

1 **Title**

2 Evaluation of pneumococcal serotyping in nasopharyngeal carriage isolates by latex  
3 agglutination, whole genome sequencing (PneumoCaT) and DNA microarray in a  
4 high pneumococcal carriage prevalence population in Malawi

5 **Authors**

Todd D. Swarthout (MSc)<sup>a,b,\*,#</sup>, Andrea Gori (PhD)<sup>b,\*</sup>, Naor Bar-Zeev (PhD)<sup>c,a</sup>, Arox  
W. Kamng'ona (PhD)<sup>d</sup>, Thandie S. Mwalukomo (MMed)<sup>e</sup>, Farouk Bonomali (BSc)<sup>a</sup>,  
Roseline Nyirenda (BSc)<sup>a</sup>, Comfort Brown<sup>a</sup>, Jacqueline Msefula (BSc)<sup>a</sup>, Dean Everett  
(PhD)<sup>f</sup>, Charles Mwansambo (MBChB)<sup>g</sup>, Katherine Gould (PhD)<sup>h,i</sup>, Jason Hinds  
(PhD)<sup>h,i</sup>, Robert S. Heyderman (PhD)<sup>a,b,\*</sup>, Neil French (PhD)<sup>a,j,†</sup>

6 a. Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre,  
7 Malawi

8 b. NIHR Global Health Research Unit on Mucosal Pathogens, Division of Infection  
9 and Immunity, University College London, London, United Kingdom

10 c. International Vaccine Access Center, Department of International Health, Johns  
11 Hopkins University, Baltimore, USA

12 d. Department of Biomedical Sciences, College of Medicine, University of Malawi,  
13 Blantyre, Malawi

14 e. Department of Medicine, College of Medicine, University of Malawi, Blantyre,  
15 Malawi

16 f. The Queens Medical Research Institute, University of Edinburgh, Edinburgh

17 g. Ministry of Health, Lilongwe, Malawi

18 h. Institute for Infection and Immunity, St. George's, University of London, London,  
19 United Kingdom.

20 i. BUGS Bioscience, London Bioscience Innovation Centre, London, United  
21 Kingdom

22 j. Centre for Global Vaccine Research, Institute of Infection and Global Health,  
23 University of Liverpool, Liverpool, United Kingdom

#Corresponding author: Todd Swarthout, Malawi-Liverpool-Wellcome Trust Clinical  
Research Programme, P.O. Box 30096, Chichiri Blantyre 3, Malawi; Email:  
[t.swarthout@ucl.ac.uk](mailto:t.swarthout@ucl.ac.uk); Phone: +265 (0)1 874 628

24 ¥ Todd D. Swarthout and Andrea Gori contributed equally to this work. Author order  
25 was determined by seniority at site of study implementation.

26 ¶ Neil French and Robert S. Heyderman contributed equally to this manuscript.  
27 Author order was determined by alphabetical order of surname.

28

### 29 **Keywords (5/5)**

30 *Streptococcus pneumoniae*, serotyping, latex agglutination, microarray, whole  
31 genome sequencing, methodology, Africa

### 32 **Running title**

33 Serotyping for pneumococcal carriage

### 34 **Summary of the article's main points**

35 Assessment of pneumococcal serotype distribution associated with colonization and  
36 disease is essential for evaluation of pneumococcal vaccines. Latex serotyping is  
37 adequate for surveillance, but whole genome sequencing and microarray should be  
38 considered at regional reference laboratories.

39 **Abstract**

40 **Background.** Accurate assessment of the serotype distribution associated with  
41 pneumococcal colonization and disease is essential for the evaluation and  
42 formulation of pneumococcal vaccines and informing vaccine policy.

43

44 **Methods.** We evaluated pneumococcal serotyping concordance between latex  
45 agglutination, PneumoCaT by whole genome sequencing (WGS) and DNA  
46 microarray using samples from community carriage surveillance in Blantyre, Malawi.  
47 Nasopharyngeal swabs were collected, following WHO recommendations, between  
48 2015 and 2017, using stratified random sampling among study populations.  
49 Participants included healthy children 3–6 years old (PCV13 vaccinated as part of  
50 EPI), healthy children 5–10 years (age-ineligible for PCV13), and HIV-infected adults  
51 (18–40yrs) on ART. For phenotypic serotyping we used a 13-valent latex kit (SSI,  
52 Denmark). For genomic serotyping we applied PneumoCaT pipeline to whole  
53 genome sequence libraries. For molecular serotyping by microarray we used the  
54 BUGS Bioscience Senti-SP microarray.

55

56 **Results.** 1347 samples were analysed. Concordance was 90.7% (95% CI: 89.0–  
57 92.2) between latex and PneumoCaT; 95.2% (93.9–96.3) between latex and  
58 microarray; and 96.6% (95.5–97.5) between microarray and PneumoCaT. By  
59 detecting additional vaccine serotype (VT) pneumococcus carried at low relative  
60 abundance (median 8%), microarray increased VT detection by 31.5% compared to  
61 latex serotyping.

62

63 **Conclusion.** All three serotyping methods were highly concordant in identifying  
64 dominant serotypes. Latex serotyping is accurate in identifying vaccine-serotypes  
65 and requires the least expertise and resources for field-implementation and analysis.  
66 However, WGS, which adds population structure, and microarray, which adds  
67 multiple-serotype carriage, should be considered at regional reference laboratories  
68 while investigating the importance of VT in low relative abundance in transmission  
69 and disease.

