Title:

Full: An empirical Bayes method for serotype case-carrier ratios, with an application to Group B streptococcus

Running head: Epidemiologic properties of GBS serotypes

Authors:

Joseph A. Lewnard[1,*], Lauren A. Cowley[2]

Affiliations:

1. Division of Epidemiology & Biostatistics, School of Public Health, University of California, Berkeley, California, United States
2. Milner Center for Evolution, Department of Biology & Biochemistry, University of Bath, Bath, United Kingdom

* *Corresponding author:*
Joseph Lewnard, PhD
Division of Epidemiology & Biostatistics
School of Public Health
University of California, Berkeley
2121 Berkeley Way #5412
Berkeley, California 94720
jlewnard@berkeley.edu
+1 (510) 664-4050

Conflicts of interest:

JAL declares receipt of research grants from Pfizer to the University of California, Berkeley and to Harvard University for unrelated work.

Acknowledgements:

The authors thank Dr. Ruth Lynfield for helpful comments on an earlier version of the manuscript.

ABSTRACT

Background: Case-carrier ratios quantifying the relative pathogenicity of serotypes can inform vaccine formulations for antigenically-diverse pathogens. However, sparse serotype-specific counts in epidemiologic datasets may undermine such analyses, most notably for rare serotypes that pose emergence risks in vaccinated populations. This challenge is well-illustrated in Group B streptococcus (GBS), where serotype III dominates in both carriage and disease.

Methods: We develop an empirical Bayes random-effects model based on conjugate Dirichlet-multinomial distributions of serotype frequencies in carriage and disease states. We validate the model using simulated datasets, and apply it to data from 15 paired sets of GBS isolates from intrapartum recto-vaginal colonization (n=3403) and neonatal invasive disease (NID; n=1088), 16 from blood (n=2352) and cerebrospinal fluid (n=780) neonatal specimens, and 3 from fatal (n=173) and non-fatal (n=1684) neonatal invasive infections.

Results: Our method accurately recovers parameters in simulated datasets. Using this approach, we confirm that GBS serotype III exhibits the greatest invasiveness, followed by serotype Ia with a 75.3% (95%CrI: 43.7-93.8%) lower estimate. Enhanced invasiveness of serotypes III and Ia is most evident in late-onset disease. Non–hexavalent-vaccine serotypes, which are rare in carriage and disease, generally show lower invasiveness; serotype IX/non-typeable GBS, the most prevalent cause of non–vaccine-preventable disease, is 98.7% (81.7-99.9%) and 94.2% (13.9-99.6%) less invasive than serotypes III and Ia, respectively.

Conclusions: We present a strategy for measuring associations of serotype with carrier and disease states in the presence of sparse counts, avoiding biases that exist in common ad-hoc approaches.

Keywords: Group B Streptococcus; *Streptococcus agalactiae*; vaccine; serotype; Empirical Bayes; meta-analysis

INTRODUCTION

Group B streptococcus (GBS; *Streptococcus agalactiae*) is the leading cause of neonatal invasive infections, with 0.53 cases per 1000 live-births globally[1] and 2.03 cases per 1000 live-births in sub-saharan Africa.[2] While intrapartum antimicrobial prophylaxis prevents early-onset GBS disease (EOD; occurring 0-6 days after birth),[3] additional preventive strategies are needed to combat late-onset disease (LOD; 7-90 days), and to serve populations with limited access to antenatal care. There are ten recognized serotypes of GBS defined by distinct polysaccharide capsular antigens (Ia, Ib, II-IX) as well as non-typeable strains (NT-GBS). Anticapsular conjugate vaccines intended for pregnant women are currently under development. A trivalent conjugate vaccine (GlaxoSmithKline) targeting serotypes Ia, Ib, and III has completed phase-2 trials,[4,5] and a hexavalent vaccine (Pfizer) against serotypes Ia, Ib, and II-V is entering human studies.

Considerations around GBS vaccine compositions have centered around serotype frequencies in neonatal invasive disease, with serotypes Ia, Ib, and III alone comprising 85% of cases worldwide.[6] However, concerns surround the possibility that vaccine-derived immunity against targeted GBS serotypes may exert selective pressure on the circulating GBS population, enabling non-vaccine serotypes that are currently rare to expand in prevalence.[7] Such "serotype replacement" in *Streptococcus pneumoniae*—whereby serotypes not targeted by pneumococcal conjugate vaccines (PCVs) have overtaken the ecological niche formerly inhabited by vaccine-targeted serotypes—has offset the impact of PCVs against invasive pneumococcal disease.[8] Monitoring for serotype replacement has also been central to postlicensure surveillance of the impact of vaccines against human papillomavirus,[9] meningococcus,[10] and *Haemophilus influenzae* type b.[11]

Estimates of the relative capacity of serotypes to cause infections, severe clinical manifestations, or death provide a basis for forecasting the impact of vaccines on the epidemiology of antigenically-diverse pathogens. Experience with PCVs has illustrated that these estimates can lead to better vaccine compositions than historical approaches,[12] such as the selection of the most prevalent serotypes as vaccine targets,[13,14] under the scenario of serotype replacement. Such estimates also provide a baseline for monitoring post-vaccination epidemiological changes.[15,16] While historical data suggest enhanced invasiveness of GBS serotype III relative to serotypes Ia, Ib, and II, these studies pre-date the emergence or discovery of serotypes IV-IX and nontypeable GBS (NT-GBS) lineages, and are of limited size, statistical rigor, and geographic scope.[17–19] Thus, quantifications of the relative pathogenicity of GBS serotypes are necessary to assess the potential impact of anticapsular vaccine candidates.

