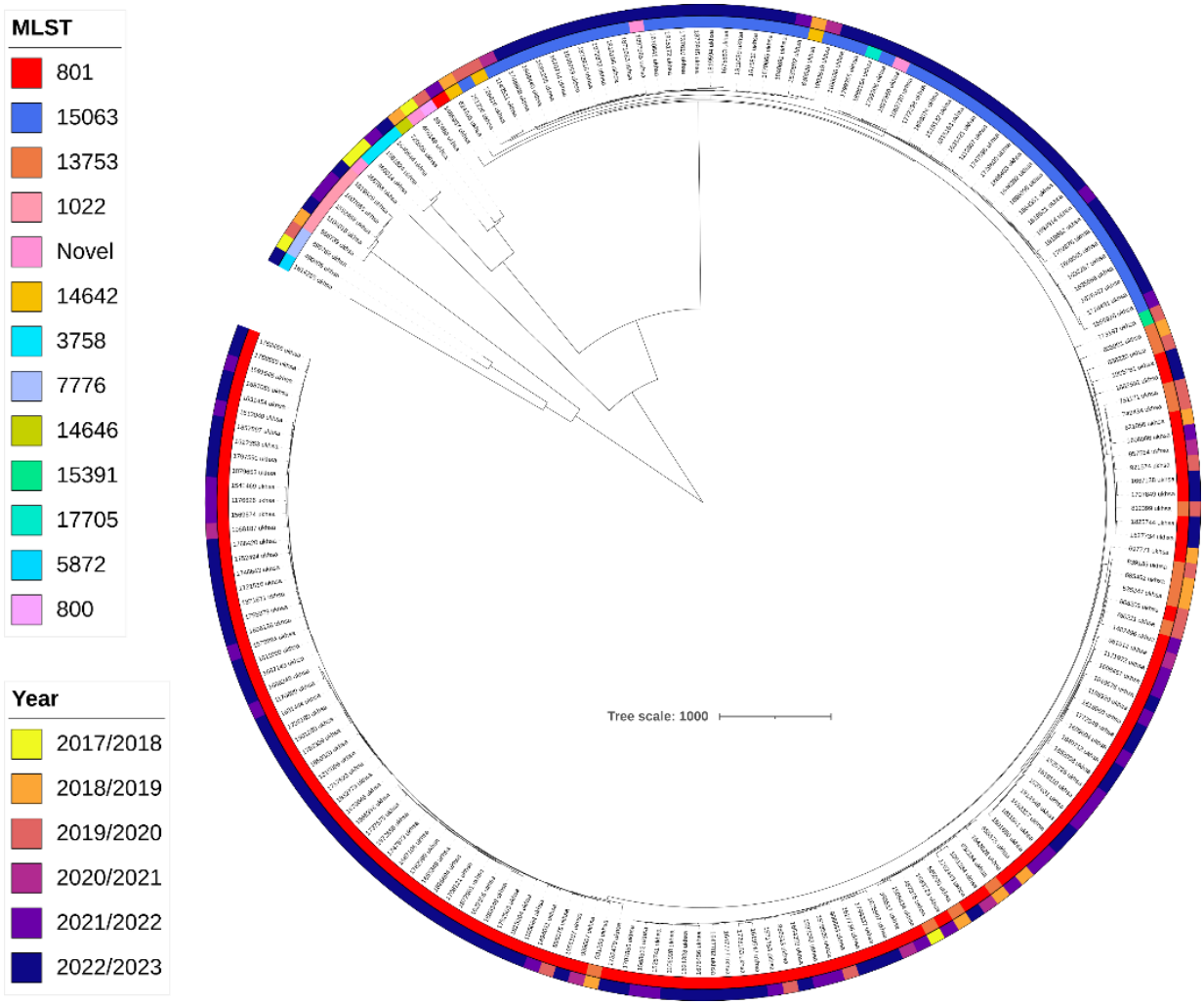


Figure S12. Maximum likelihood phylogeny formed from the non-recombinant regions of the 196 isolate GPSC162 alignment for serotype 4 isolates. Tips are labelled with isolate name. The inner annotation ring represents the MLST of the tip, with the legend ordered by overall MLST frequency in the population. The outer ring represents the year of isolation for the tip. The branch scale is in SNPs (bottom)



Additional references

1. Tewolde R, Dallman T, Schaefer U, et al. MOST: a modified MLST typing tool based on short read sequencing. *PeerJ* 2016; **4**: e2308.
2. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018; **3**: 124.
3. Kapatai G, Sheppard CL, Al-Shahib A, et al. Whole genome sequencing of Streptococcus pneumoniae: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ* 2016; **4**: e2477-e.
4. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**(8): 1072-5.
5. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**(7): 1043-55.
6. Kwun MJ, Ion AV, Cheng H-C, et al. Post-vaccine epidemiology of serotype 3 pneumococci identifies transformation inhibition through prophage-driven alteration of a non-coding RNA. *Genome Medicine* 2022; **14**(1): 144.
7. Simonsen M, Mailund T, Pedersen CNS. Inference of Large Phylogenies Using Neighbour-Joining. 2011; Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 334-44.
8. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**(9): 1312-3.
9. Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2**(11): e000093.
10. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 2021; **49**(W1): W293-W6.
11. Gladstone RA, Siira L, Brynildsrud OB, et al. International links between Streptococcus pneumoniae vaccine serotype 4 sequence type (ST) 801 in Northern European shipyard outbreaks of invasive pneumococcal disease. *Vaccine* 2022; **40**(7): 1054-60.
12. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 2009; **25**(15): 1968-9.
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 Research* 2014; **43**(3): e15-e.