70

## 71 **Introduction**

72 *Streptococcus pneumoniae* colonises the nasopharynx of healthy individuals.  
73 Although carriage is usually asymptomatic, nasopharyngeal (NP) colonization is a  
74 prerequisite for disease including otitis media, sinusitis, pneumonia, bacteraemia,  
75 and meningitis. (1) The pneumococcus is estimated to be responsible for over 318  
76 000 (uncertainty ratio [UR]: 207 000–395 000) deaths every year in children aged 1  
77 to 59 months, with the highest mortality burden among African children.(2) Evidence  
78 also shows that HIV-infected children and adults are at significantly higher risk of  
79 invasive pneumococcal disease (IPD) than their HIV-uninfected counterparts. (3, 4)

80

81 Current multivalent pneumococcal conjugate vaccines (PCV) target subsets of the  
82 100 capsular serotypes known to be expressed by the pneumococcus. PCV reduces  
83 nasopharyngeal carriage of the sub-set of pneumococcal serotypes they contain,  
84 known as vaccine serotypes (VT). With reduced carriage among the vaccinated  
85 there is then reduced risk of VT-IPD (direct protection) and reduced transmission,  
86 therefore reduced risk of VT-IPD among those PCV-unvaccinated (indirect  
87 protection). However, non-vaccine serotypes (NVT) have the potential to fill the

88 ecological niche, becoming more common in carriage and disease. (5-7) This  
89 phenomenon, known as serotype replacement, may be more pronounced in low-  
90 income settings because of higher prevalence, density and diversity of  
91 pneumococcal carriage, and represents a considerable risk to the global  
92 pneumococcal immunisation strategy.(8) Serotype distribution differs between  
93 continents as well as individual countries. (9) Given these differences, accurate  
94 assessment of the serotype distribution associated with both pneumococcal  
95 colonization and pneumococcal disease is needed in the evaluation, formulation and  
96 delivery of pneumococcal vaccines.

97

98 A pneumococcal serotyping method suitable for use in robust carriage and  
99 surveillance studies should therefore, at minimum, be accurate in its serotype  
100 assignment, particularly in relation to VTs. Additional desirable parameters include  
101 detection of most or all serotypes, ability to detect multiple serotypes in carriage  
102 (common in high burden settings (10, 11), more in-depth information on genotype,  
103 suitable to scale up for large projects, and practical for resource-poor settings.  
104 Unfortunately, work in resource-poor settings can too often limit the number of these  
105 parameters that can be achieved.

106

107 The gold-standard serotyping method, the Quellung reaction, was developed in the  
108 early 1900s and is performed by testing colonies with a set of type-specific antisera.  
109 (12) Bacteria are observed by microscopy, with serotype defined by observing  
110 apparent capsular swelling in reaction to the type-specific antisera. It is laborious,  
111 requires frequent use to maintain skills, requires a complete set of type-specific  
112 antisera, and is therefore mainly performed by reference laboratories. The

113 PneuCarriage project, a large, multi-centre study, was established with the aim of  
114 identifying the best pneumococcal serotyping methods for carriage studies. (13) The  
115 Project identified microarray with a culture amplification step as the top-performing  
116 method. While robust and systematic, their decision algorithm did not take into  
117 account parameters such as cost, skill level and resources needed for assay  
118 implementation and maintenance, as well as output processing and interpretation.

119

120 Here we describe, in the context of an ongoing field-based study, (14) the level of  
121 concordance between three methods commonly used during ongoing routine  
122 pneumococcal surveillance activities in our work: latex agglutination, microarray and  
123 serotyping-by-sequencing. We also address parameters that researchers and  
124 policymakers can consider when deciding which assay to implement in their local  
125 setting.

126

## 127 **Materials and Methods**

### 128 **Study Setting**

129 Blantyre is located in southern Malawi with an urban population of approximately 1.3  
130 million.

131

### 132 **Study Population and Recruitment**

133 Samples were collected as part of a larger 3.5-year pneumococcal carriage  
134 surveillance project, as described elsewhere. (14) In brief, this was a prospective  
135 rolling cross-sectional observational study using stratified random sampling to  
136 measure pneumococcal nasopharyngeal carriage in Blantyre, Malawi. Samples used  
137 in this analysis were collected during the first two years of twice-annual cross-

138 sectional surveys, from June 2015 to April 2017. Recruitment included three groups:  
139 i) healthy children 3–6 years old who received PCV13 as part of routine EPI, ii)  
140 healthy children 5–10 years old who were age-ineligible to receive PCV13 as part of  
141 EPI, and iii) HIV-infected adults (18–40yrs) on ART.

142

### 143 **Sample Selection**

144 For concordance analyses between the three methods, all samples were included  
145 that had serotyping results available from each of the three methods (latex,  
146 microarray, serotyping-by-sequencing). From the total nasopharyngeal swab (NPS)  
147 samples collected during the larger surveillance project (including 1,044 from  
148 children 3–6 years old [PCV-vaccinated], 531 children 5–10 years old [PCV-  
149 unvaccinated, age-ineligible] and 428 HIV-infected adults on ART), 1347 samples  
150 were culture-confirmed for *S. pneumoniae* and also had results available from the  
151 microarray and serotyping-by-sequencing. The final concordance analysis included  
152 846 children 3–6 years old (PCV13-vaccinated), 422 children 5–10 years old (age-  
153 ineligible for PCV13 vaccination) and 79 adults (HIV-infected and PCV13-  
154 unvaccinated). (Figure 1) Sample selection for microarray and serotyping-by-  
155 sequencing was done independently and blind to latex serotype data.

156

### 157 **Nasopharyngeal Swab Collection**

158 Collection of NP swabs is described elsewhere. (14) In brief, an NP swab sample  
159 was collected from each participant using a nylon flocced swab (FLOQSwabs™,  
160 Copan Diagnostics, Murrieta, CA, USA) and then immediately placed into 1.5mL  
161 skim milk-tryptone-glucose-glycerol (STGG) medium and processed at the Malawi–

162 Liverpool–Wellcome Trust (MLW) laboratory in Blantyre, according to WHO  
163 recommendations. (15) Samples were frozen on the same day at  $-80^{\circ}\text{C}$ . (Figure 2)

164

#### 165 **NPS Culture for Pneumococcal Screening & Serotyping**

166 30 $\mu\text{L}$  NPS–STGG was plated on a sterile sheep blood-gentamicin (SBG; 7% SBA, 5  
167  $\mu\text{L}$  gentamicin/mL) agar plate (primary plate) and incubated overnight at  $37^{\circ}\text{C}$  in ~5%  
168  $\text{CO}_2$ . Plates showing no *S. pneumoniae* growth were incubated overnight a second  
169 time before being reported as negative. *S. pneumoniae* was identified by colony  
170 morphology and optochin disc (Oxoid, Basingstoke, UK) susceptibility. The bile  
171 solubility test was used on isolates with no or intermediate (zone diameter  $<14\text{mm}$ )  
172 optochin susceptibility. A single colony of confirmed pneumococcus was selected  
173 and grown on a sterile SBG plate (secondary plate), following the same process as  
174 the primary plate. (Figure 1)