The rare occurrence of certain GBS serotypes in maternal carriage and neonatal disease, however, makes it difficult to estimate these quantities with certainty, even in large samples.[20,21] While random-effects meta-analysis of data from multiple epidemiologic studies may thus present an attractive analysis strategy,[22] investigators remain likely to encounter challenges when generating pooled risk ratios or odds ratios in the presence of zero-valued counts within studies. Common ad-hoc workarounds, such as "continuity corrections" adding a fixed value to cells[23–25] and simple summing or averaging of values across studies,[13] risk introducing bias in both point estimates and measures of uncertainty.[26] While many studies exclude analyses of rare serotypes to avoid such problems,[27–31] rare serotypes may in fact be of interest for serotype replacement; moreover, zero-valued cell counts within studies present valuable information on overall serotype prevalence. Here, we develop an empirical Bayes method to infer serotype distribution parameters in pooled analyses of sparse datasets, providing a basis for estimating case-carrier ratios. We apply this approach to estimate the association of GBS serotype with invasive potential, risk of presenting as meningitis, and risk of death in neonatal infections.

METHODS

Outcome definition

We aimed to measure the association of serotype with the following epidemiologic features of neonatal invasive GBS:

1. Risk of causing invasive disease given maternal colonization: We defined invasive disease as GBS isolation from an ordinarily sterile site (e.g., blood or cerebrospinal fluid), distinguishing between EOD and LOD;
2. Risk for meningitis during neonatal invasive infection: We defined cases with GBS isolation from cerebrospinal fluid as meningitis; and
3. Risk for fatal outcome of neonatal invasive infection: We defined fatal cases as deaths attributable to invasive GBS among live-born infants (thus excluding stillbirths).

For analyses of invasive potential, we compared serotype frequencies in neonatal disease against serotype frequencies in recto-vaginal colonization among pregnant women; we did not restrict colonization samples according to time of sampling within pregnancy due to limited metadata on when swabs were obtained.

Study identification

We updated previously-undertaken systematic literature reviews[1,6,] and extracted serotype-stratified data on maternal carriage, disease timing (EOD/LOD), disease manifestation (meningitis or other), and outcome (fatal or non-fatal). For analyses of serotype associations with neonatal invasive potential, we aimed to include all studies monitoring both carriage and disease serotype distributions within a single source population or setting; for analyses of serotype associations with meningeal invasion and fatal outcome, we sought all studies presenting serotype frequencies in neonatal meningitis and sepsis, and in fatal and non-fatal neonatal invasive disease. We examined full texts of all studies included in the previous reviews to determine eligibility, and conducted a PubMed search as well as citation tracking of these reviews and two other relevant studies[20,21] to identify new articles for screening. We detail the search strategy and inclusion/exclusion criteria in the **Supplementary information**.

Statistical framework: overview

From each study with paired carriage and disease data, we estimated the relative odds of a serotype being isolated from neonatal invasive disease (all neonatal invasive disease, EOD, LOD) versus maternal colonization; we derive this measure on the risk scale below. From studies reporting GBS isolation from blood or CSF, we estimated the relative risk that invasive disease caused by each serotype would manifest as meningitis. Last, for studies reporting fatal and non-fatal outcome, we estimated the risk ratio for fatal outcome (case-fatality ratio; CFR), given invasive disease, for each serotype.

We pooled these measures of association across studies under an empirical Bayes framework, described below in technical detail. Briefly, we took serotype frequencies in carriage, all neonatal invasive disease, EOD, LOD, meningitis, sepsis, non-fatal disease, and fatal disease, within each study, to be multinomially-distributed. We took the parameters of each setting-specific multinomial distribution, in turn, to be random draws from a global Dirichlet conjugate prior distribution. This enabled us to use data from all studies in estimating Dirichlet hyperparameters describing global serotype frequencies in GBS carriage and disease, and thus to generate positive-valued estimates of setting-specific multinomial probabilities in the presence of zero-valued counts. In turn, we used these multinomial probabilities to generate pooled measures of association under a random effects model.

Empirical Bayes random effects model

The invasive potential of serotype *j*, measured relative to serotype *k*, from serotype frequencies in carriage and disease ($\mathbf{x}^C$ and $\mathbf{x}^D$, respectively), is typically estimated as

$$IP(j,k) = \frac{x^D(j)x^C(k)}{x^D(k)x^C(j)} = \frac{\left(\frac{x^D(j)}{x^C(j)}\right)}{\left(\frac{x^D(k)}{x^C(k)}\right)}.$$

We derive this in a later section ("**Measures of association**"). Challenges arise when $\mathbf{x}^C$ and $\mathbf{x}^D$ include zero-valued counts. We propose the following solution: Within the ℓth study, take observed serotype-specific frequencies in carriage and disease (vectors $\mathbf{x}_\ell^C$ and $\mathbf{x}_\ell^D$, respectively) to be random draws from a multinomial distribution, specified in full by the number of carriage and disease isolates obtained (n_ℓ^C and n_ℓ^D, respectively) and the true underlying proportions of carriage and disease (vectors $\mathbf{p}_\ell^C$ and $\mathbf{p}_\ell^D$, respectively) attributable to each serotype within the populations from which cases and carriers were sampled:

$$\mathbf{x}_\ell^C \sim \text{Multinom}\left(n_\ell^C, \mathbf{p}_\ell^C\right)$$

and

$$\mathbf{x}_\ell^D \sim \text{Multinom}(n_\ell^D, \mathbf{p}_\ell^D).$$

The Dirichlet conjugate prior for the population proportion parameters, $\mathbf{p}_\ell^C$ and $\mathbf{p}_\ell^D$, defines these study-level proportions as random deviates of global serotype distributions with parameters $\boldsymbol{\alpha}^C$ and $\boldsymbol{\alpha}^D$:

$$\mathbf{p}_\ell^C \sim \text{Dir}(\boldsymbol{\alpha}^C)$$

and

$$\mathbf{p}_\ell^D \sim \text{Dir}(\boldsymbol{\alpha}^D).$$

This yields positive estimates of $\mathbf{p}_\ell^C$ and $\mathbf{p}_\ell^D$ across the included studies. We verified the identifiability of the parameters in a simulation study described below.

Define study-level estimates of neonatal invasive potential using the inferred population proportions of carriage and disease attributable to each serotype. Within the ℓth study,

$$IP_\ell(j,k) = \frac{p_\ell^D(j)p_\ell^C(k)}{p_\ell^D(k)p_\ell^C(j)}.$$

We calculate invasive potential relative to serotype III (k), taking measures from each study to be deviates from a global serotype-specific value, $\theta(j,k)$, with associated variance across studies $\sigma^2(j,k)$:

$$\log IP_\ell(j,k) \sim \text{N}\left(\log\theta(j,k), \sigma^2(j,k)\right).$$

Estimation

We sampled from the joint distributions $f(\boldsymbol{\alpha}^C, \mathbf{p}_1^C, \mathbf{p}_2^C, \dots, \mathbf{p}_g^C | \mathbf{x}_1^C, \mathbf{x}_2^C, \dots \mathbf{x}_g^C)$ and $f(\boldsymbol{\alpha}^D, \mathbf{p}_1^D, \dots, \mathbf{p}_g^D | \mathbf{x}_1^D, \dots \mathbf{x}_g^D)$ via Markov chain Monte Carlo (MCMC) using a Metropolis-Hastings updating procedure. We ran chains for 600,000 iterations after the burn-in period, saving the state of the chain at every 500th sample to reduce autocorrelation.

We estimated the global invasiveness parameters according to

$$\Pr\left(\theta(j,k), \sigma(j,k) | \mathbf{IP}(j,k)\right) \propto \prod_\ell f(\log IP_\ell(j,k) | \theta(j,k), \sigma^2(j,k)),$$

which provided a basis for sampling from the joint distribution of $\{\theta(j,k), \sigma^2(j,k)\}$ for each serotype j, relative to serotype III (k).

Measures of association

Odds ratio estimator for neonatal invasive potential

Define the neonatal invasive potential (IP) of serotype j relative to serotype k as the ratio

$$IP(j,k) = \frac{\pi(j)}{\pi(k)}$$

of the probabilities $\pi(j)$ and $\pi(k)$ for neonates exposed to maternal carriage of serotypes *j* and *k*, respectively, to experience invasive disease, such that

$$p^D(j) = \pi(j)p^C(j)$$
$$p^D(k) = \pi(k)p^C(k)$$

for the variables $\mathbf{p}^D$ and $\mathbf{p}^C$ defined above. We note that π represents the joint probability of both GBS transmission from mother (or other source) to neonate and of disease progression given transmission. Thus,

$$IP(j,k) = \frac{x^D(j)x^C(k)}{x^C(j)x^D(k)} = \frac{\left(\frac{x^D(j)}{x^C(j)}\right)}{\left(\frac{x^D(k)}{x^C(k)}\right)} = \frac{\left(\frac{p^D(j)N^D}{p^C(j)N^C}\right)}{\left(\frac{p^D(k)N^D}{p^C(k)N^C}\right)} = \frac{\pi(j)}{\pi(k)}.$$