175

#### 176 **Latex Serotyping**

177 Pneumococcal growth from secondary plates was used for serotyping by latex  
178 agglutination (ImmuLex™ 7-10-13-valent Pneumotest; Statens Serum Institute,  
179 Denmark), following manufacturer guidelines. Using a reaction card and a sterile  
180 inoculation loop, a small sweep of an overnight bacterial culture was mixed with  
181 saline and a series of individual Pneumotest-Latex reagents in suspension. The card  
182 was rocked manually and observed for agglutination. A Pneumotest-Latex  
183 chessboard was used to determine which serotype is associated with the observed  
184 set of agglutination reactions. The kit allows for differential identification of each  
185 PCV13 VT (1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, 23F). Other than for a  
186 limited number of serogroups (6, 7, 9, 18, 19, 23) for which the kit provides



187 serogroup differentiation, there is no further differential identification of NVT  
188 serogroups or serotypes. NVT and non-typeable isolates were reported as NVT.  
189 Samples were batch-tested on a weekly basis, with technicians blinded to the  
190 sample source. After serotyping was complete, the remaining growth from each  
191 secondary plate was archived at -80°C in sterile STGG. Refer to Text S1 in the  
192 supplemental material for a more detailed description of latex serotyping.

193

#### 194 **Molecular Serotyping by Microarray**

195 For samples with culture-confirmed pneumococcal carriage, the original inoculated  
196 STGG was thawed and vortexed. Aliquots of 100µl were shipped in 1.8mL cryovials  
197 to BUGS Bioscience (BUGS Bioscience Ltd., London, United Kingdom) on dry ice.  
198 (Figure 1) The remaining steps for microarray serotyping (including sample  
199 processing, culturing, DNA extraction, microarray and analysis) were completed  
200 entirely by BUGS Bioscience. (16, 17) Final microarray results were retrieved by the  
201 study team from BUGS Bioscience web-based SentiNET platform and imported into  
202 STATA 13.1 (StataCorp, College Station, TX, USA) for analysis. Refer to Text S1 in  
203 the supplemental material for a more detailed description of microarray serotyping.

204

#### 205 **DNA Extraction and WGS**

206 Archived secondary growth isolates were used to develop sequence libraries for  
207 serotyping-by-sequencing. To optimise total retrieved DNA, 30µl of thawed isolate-  
208 STGG was incubated overnight in 6mL THY (Todd Hewitt broth + yeast) enrichment  
209 culture. DNA was extracted from the overnight culture using the Qiagen® QIAamp™  
210 DNA Mini Kit, following manufacturer guidelines for bacterial DNA. Quality control  
211 (QC) measures, as required by the guidelines of the sequencing institution, included

212 DNA quantification (Qubit™, Thermo Fisher Scientific, Massachusetts, USA) of all  
213 DNA samples and gel electrophoresis imaging on 0.7% agarose to assess DNA  
214 integrity. After attaining quantity and quality requirements, 100µL of extracted DNA  
215 were aliquoted into skirted 96-well microwell plates and stored at -80°C until shipped  
216 on dry ice to the Oxford Genomics Centre (University of Oxford, United Kingdom) for  
217 sequencing. Whole genome sequencing was performed at the Oxford Genomics  
218 Centre on a HiSeq4000 platform (Illumina™), with paired-end libraries and a read  
219 length of 150bp.

220

### 221 **Serotyping-by-sequencing**

222 WGS data was retrieved by the study team from a web-based FTP link. Serotype  
223 was inferred from the isolates' genome sequences using the PneumoCaT software  
224 pipeline, an opensource bioinformatic tool. (18) PneumoCaT requires raw  
225 sequencing reads for each isolate which were trimmed and cleaned. Reads were  
226 trimmed of the illumina adapters and cleaned of low-quality ends using Trimmomatic  
227 (ver. 0.38; available at <http://www.usadellab.org/cms/?page=trimmomatic>). Minimum  
228 read length after trimming was 80 base pairs (bp), with the minimum average quality  
229 for a sliding window of 4 nucleotide being 15. A subset of 700,000 reads per end (1.4  
230 million total) was used for any subsequent analysis. XML result files were parsed  
231 with ad-hoc bash scripts, in order to extract and tabulate the serotyping result for  
232 each isolate. PneumoCaT was installed and used on a Linux machine at the MRC  
233 Cloud Infrastructure for Microbial Bioinformatics (CLIMB; <https://www.climb.ac.uk/>).  
234 Each serotype identification required an average 5-8 minutes. Refer to Text S1 in the  
235 supplemental material for a more detailed description of serotyping-by-sequencing.

236

237 **Definitions**

238 Concordance was calculated with all samples aggregated and according to the level  
239 of discrimination provided by the method. Concordance is reported using two criteria:  
240 i) a criterion based on whether both assays reported NVT or both reported VT  
241 (VT/NVT criterion) and ii) a criterion based on whether the final serotype reported by  
242 each assay is equivalent (serotype-specific criterion).

243

244 Concordance between latex and serotyping-by-sequencing (PneumoCaT): Other  
245 than a limited number of serogroups (6, 7, 9, 18, 19, 23) for which the latex kit  
246 provides serogroup differentiation, there is no further differential identification of NVT  
247 serogroups to serotype. NVT and non-typeable isolates were reported as NVT.  
248 Concordance at serotype level (serotype-specific criterion) was reported only if latex  
249 reported VT carriage. If latex reported NVT, any NVT reported by PneumoCaT was  
250 considered concordant. For example: 23F reported by both latex and PneumoCaT  
251 was considered concordant, as was NVT and 15B. However, 19F and 19A was  
252 considered discordant, as was NVT and 6B.