Risk ratios for meningitis and fatal outcome

The conventional relative risk estimator for invasive disease cause by serotype *j* to manifest as meningitis, as compared to invasive disease caused by serotype *k*, is

$$RR^M(j,k) = \frac{x^M(j)x^D(k)}{x^D(j)x^M(k)} = \frac{\left(\frac{x^M(j)}{x^D(j)}\right)}{\left(\frac{x^M(k)}{x^D(k)}\right)}$$

for the vectors of serotype frequencies in meningitis ($\mathbf{x}^M$), here specifying that $\mathbf{x}^D$ includes cases with known clinical manifestations. Similarly, the relative risk for invasive disease caused by serotype *j* to result in fatality, as compared to invasive disease caused by serotype *k*, is conventionally measured as

$$RR^F(j,k) = \frac{x^F(j)x^D(k)}{x^D(j)x^F(k)} = \frac{\left(\frac{x^F(j)}{x^D(j)}\right)}{\left(\frac{x^F(k)}{x^D(k)}\right)}$$

where $\mathbf{x}^F$ indicates serotype frequencies in fatal cases of illness, while $\mathbf{x}^D$ includes cases with known clinical outcomes (fatal or nonfatal). Consistent with the approach taken for invasive potential, we sampled from the distributions of serotype-specific proportions in all meningitis cases ($\mathbf{p}_{\boldsymbol{\ell}}^M$) and fatal invasive disease cases ($\mathbf{p}_{\boldsymbol{\ell}}^F$) via

$$\mathbf{x}_\ell^M \sim \text{Multinom}(n_\ell^M, \mathbf{p}_{\boldsymbol{\ell}}^M)$$

$$\mathbf{x}_\ell^F \sim \text{Multinom}(n_\ell^F, \mathbf{p}_{\boldsymbol{\ell}}^F)$$

$$\mathbf{p}_\ell^M \sim \text{Dir}(\boldsymbol{\alpha}^M)$$

$$\mathbf{p}_\ell^F \sim \text{Dir}(\boldsymbol{\alpha}^F),$$

where n_ℓ^M and n_ℓ^F indicate the total number of meningitis isolates and fatal case isolates, respectively, in the ℓth study, and $\boldsymbol{\alpha}^M$ and $\boldsymbol{\alpha}^F$ serve as hyperparameters for the underlying distributions. We define global

estimates for serotype-specific relative risks from the normal distribution of log-transformed estimates from each study as above:

$$RR_{\ell}^{M}(j,k) = \frac{p_{\ell}^{M}(j)p_{\ell}^{D}(k)}{p_{\ell}^{M}(k)p_{\ell}^{D}(j)}$$

$$RR_{\ell}^{F}(j,k) = \frac{p_{\ell}^{F}(j)p_{\ell}^{D}(k)}{p_{\ell}^{F}(k)p_{\ell}^{D}(j)},$$

and use the same framework to estimate $\theta(j,k)$ and $\sigma(j,k)$ as we describe above for the analysis of invasive potential.

Presentation

For all analyses, we report relative measures (odds ratios and risk ratios) using the serotype with the greatest absolute odds or risk of the endpoint as the referent category.

Simulation study

We conducted a simulation study to verify identifiability of the $\mathbf{p}_{\boldsymbol{\ell}}$ parameters, defined above, that we used as inputs to calculation of case-carrier ratios. We generated simulated distributions of 10 serotypes from the Dirichlet conjugate prior distribution. We considered various parameterizations to allow scenarios of differing serotype diversity, defining $\alpha(j)$ as $\exp(-\lambda k(j))b$, with **k** comprising randomly-ordered values of 1, 2, ..., 10. We considered $\lambda \in (0.1, 0.2, \ldots, 0.6)$ to allow differing degrees of serotype dominance, and $b \in (1, 5, 10)$ to allow differing degrees of between-study variance in serotype-specific prevalence among isolates. From each of the resulting Dirichlet parameterizations for $\boldsymbol{\alpha}$, we sampled twenty vectors $\mathbf{p}_{\boldsymbol{\ell}}$ for differing study settings, and generated observations

$$\mathbf{x}_{\boldsymbol{\ell}} \sim \mathrm{Multinom}(n_{\ell}, \mathbf{p}_{\boldsymbol{\ell}})$$

with sample sizes n_{ℓ} distributed uniformly, on the log scale, between 10 and 1000. We used the simulated data $\mathbf{x}_{\boldsymbol{\ell}}$ to estimate $\mathbf{p}_{\boldsymbol{\ell}}$ and $\boldsymbol{\alpha}$ via Markov chain Monte Carlo sampling, consistent with the approach described above. We calculated parameter coverage probabilities at the 2.5% credible threshold (an analog to statistical power for continuous measurements, indicating the probability that true values of these parameters were contained within 95% credible intervals of the estimates).

RESULTS

Model performance on simulated data

We illustrate the proportion of counts expected to take on values of zero under differing values of *b* and λ in **Figure 1A-C**, for sample sizes between 10 and 1000 isolates. Zero-valued counts are most probable in small samples, at low values of *b* (corresponding to greater variance across settings in multinomial serotype prevalence), and at high values of λ (corresponding to a situation where one or several serotypes is consistently dominant among carriage or disease isolates).

We identified 80% or greater probability for true estimates of setting-specific serotype prevalences ($\mathbf{p}_{\boldsymbol{\ell}}$) to fall within 95% credible intervals of estimates across the range of values explored for *b* and λ (**Figure 1D**); these coverage probabilities were poorest under the circumstances that would lead to the greatest probability for zero-valued counts as described above (low *b*, high λ). Simulated coverage probabilities for $\mathbf{p}_{\boldsymbol{\ell}}$ fell within the range of 80.5-91.5% for *b*=1, and 90.5-95.5% for *b*=10. A similar trend toward better coverage probabilities with higher *b* and lower λ was evident in estimates of the hyperparameter $\boldsymbol{\alpha}$ (**Figure 1E**); lower accuracy in estimates of $\boldsymbol{\alpha}$ (with coverage probabilities reaching as low as 40%) did not prevent the model from recovering valid estimates of the target parameters $\mathbf{p}_{\boldsymbol{\ell}}$.

Data abstraction

We assessed full texts of 599 studies covered in previous systematic reviews, and screened abstracts of 50 articles identified via a PubMed search and 529 articles citing the previous reviews and two other relevant articles, not exclusive of duplicates (**Figure 2**). From these, we identified 15 paired datasets of GBS serotype distributions in carriage and neonatal invasive disease (3403 and 1088 isolates, respectively; **Table 1**); 16 datasets presenting serotype distributions in meningitis and other neonatal invasive disease (3132 isolates; **Table 2**); and 3 datasets presenting serotypes isolated from fatal and non-fatal neonatal invasive disease cases (1684 isolates; **Table 2**).

Of the paired maternal carriage and neonatal disease datasets, two were obtained from maternal-newborn dyads; all others were from geographically- and temporally-matched samples of maternal colonization and neonatal invasive disease isolates. Latin American (1 study, 75 isolates), East Asian (1 study, 101 isolates), and South/Southeast Asian (0 studies) populations were under-represented; nine datasets were from North America, Europe, and Oceania. Studies reporting disease manifestation and fatal outcome were from North America, Europe, Oceania, East Asia, and sub-Saharan Africa; no data from Latin America or South/Southeast Asia were available.

Relative neonatal invasive potential across GBS serotypes

Serotype III caused 48-95% of neonatal invasive disease cases across the included studies, followed by serotype Ia (1-28%), serotype V (0-15%), and serotype II (0-10%). This hierarchy of serotype frequencies was less evident in colonization (**Figure 3**). Serotypes Ia, Ib, and II-VII and IX/NT-GBS were each associated with lower neonatal invasive potential in comparison to serotype III; because no studies reported serotype VIII in disease, we did not generate pooled estimates for this serotype. Relative to serotype III, we estimated serotypes Ia and V were 24.7% (6.2-56.3%) and 5.6% (0.9%-21.8%) as likely, respectively, to cause neonatal invasive disease given maternal carriage. These pooled estimates were consistent with findings of individual studies; nine and eleven studies, respectively, provided evidence of lower neonatal invasive potential of serotypes Ia and V (relative to serotype III) at the 97.5% confidence threshold. We estimated the neonatal invasive potential for serotypes Ib, II, and IV—the remaining serotypes included in the candidate hexavalent vaccine—to be 3.9% (0.4-17.4%), 2.9% (0.4-11.6%), and 8.6% (1.0-70.5%) that of serotype III, respectively, and 12.2% (1.6-70.5%), 16.1% (1.5-97.5%), and 36.8% (3.7–459.5%) that of serotype Ia.

Serotype VI caused only 0-2% of neonatal invasive disease within studies. We identified 87.9% posterior probability that serotype VI was associated with reduced serotype potential relative to serotype III, but found its invasiveness to be statistically indistinguishable from that of serotype Ia at levels of statistical power afforded by our meta-analysis; we identified ≥80% probability that serotype VI invasiveness exceeded that of vaccine-targeted serotypes Ib and II. We estimated the neonatal invasive potential of serotype VII to be 4.5% (0.3-105.1%) that of serotype III. Serotype IX/NT-GBS was associated with the lowest neonatal invasive potential, at 1.3% (<0.1-18.3%) that of serotype III and 5.8% (0.4-86.1%) that of serotype Ia; moreover, we identified ≥80% posterior probability that this reflected lower neonatal invasive potential than each serotype except Ib, II, and VII.