253

254 Concordance between latex and microarray: Concordance at serotype level  
255 (serotype-specific criterion) was reported only if latex reported VT carriage. If latex  
256 reported NVT, any NVT reported by microarray was considered concordant.  
257 Because microarray reports multiple serotype carriage, 23F reported by latex and  
258 23F+34 reported by microarray was considered concordant, as was NVT and  
259 18C+33D. However, 19F and 33D+19A was considered discordant, as was NVT and  
260 3+7F.

261

262 Note that for microarray, some closely related serotypes are reported as a group,  
263 with the final individual serotype call in brackets (e.g., 6A/B [6B]). In this case, results  
264 were analysed using the individual serotype call. For example, if microarray reported  
265 6A/B [6B], this was considered discordant to a 6A latex result and concordant to a  
266 6B latex result. For simplicity of analysis, if a method did not claim to detect a  
267 serotype but the sample contained that serotype, this result was deemed discordant.  
268 For example, if microarray detected both serotypes 23F and 19A but latex detected  
269 only serotype 3, this result was considered discordant.

270

271 Concordance between microarray and serotyping-by-sequencing (PneumoCat):  
272 Microarray and PneumoCat both differentiate VT and NVT to serotype level, allowing  
273 concordance to be calculated on serotype concordance (serotype-specific criterion)  
274 for both VT and NVT *S. pneumoniae*.

275

### 276 **Statistical analysis**

277 The formula for percent increase in VT prevalence was:  $([VT \text{ prevalence using latex}$   
278  $- VT \text{ prevalence using microarray}] / VT \text{ prevalence using latex}) * 100\%$ . Confidence  
279 intervals are binomial exact. Statistical significance was inferred from two-sided  
280  $p < 0.05$ . Participant data collection was completed using Open Data Kit (ODK)  
281 Collect open source software. (v1.24.0). Statistical analyses were completed using  
282 Stata 13.1 (StataCorp, College Station, TX, USA).

283

### 284 **Ethics Considerations**

285 The study protocol was approved by the College of Medicine Research and Ethics  
286 Committee, University of Malawi (P.02/15/1677) and the Liverpool School of Tropical

287 Medicine Research Ethics Committee (14.056). Adult participants and  
288 parents/guardians of child participants provided written informed consent; children 8-  
289 years and older provided informed assent. This included consent for publication.

290

## 291 **Results**

292 Pneumococcal carriage prevalence results from the larger surveillance project are  
293 reported elsewhere. (14) Comparing latex and PneumoCat, the adjusted  
294 concordance of correctly identifying pneumococcal carriage as VT or NVT was  
295 90.7% (1216/1341; 95% CI: 89.0–92.2). (Figure 3) Based on the serotype-specific  
296 criterion, concordance between latex and PneumoCaT was 87.5%; (1174/1341)  
297 (95% CI: 85.7–89.3). Comparing latex and microarray, the concordance based on  
298 correctly identifying pneumococcal carriage as VT or NVT was 97.3% 1311/1347  
299 (95% CI: 96.3–98.1). Based on a serotype-specific criterion, the concordance was  
300 95.2% (1282/1347; 95% CI: 93.9–96.3). Comparing microarray and PneumoCaT,  
301 concordance based on correctly identifying pneumococcal carriage as VT or NVT  
302 was 96.6% (1295/1341; 95% CI: 95.5–97.5). Based on a serotype-specific criterion,  
303 the concordance was 82.8% (1110/1341; 95% CI: 80.6–84.8).

304

## 305 **Increased VT Detection Using Microarray**

306 Using a larger study database of 1,949 samples from the same study, we evaluated  
307 latex and microarray data. Aggregating all ages (i.e. child and adult), there was an  
308 increase of 31.5% in VT prevalence by microarray compared to latex serotyping:  
309 43.0% increase in VT carriage among children 3–6 years old, 21.7% among children  
310 5–10 years old and 10.8% among HIV-infected adults on ART (Table 1). This was  
311 due to samples reporting NVT by latex but that also carried VT, as detected by

312 microarray. These VT, undetected by latex, were carried in lower relative abundance  
313 (median 8%, range: 2% - 48%). The prevalence of multiple serotype carriage (range  
314 2-6 serotypes) was 35.2% (686/1949). The prevalence among respective age groups  
315 was 44.4% (457/1029), 32.8% (169/515), and 14.8% (60/405). Among samples with  
316 multiple serotype carriage, latex identified the dominant serotype in 85.3% (585/686;  
317 95%CI: 82.4–87.8) of samples. Despite the overall increase in detection of VT  
318 carriage, the proportion of individual VT serotypes detected is not different when  
319 comparing microarray to latex (Figure 4). Refer to Table S1 in the supplemental  
320 material for the reported frequency of each vaccine serotype (VT) detected by  
321 microarray and latex.

322

### 323 **Key Parameters of selected serotyping methods.**

324 Table 2 presents key parameters to further consider when deciding which assay is  
325 appropriate for a particular setting. Estimated costs and feasibility of implementation  
326 and maintenance are specific to the setting in Malawi at the Malawi-Liverpool-  
327 Wellcome Trust Clinical Research Programme in Blantyre. Extrapolation would need  
328 further validation outside the scope of this evaluation. Though more limited in its  
329 reporting only a single serotype, latex is highly accurate while being less costly and  
330 requiring less expertise and resources for field-implementation and analysis. While  
331 microarray is the costliest option, it provides greater accuracy of total pneumococcal  
332 carriage, including multiple serotype carriage and relative abundance of individual  
333 serotypes in carriage. Whole genome sequencing is a strong alternative to latex and  
334 would be nearly cost-free if the sequence libraries were already available. In  
335 addition, WGS provides opportunity for further analyses, including population  
336 structure and antibiotic resistance.

337

338 **Discussion**

339 We report high concordance between three serotyping techniques applicable to  
340 routine pneumococcal surveillance. Importantly, we have extended the analysis to  
341 include relevant parameters beyond accuracy including cost, time to result, and  
342 measures of input required for assay implementation and maintenance. These are  
343 parameters that researchers and policy makers should consider when deciding  
344 which assay to implement. All three assays appear accurate and concordant in  
345 identifying the dominant serotype.