Serotype differences in invasive potential were less apparent in EOD than in LOD. Point estimates of EOD potential again suggested exceptional invasiveness for serotype III; we estimated the EOD potential of serotypes Ia and IV to be 67.0% (38.7-113.3%) and 23.5% (4.5-186.8%) that of serotype III, respectively, whereas these serotypes showed 14.6% (9.0-25.9%) and 3.3% (1.2-9.1%) the LOD potential of serotype III. For serotypes Ib, II, and V, we estimated EOD potentials to be 16.5% (4.3-54.9%), 16.5% (5.1-44.9%), and 34.1% (16.5-68.7%) as high as serotype III, respectively, whereas the LOD potentials of these serotypes were 0.8% (0.4-1.5%), 0.5% (0.3-1.0%), and 0.4% (0.2-0.9%) that of serotype III. Low serotype-specific frequencies in both carriage and disease (beneath the threshold at which our simulation study predicted reliable coverage) prohibited us from distinguishing EOD and LOD potentials for the non-vaccine serotypes.

Relative risks of meningitis and death across GBS serotypes

Meningitis occurred in 10–77% of serotype III neonatal infections across the included studies; as compared to serotype III, serotypes Ia, Ib, II, and V showed 10-13% as high a risk of meningitis, given invasive disease (**Figure 4**). Serotypes IV, VI, and IX/NT-GBS were isolated in a lower proportion of studies, and caused meningitis in 0-100% of invasive disease episodes within these studies. Pooled estimates suggested they were each less likely than serotype III to manifest as meningitis in invasive disease cases, albeit with weaker precision as compared to estimates for serotypes Ia, Ib, II and V. Although we do not stratify estimates by EOD/LOD, we present data in **Figure 4** from studies that defined serotype frequencies within these strata.

Risk of death showed lower degrees of variation among serotypes, with no estimated differences meeting conventional thresholds of statistical significance (**Figure 5**). Point estimates suggested the highest CFR for neonatal invasive disease caused by serotype Ia, similar to the estimtes for serotype V and VI with relative risks of 90.5% (15.7-376.2%) and 77.9% (5.0-902.5%), respectively. We estimated the CFR for disease caused by serotype III to be 46.1% (5.6-271.8%) that of serotype Ia; point estimates for the relative risk of death given invasive infection were in the range of 31-59% for the remaining vaccine-targeted serotypes. No fatal cases of disease caused by serotypes VII, VIII, and IX/NT-GBS were reported in the original studies.

DISCUSSION

Because vaccines targeting antigenically-diverse pathogens may facilitate expansion of non-vaccine serotypes, assessments of serotype-specific pathogenicity provide key insight into the potential impacts of serotype replacement on disease burden. We present a method that circumvents key limitations to such analyses. These include the need to limit such analyses to only the most prevalent serotypes due to zero-valued counts of serotype frequencies in datasets, and to implement ad-hoc continuity corrections that bias measures of statistical association and uncertainty.[26] Because non-vaccine serotypes posing emergence risks in vaccinated populations may be uncommon in the pre-vaccine era, unbiased quantification of rare serotype parameters is of particular importance.[12] Our empirical Bayes approach uses estimates of serotype-specific prevalence across included studies to generate a “global” Dirichlet conjugate prior for serotype prevalence among GBS isolates in study populations, treating the observed serotype frequencies within each setting, in turn, as independent multinomial draws.

Applying this method to data from a systematic review of GBS carriage and disease, our analysis confirms longstanding clinical insight that GBS serotypes III and Ia pose the greatest risk of neonatal disease when carried by mothers, and that neonatal invasive infections caused by serotype III are the most likely to present as meningitis.[17–19] Because this evidence preceded discovery of now-prevalent serotypes, the risks of serotype replacement under differing vaccine formulations have to date been uncertain.[7] The enhanced association of serotypes III and Ia with invasive potential in the later neonatal period underscores the potential for GBS vaccination to reduce the burden of LOD, which cannot be prevented by screening-based intrapartum antimicrobial prophylaxis. We found serotypes VII and IX/NT-GBS to pose relatively limited risk of disease when carried by mothers. Although serotype VI has not historically been a prevalent cause of disease, its invasiveness appears on par with that of hexavalent vaccine serotypes other than III, suggesting serotype VI may be important to monitor in postlicensure surveillance. Taken together, our findings suggest non-vaccine types are unlikely to offset vaccine-preventable neonatal disease burden, even if they replace vaccine-targeted serotypes in carriage.

It is important to note that anticapsular vaccines may have impacts on GBS population structure beyond serotype replacement. The emergence of virulent vaccine-escape lineages by recombination constitutes an important threat foreshadowed by evolutionary responses of *S. pneumoniae* to PCVs.[32,33] A majority of recombination events in GBS has involved genes encoding surface protein antigens and capsular biosynthesis proteins, suggesting immune escape via horizontal gene transfer.[34] The acquisition of serotype IV capsule by clonal complex (CC) 17 lineages historically belonging to serotype III illustrates the capacity for invasive lineages to escape immunity via capsular switching. Expression of ≥7 distinct serotypes in CC1 isolates obtained from surveillance in Toronto underscores the frequency of serotype-

switching events in GBS, and the need to understand the contribution of noncapsular antigens to epidemiologic parameters such as invasiveness.[35–37] While our model can be extended to incorporate both clone and serotype as determinants of epidemiologic properties, molecular typing information in GBS is only beginning to become routinely available.[38]

The method we propose has value for analyses of multiple antigenically-diverse pathogens beyond GBS. Assessments of serotype replacement risk in pneumococcal disease remain important as vaccine manufacturers develop extended-valency anticapsular vaccines for *S. pneumoniae*. Increasing attention to the impact of PCVs on non-invasive pneumonia also underscores the need for determinations of the role of serotype in such disease.[39] The method is also applicable to analyses of differential pathogenicity of *emm* types in group A Streptococcus, for which vaccines are currently in development;[40,41] knowledge of *emm* type associations with carriage and disease states remains limited.[42,43] Similarly, our approach may provide a strategy for understanding epithelial tropisms of the many human papillomavirus genotypes, and their role in cancer progression.[44]

Several features of our method can be extended. While not available for this analysis, prior distributions of both serotype frequencies and measures of pathogenicity can be incorporated into the calculation of posterior probabilities. For instance, phenotypic aspects of *S. pneumoniae* such as capsule width, molecular content, and charge predict serotype associations with carriage prevalence as well as risk of invasive infection,[45] death,[25] and otitis media;[15,46] estimates of such associations could provide prior distributions on prevalence or invasiveness parameters for rare serotypes. Whereas we have assumed statistically-independent distributions of serotypes in carriage and disease within each setting, the model could also be extended to account for Dirichlet prior distributions of serotypes by geographic setting in addition to carriage or disease status.

In conclusion, we have provided a flexible empirical Bayes method for the estimation of serotype associations with carriage and disease states in meta-analyses. This method removes the need for ad-hoc continuity corrections that are commonly undertaken in such analyses due to sparse observations. Simulation studies reveal the method performs well under realistic conditions, providing 80% or greater coverage probability for multinomial probabilities from which we estimate measures of association. Applying this analytic framework, we identify low pathogenicity of non-vaccine GBS serotypes relative to serotype III, suggesting limited risk for vaccine-driven serotype replacement to offset the burden of vaccine-preventable GBS infections in neonates.

REFERENCES

1. Edmond KM, Kortsalioudaki C, Scott S, et al. Group B streptococcal disease in infants aged younger than 3 months: Systematic review and meta-analysis. *Lancet*. 2012;379(9815):547-556. doi:10.1016/S0140-6736(11)61651-6.

2. Sinha A, Russell LB, Tomczyk S, et al. Disease Burden of Group B streptococcus among infants in Sub-Saharan Africa: A systematic literature review and meta-analysis. *Pediatr Infect Dis J*. 2016;35(9):933-942. doi:10.1097/INF.0000000000001233.

3. Moore MR, Schrag SJ, Schuchat A. Effects of intrapartum antimicrobial prophylaxis for prevention of Group B streptococcal disease on the incidence and ecology of early-onset neonatal sepsis. *Lancet Infect Dis*. 2003. doi:10.1016/S1473-3099(03)00577-2.

4. Heyderman RS, Madhi SA, French N, et al. Group B streptococcus vaccination in pregnant women with or without HIV in Africa: A non-randomised phase 2, open-label, multicentre trial. *Lancet Infect Dis*. 2016;16(5):546-555. doi:10.1016/S1473-3099(15)00484-3.

5. Madhi SA, Cutland CL, Jose L, et al. Safety and immunogenicity of an investigational maternal trivalent Group B streptococcus vaccine in healthy women and their infants: a randomised phase 1b/2 trial. *Lancet Infect Dis*. 2016;16(8):923-934. doi:10.1016/S1473-3099(16)00152-3.

6. Madrid L, Seale AC, Kohli-Lynch M, et al. Infant Group B streptococcal disease incidence and serotypes worldwide: Systematic review and meta-analyses. *Clin Infect Dis*. 2017;65:S160-S172. doi:10.1093/cid/cix656.

7. Vekemans J, Moorthy V, Friede M, et al. Maternal immunization against Group B streptococcus: World Health Organization research and development technological roadmap and preferred product characteristics. *Vaccine*. 2018.

8. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. *Lancet*. 2011;378:1962-1973. doi:10.1016/S0140-6736(10)62225-8.

9. Tota JE, Ramanakumar A V, Jiang M, et al. Epidemiologic approaches to evaluating the potential for human papillomavirus type replacement postvaccination. *Am J Epidemiol*. 2013. doi:10.1093/aje/kwt018.

10. Maiden MCJ, Spratt BG. Meningococcal conjugate vaccines: New opportunities and new challenges. *Lancet*. 1999. doi:10.1016/S0140-6736(99)00252-4.