346

347 While latex agglutination is accurate and requires the least expertise and resources  
348 for field-implementation and analysis and provides rapid results, standard latex  
349 approaches is not optimal for optimal surveillance of vaccine impact, including the  
350 detection of multiple serotype carriage and VT in low relative abundance. (19) There  
351 have been attempts to implement latex for detection of multiple serotype carriage.  
352 Gratten et al. serotyped up to six colonies from nasal-swab culture plates and found  
353 multiple-serotype carriage in 29.5% of Papua New Guinean children. (20) The  
354 authors went on to serotype at least 50 colonies from 10 selected nasal-swab  
355 cultures and concluded that the minor carried serotype accounted for 4 to 27% of the  
356 total pneumococcal population. A review of published data on multiple carriage  
357 concluded that, to detect a minor carried serotype it would be necessary to serotype  
358 at least five colonies to have a 95% chance of detecting the serotype if it accounted  
359 for 50% of the total pneumococcal population, and one would need to examine 299  
360 colonies if the serotype was present at a relative abundance of 1%. As part of the  
361 PneuCarriage project, to thoroughly characterise samples, up to 120 colonies from

362 each sample were selected to achieve >99% power to detect a minor serotype of 5%  
363 abundance. (13) This approach would not be cost- or time-effective. Though  
364 dependent on technical capacity to develop in-house reagents, researchers in The  
365 Gambia developed a latex agglutination technique in which colonies from the primary  
366 culture plate are suspended in saline and serotyped by latex agglutination. (21)  
367 While not differentiating NVT serotypes, they did show that up to 10.4% of  
368 pneumococcal acquisitions were found to be of multiple serotypes in a longitudinal  
369 infant cohort study. While latex is limited in its output, the process can be leveraged  
370 for additional endpoints including, for example, measuring carriage density through  
371 counting of colony-forming units (CFU) on agar culture plates. What's more, though  
372 less cost- and time-effective, use of both commercial products (including Statens  
373 Serum Institute, Denmark) and production of in-house latex serotyping reagents has  
374 been well-documented to significantly expand the number of NVT serotypes that can  
375 be differentiated by latex, including quality control procedures. (11, 22)

376

377 With opensource bioinformatic tools such as PneumoCaT, serotyping-by-sequencing  
378 can be less costly than microarray, even accounting for costs of DNA extraction and  
379 WGS, while still being able to differentiate non-typeable and nearly every known VT  
380 and NVT. Though we would not recommend initiating DNA extraction and WGS for  
381 the use of PneumoCaT alone, sequence libraries can be further leveraged for  
382 extensive informative bioinformatic analyses, useful in population biology,  
383 antimicrobial resistance investigations and vaccine monitoring. Moreover, using  
384 PneumoCaT for serotyping would be essentially cost-free if the sequence libraries  
385 were already available, apart from the limited bioinformatic skills needed. While  
386 microarray is more costly, it differentiates NVT and multiple serotype carriage with



387 relative abundance, as well as non-*S. pneumoniae* contaminants (i.e. *S. mitis*, *S.*  
388 *salivarius*, *Staphylococcus aureus*) with a degree of precision. This technique stands  
389 out for its sensitivity, being able to detect serotypes in low relative abundance, which  
390 is of critical importance for understanding the transmission patterns of *S.*  
391 *pneumoniae*. Having the extra counts for each serotype from the same number of  
392 samples, as provided by microarray, also has the advantage of adding power to a  
393 study's statistics.

394

395 There are a number of limitations to mention, including the number of serotyping  
396 methods which were not evaluated, including PCR and the SeroBA pipeline. SeroBA  
397 is a relatively new serotyping-by-sequencing software. With similar accuracy to  
398 PneumoCaT, SeroBA does have operational advantages. (23) SeroBA can correctly  
399 call a serotype with a read coverage as low as 10X (20X is required for  
400 PneumoCaT). Using a k-mer based approach, rather than the raw sequence  
401 alignment, SeroBA requires much lower computational power and time. On the other  
402 hand, the PneumoCaT source code can be easily adapted to the operator needs,  
403 and both software are likely to run on a standard server configuration. Alternative  
404 culture-independent methods, such as isolation-independent conventional multiplex  
405 PCR-serotyping (cmPCR) could be important for confirming carriage when re-  
406 culturing of original NP swab samples is not feasible. Though cmPCR has been  
407 successfully applied on DNA extracted directly from NPS-STGG, evidence suggests  
408 that cmPCR serotyping after culture enrichment returns a higher sensitivity and an  
409 ability to identify multiple serotype carriage. (9) Nonetheless, cmPCR can be  
410 confounded by non-pneumococcal streptococci (including *S. mitis*, *S. oralis* and *S.*  
411 *parasanguinis*). (24) Due to a high sequence similarity with target serotype-specific

412 amplicons, cmPCR can overestimate pneumococcal carriage. Carvalho *et al.*, for  
413 example, reported that 82.5% of samples (combined nasopharyngeal and  
414 oropharyngeal swabs) that were positive for pneumococcus by cmPCR were culture-  
415 negative by latex. Similarly, 35.0% of those positive for pneumococcus by cmPCR  
416 were negative by *lytA*-pneumococcal specific PCR. This is particularly problematic  
417 when bacterial DNA is extracted from culture-enriched naso- and oropharyngeal  
418 samples, requiring PCR results to be confirmed by latex agglutination or other  
419 serotyping procedures. (24) Additional PCR limitations include the need for region-  
420 specific reaction protocols, implementing a high number of primer pairs to identify the  
421 same range of serotypes identified by microarray or WGS, and the increased risk of  
422 detecting non-viable pneumococci. As there is no evidence of a viable but non-  
423 culturable (VBNC) state in *S. pneumoniae*, (25) identifying non-viable pneumococci  
424 could be disadvantageous for field-based research. While a formal economic  
425 analysis of the methods would be justified, we were unable to extrapolate the  
426 individual costing components between sites. Such components would include local  
427 salaries and additional labour costs, procurement and shipping of equipment and  
428 consumables, equipment maintenance, local health and safety requirements, and  
429 institutional costs. For this reason, comparative costing is grossly categorized.  
430 Though we did not include invasive isolates (from blood or cerebrospinal fluid, for  
431 example), it is important to identify serotypes associated with IPD, including in post-  
432 PCV impact studies. For invasive isolates, with a single-serotype sample expected,  
433 microarray would have limited advantage. Application of serotyping-by-sequencing  
434 would then be the most informative option, including insight into population structure,  
435 antimicrobial resistance patterns and serotype replacement disease.

436

437 **CONCLUSION**

438 Selection of the appropriate assay should be based on the intended analysis and  
439 endpoint. While accuracy and concordance is high between the three assays,  
440 parameters of field-implementation and cost vary significantly. In a setting of limited  
441 resources, as is true throughout much of sub-Saharan Africa, latex is the best overall  
442 option for decentralised surveillance of vaccine impact. However, WGS, which adds  
443 population structure, and microarray, which adds multiple-serotype carriage, should  
444 be considered at regional reference laboratories while investigating the importance  
445 of VT in low relative abundance in transmission and disease.

446

447 **Supplemental Material**

448 Supplemental material is available online. Consisting of data provided by the authors  
449 to benefit the reader, the posted materials are not copyedited and are the sole  
450 responsibility of the authors, so questions or comments should be addressed to the  
451 corresponding author.

452

453 **Data availability.** The data supporting the findings of this study has been deposited  
454 in the Figshare repository. (26)

455

456 **Acknowledgements.** The authors thank the individuals who participated in this  
457 study and the local authorities for their support. We are grateful to the hospitality of  
458 the QECH ART Clinic, led by Ken Malisita. Our thanks also extend to the MLW  
459 laboratory management team, led by Brigitte Denis and George Selemani.

460

461 **References**

- 462 1. Simell B, Auranen K, Kayhty H, Goldblatt D, Dagan R, O'Brien KL, Pneumococcal  
463 Carriage G. 2012. The fundamental link between pneumococcal carriage and disease.  
464 *Expert Rev Vaccines* 11:841-55.
- 465 2. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, Luksic I, Nair H,  
466 McAllister DA, Campbell H, Rudan I, Black R, Knoll MD. 2018. Burden of  
467 *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in  
468 the era of conjugate vaccines: global, regional, and national estimates for 2000-15.  
469 *Lancet Glob Health* 6:e744-e757.
- 470 3. von Gottberg A, Cohen C, de Gouveia L, Meiring S, Quan V, Whitelaw A, Crowther-  
471 Gibson P, Madhi SA, Whitney CG, Klugman KP. 2013. Epidemiology of invasive  
472 pneumococcal disease in the pre-conjugate vaccine era: South Africa, 2003-2008.  
473 *Vaccine* 31:4200-8.
- 474 4. Nunes MC, von Gottberg A, de Gouveia L, Cohen C, Kuwanda L, Karstaedt AS,  
475 Klugman KP, Madhi SA. 2011. Persistent high burden of invasive pneumococcal  
476 disease in South African HIV-infected adults in the era of an antiretroviral treatment  
477 program. *PLoS One* 6:e27929.
- 478 5. Singleton RJ, Hennessy TW, Bulkow LR, Hammitt LL, Zulz T, Hurlburt DA, Butler JC,  
479 Rudolph K, Parkinson A. 2007. Invasive pneumococcal disease caused by nonvaccine  
480 serotypes among alaska native children with high levels of 7-valent pneumococcal  
481 conjugate vaccine coverage. *JAMA* 297:1784-92.
- 482 6. Feikin DR, Kagucia EW, Loo JD, Link-Gelles R, Puhon MA, Cherian T, Levine OS,  
483 Whitney CG, O'Brien KL, Moore MR, Serotype Replacement Study G. 2013. Serotype-  
484 specific changes in invasive pneumococcal disease after pneumococcal conjugate

- 485 vaccine introduction: a pooled analysis of multiple surveillance sites. *PLoS Med*  
486 10:e1001517.
- 487 7. Weinberger DM, Malley R, Lipsitch M. 2011. Serotype replacement in disease after  
488 pneumococcal vaccination. *Lancet* 378:1962-73.
- 489 8. Adegbola RA, DeAntonio R, Hill PC, Roca A, Usuf E, Hoet B, Greenwood BM. 2014.  
490 Carriage of *Streptococcus pneumoniae* and other respiratory bacterial pathogens in  
491 low and lower-middle income countries: a systematic review and meta-analysis.  
492 *PLoS One* 9:e103293.
- 493 9. da Gloria Carvalho M, Pimenta FC, Jackson D, Roundtree A, Ahmad Y, Millar EV,  
494 O'Brien KL, Whitney CG, Cohen AL, Beall BW. 2010. Revisiting pneumococcal carriage  
495 by use of broth enrichment and PCR techniques for enhanced detection of carriage  
496 and serotypes. *J Clin Microbiol* 48:1611-8.
- 497 10. Swarthout TD MT, Everett D, Bar Zeev N, Kamng'ona A, Chipasula T, Malisita K,  
498 Mwansambo C, Banda M, Hinds J, Gould KA, French N, Heyderman RS. 2016.  
499 Persistent vaccine-type carriage of *Streptococcus pneumoniae* four years after  
500 introducing PCV13 in a 3+0 schedule in Malawi, abstr International Symposium on  
501 Pneumococci and Pneumococcal Disease, Glasgow, Scotland, 2016.
- 502 11. Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, Nosten F, Goldblatt D.  
503 2011. Improved detection of nasopharyngeal cocolonization by multiple  
504 pneumococcal serotypes by use of latex agglutination or molecular serotyping by  
505 microarray. *J Clin Microbiol* 49:1784-9.
- 506 12. Austrian R. 1976. The quellung reaction, a neglected microbiologic technique. *Mt*  
507 *Sinai J Med* 43:699-709.

- 508 13. Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK, PneuCarriage project g.  
509 2015. The PneuCarriage Project: A Multi-Centre Comparative Study to Identify the  
510 Best Serotyping Methods for Examining Pneumococcal Carriage in Vaccine  
511 Evaluation Studies. *PLoS Med* 12:e1001903; discussion e1001903.
- 512 14. Swarthout TD, Fronterre C, Lourenco J, Obolski U, Gori A, Bar-Zeev N, Everett D,  
513 Kamng'ona AW, Mwalukomo TS, Mataya AA, Mwansambo C, Banda M, Gupta S,  
514 Diggle P, French N, Heyderman RS. 2020. High residual carriage of vaccine-serotype  
515 *Streptococcus pneumoniae* after introduction of pneumococcal conjugate vaccine in  
516 Malawi. *Nat Commun* 11:2222.
- 517 15. Satzke C, Turner P, Virolainen-Julkunen A, Adrian PV, Antonio M, Hare KM, Henao-  
518 Restrepo AM, Leach AJ, Klugman KP, Porter BD, Sa-Leao R, Scott JA, Nohynek H,  
519 O'Brien KL, Group WHOPCW. 2013. Standard method for detecting upper respiratory  
520 carriage of *Streptococcus pneumoniae*: updated recommendations from the World  
521 Health Organization Pneumococcal Carriage Working Group. *Vaccine* 32:165-79.
- 522 16. Newton R, Hinds J, Wernisch L. 2011. Empirical Bayesian models for analysing  
523 molecular serotyping microarrays. *BMC Bioinformatics* 12:88.
- 524 17. Hinds J GK, Witney AA, Baldry SJ, Lambertsen L, Hannage WP, Antonio M, Turner P,  
525 Newton R, Wernisch L, Aanensen DM, Bentley S. 2009. Molecular serotyping of  
526 *Streptococcus pneumoniae*: a microarray-based tool with enhanced utility for isolate  
527 typing, novel serotype discovery, non-typeable investigation, multiple carriage  
528 detection, and direct analysis of nasopharyngeal swabs, abstr 9th European Meeting  
529 on the Molecular Biology of the Pneumococcus (Europneumo), Bern, Switzerland
- 530 18. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK.  
531 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development,

- 532 evaluation and verification of targets for serogroup and serotype prediction using an  
533 automated pipeline. PeerJ 4:e2477.
- 534 19. O'Brien KL, Nohynek H, World Health Organization Pneumococcal Vaccine Trials  
535 Carraige Working G. 2003. Report from a WHO working group: standard method for  
536 detecting upper respiratory carriage of *Streptococcus pneumoniae*. *Pediatr Infect Dis*  
537 *J* 22:133-40.
- 538 20. Gratten M, Montgomery J, Gerega G, Gratten H, Siwi H, Poli A, Koki G. 1989. Multiple  
539 colonization of the upper respiratory tract of Papua New Guinea children with  
540 *Haemophilus influenzae* and *Streptococcus pneumoniae*. *Southeast Asian J Trop*  
541 *Med Public Health* 20:501-9.
- 542 21. Hill PC, Cheung YB, Akisanya A, Sankareh K, Lahai G, Greenwood BM, Adegbola RA.  
543 2008. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a  
544 longitudinal study. *Clin Infect Dis* 46:807-14.
- 545 22. Ortika BD, Habib M, Dunne EM, Porter BD, Satzke C. 2013. Production of latex  
546 agglutination reagents for pneumococcal serotyping. *BMC Res Notes* 6:49.
- 547 23. Epping L, van Tonder AJ, Gladstone RA, The Global Pneumococcal Sequencing C,  
548 Bentley SD, Page AJ, Keane JA. 2018. SeroBA: rapid high-throughput serotyping of  
549 *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genom* 4.
- 550 24. Carvalho Mda G, Pimenta FC, Moura I, Roundtree A, Gertz RE, Jr., Li Z, Jagero G,  
551 Bigogo G, Junghae M, Conklin L, Feikin DR, Breiman RF, Whitney CG, Beall BW. 2013.  
552 Non-pneumococcal mitis-group streptococci confound detection of pneumococcal  
553 capsular serotype-specific loci in upper respiratory tract. *PeerJ* 1:e97.
- 554 25. Li L, Mendis N, Trigui H, Oliver JD, Faucher SP. 2014. The importance of the viable but  
555 non-culturable state in human bacterial pathogens. *Front Microbiol* 5:258.

- 556 26. Swarthout TD, Gori A, Bar-Zeev N, Kamng'ona AW, Mwalukomo TS, Bonomali F,  
557 Nyirenda R, Brown C, Msefula J, Everett D, Mwansambo C, Gould K, Hinds J,  
558 Heyderman RS, French N. 2020. Study data and dictionary - evaluation of  
559 pneumococcal serotyping methods, Blantyre, Malawi, 2015-2017  
560 doi:10.6084/m9.figshare.12807872.

561

562



563 **Figure legends**

564 **Figure 1** Sample selection for analysis. Sample were collected during 4 rolling cross-  
565 sectional surveys from June 2015 to April 2017. From the total nasopharyngeal swab  
566 samples collected 1347 samples had results available from the three assays under  
567 review. The method of selection for microarray and PneumoCaT was done  
568 independent of available latex serotype data.

569

570 **Figure 2** Laboratory procedures. Nasopharyngeal swabs (NPS) were inoculated into  
571 STGG and subsequently plated on a growth agar of sheep blood and gentamycin.  
572 Bacteria growth (from single-colony picks) from samples culture-confirmed for  
573 *Streptococcus pneumoniae* were used for latex serotyping. Remaining pure-growth  
574 isolates, retained at -80°C in sterile STGG, were later grown for DNA extraction and  
575 WGS. Aliquots of original samples (NPS-STGG) that were culture-confirmed for  
576 *Streptococcus pneumoniae* were assessed by microarray. NPS=nasopharyngeal  
577 swabs, STGG=skim-milk-tryptone-glucose-glycerol, WGS=whole-genome  
578 sequencing, Spn=*Streptococcus pneumoniae*, SBG=sheep blood and gentamycin,  
579 SSI= Statens Serum Institute, 13v=13-valent, NSP-STGG=NPS inoculated into  
580 STGG.

581

582 **Figure 3** Concordance between assays. Concordance between two assays was  
583 defined as both assays identifying pneumococcal carriage as VT or both as NVT.  
584 Latex and PneumoCaT reported one result per sample, both using the same pure-  
585 growth culture. Microarray, using an aliquot of the original NPS-STGG, differentiated  
586 individual serotypes in multiple serotype carriage, when present. When comparing  
587 the three assays, concordance was based on serotype if latex reported VT carriage.

588 If latex reported NVT, this was considered concordant to any NVT reported by  
589 PneumoCaT and microarray, as long as PneumoCaT and microarray reported the  
590 same NVT.

591

592 **Figure 4** Proportion of individual VT serotypes contributing to total VT carriage. The  
593 proportion of individual VT serotypes detected is not significantly different when  
594 comparing microarray to latex. Refer to Table S1 in the supplemental material for the  
595 reported frequency of each vaccine serotype (VT) detected by microarray and latex.  
596 VT=vaccine serotype.

597

#### 598 **Notes**

599 **Financial Support.** Bill & Melinda Gates Foundation, USA (OPP117653). A project  
600 grant jointly funded by the UK Medical Research Council (MRC) and the UK  
601 Department for International Development (DFID) under the MRC/DFID Concordat  
602 agreement, also as part of the EDCTP2 programme supported by the European  
603 Union (MR/N023129/1); and a recruitment award from the Wellcome Trust (Grant  
604 106846/Z/15/Z). The MLW Clinical Research Programme is supported by a Strategic  
605 Award from the Wellcome Trust, UK (206545/Z/17/Z). National Institute for Health  
606 Research (NIHR) Global Health Research Unit on Mucosal Pathogens using UK aid  
607 from the UK Government (Project number 16/136/46).

608

609 **Disclaimer.** The funders had no role in study design, collection, analysis, data  
610 interpretation, writing of the report or in the decision to submit the paper for  
611 publication. The corresponding author had full access to the study data and, together  
612 with the senior authors, had final responsibility for the decision to submit for

613 publication. TDS, AG, RSH, and NF are supported by the National Institute for  
 614 Health Research (NIHR) Global Health Research Unit on Mucosal Pathogens using  
 615 UK aid from the UK Government. The views expressed in this publication are those  
 616 of the author(s) and not necessarily those of the NIHR or the Department of Health  
 617 and Social Care.

618

619 **Potential conflicts of interest.** Dr. Bar-Zeev reports investigator-initiated research  
 620 grants from GlaxoSmithKline Biologicals and from Takeda Pharmaceuticals outside  
 621 the submitted work. Dr. Jason Hinds is an investigator on studies undertaken on  
 622 behalf of St George's, University of London or BUGS Bioscience that are sponsored  
 623 or funded by vaccine manufacturers, including Pfizer, GlaxoSmithKline and Sanofi  
 624 Pasteur. Dr. Jason Hinds is also co-founder and shareholder of BUGS Bioscience, a  
 625 not-for-profit spin-out of St George's, University of London. No other authors declare  
 626 competing interests.

627

628 **Table 1** Increased detection of VT carriage, latex vs microarray

	Latex VT prevalence (n) 95% CI	Microarray VT prevalence (n) 95% CI	% Increase in VT prevalence
Children 3-6 years, PCV-vaccinated (n=1360)	20.0% (272) 17.9, 22.2	28.6% (389) 26.2, 31.1	43.0%
Children 5-10 years, PCV-unvaccinated (n=904)	21.1% (191) 18.5, 23.9	26.5% (240) 23.7, 29.6	21.7%
Adults, 18-40 years, HIV-infected, PCV-unvaccinated (n=963)	14.2% (137) 12.1, 16.6	16.6% (160) 14.3, 19.1	10.8%
Total (n=3227)	18.6 (600) 17.3, 20.0	24.4 (789) 23.0, 26.0	31.5%

629 PCV=pneumococcal conjugate vaccine, VT=vaccine serotype, CI=confidence

630 interval.

631 **Table 2** Key comparative parameters of serotyping methods.  
632

	Latex (phenotypic)	Microarray (genomic)	PneumoCaT (genomic)
<b>Assay implementation</b>			
Sample used in assay	• Pure growth from single isolate	• Original sample in STGG	• Pure growth from single isolate
Cost estimate <sup>1</sup>	• Lowest of three assays	• Highest of three assays	• Middle of three assays
Implementation of assay	• Least difficult (relatively simple)	• Most difficult	• Moderate difficult
Training required for implementation	• Minimal	• Advanced	• DNA extraction: moderate • WGS library manipulation: advanced • PneumoCaT tool: moderate
Training required for processing and interpretation of results	• Minimal	• Moderate	• Moderate
<b>Assay output and interpretation</b>			
Serotypes reported	• Single	• Multiple, if present	• Single
NVT differentiation	• No <sup>3</sup>	• Yes	• Yes
Relative abundance of individual serotypes reported	• No	• Yes	• No
Additional outputs	• Isolates archived and available for further analyses	• AMR profile <sup>2</sup> • NT differentiation	• WGS library accessible for further analyses, including population structure and AMR
Conclusion	• Adequate for surveillance • Limited resolution for optimal VE estimation	• Cost and technique limits ability to de-centralise implementation • Detection of VT in low relative abundance is of critical importance • Sentinel sites should be considered for regional NVT & VT resolution for optimal VE estimation	• Limited resolution for optimal VE estimation • No benefit over latex unless WGS library already available

633 <sup>1</sup>Estimated costs and feasibility of implementation and maintenance are specific to the study requirements and laboratory capacity  
634 (including no capacity for WGS or microarray) at the Malawi-Liverpool-Wellcome Trust Clinical Research Programme in Blantyre.

635 <sup>2</sup> AMR profile cannot be assigned to a single strain in a sample with multiple-serotype or multiple-pathogen carriage.

636 <sup>3</sup>NVT & NT reported as NVT. Use of both commercial products and production of in-house latex serotyping reagents can  
637 significantly increase the number of NVT serotypes that can be differentiated by latex.