11. Adam HJ, Richardson SE, Jamieson FB, Rawte P, Low DE, Fisman DN. Changing epidemiology of invasive *Haemophilus influenzae* in Ontario, Canada: Evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine*. 2010. doi:10.1016/j.vaccine.2010.03.075.

12. Nurhonen M, Auranen K. Optimal serotype compositions for pneumococcal conjugate vaccination under serotype replacement. *PLoS Comput Biol*. 2014;10(2). doi:10.1371/journal.pcbi.1003477.

13. Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serogroups cause the most invasive disease: Implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis*. 2000;30(1):100-121. doi:10.1086/313608.

14. Sniadack DH, Schwartz B, Lipman H, et al. Potential interventions for the prevention of childhood pneumonia: geographic and temporal differences in serotype and serogroup distribution of sterile site pneumococcal isolates from children—implications for vaccine strategies. *Pediatr Infect Dis J*. 1995. doi:10.1097/00006454-199506000-00007.

15. Lewnard JA, Givon-Lavi N, Weinberger DM, Lipsitch M, Dagan R. Pan-serotype reduction in progression of *Streptococcus pneumoniae* to otitis media after rollout of pneumococcal conjugate vaccines. *Clin Infect Dis*. July 2017. doi:10.1093/cid/cix673.

16. Southern J, Andrews N, Sandu P, et al. Pneumococcal carriage in children and their household contacts six years after introduction of the 13-valent pneumococcal conjugate vaccine in England. *PLoS One*. 2018;16:1-14.

17. Baker CJ, Barrett FF, Gordon RC, Yow MD. Suppurative meningitis due to streptococci of lancefield group B: A study of 33 infants. *J Pediatr*. 1973;82(4):724-729. doi:10.1016/S0022-3476(73)80606-7.

18. Baker CJ, Barrett FF. Transmission of Group B streptococci among parturient women and their neonates. *J Pediatr*. 1973;83(6):919-925. doi:10.1016/S0022-3476(73)80524-4.

19. Wilkinson HW, Facklam RR, Wortham EC. Distribution by serological type of Group B streptococci isolated from a variety of clinical material over a five year period (with special reference to neonatal sepsis and meningitis). *Infect Immun*. 1973;8(2):228-235.

20. Madzivhandila M, Adrian P V., Cutland CL, Kuwanda L, Schrag SJ, Madhi SA. Serotype distribution and invasive potential of Group B streptococcus isolates causing disease in infants and colonizing maternal-newborn dyads. *PLoS One*. 2011;6(3).

doi:10.1371/journal.pone.0017861.

21. Seale AC, Koech AC, Sheppard AE, et al. Maternal colonization with *Streptococcus agalactiae* and associated stillbirth and neonatal disease in coastal Kenya. *Nat Microbiol*. 2016;1(7). doi:10.1038/nmicrobiol.2016.67.

22. Balsells E, Dagan R, Yildirim I, et al. The relative invasive disease potential of *Streptococcus pneumoniae* among children after PCV introduction: A systematic review and meta-analysis. *Journal of Infection*. 2018.

23. Feikin DR, Kagucia EW, Loo JD, et al. Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: A pooled analysis of multiple surveillance sites. *PLoS Med*. 2013. doi:10.1371/journal.pmed.1001517.

24. Flasche S, Givon-Lavi N, Dagan R. Using pneumococcal carriage data to monitor postvaccination changes in the incidence of pneumococcal otitis media. *Am J Epidemiol*. 2016;184(9):652-659. doi:10.1093/aje/kww012.

25. Weinberger DM, Harboe ZB, Sanders E a M, et al. Association of serotype with risk of death due to pneumococcal pneumonia: a meta-analysis. *Clin Infect Dis*. 2010;51(6):692-699. doi:10.1086/655828.

26. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med*. 2004. doi:10.1002/sim.1761.

27. Brueggemann AB, Peto TEA, Crook DW, Butler JC, Kristinsson KG, Spratt BG. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. Journal of Infectious Diseases. doi:10.1086/423820.

28. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis*. 2003;187(9):1424-1432. doi:10.1086/374624.

29. Sleeman KL, Griffiths D, Shackley F, et al. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *J Infect Dis*. 2006;194(5):682-688. doi:10.1086/505710.

30. Shouval DS, Greenberg D, Givon-Lavi N, Porat N, Dagan R. Site-specific disease potential of individual *Streptococcus pneumoniae* serotypes in pediatric invasive disease, acute otitis media and acute conjunctivitis. *Pediatr Infect Dis J*. 2006;25(7):602-607. doi:10.1097/01.inf.0000220231.79968.f6.

31. Hanage WP, Auranen K, Syrjänen R, et al. Ability of pneumococcal serotypes and clones to cause acute otitis media: Implications for the prevention of otitis media by conjugate vaccines. *Infect Immun*. 2004;72(1):76-81. doi:10.1128/IAI.72.1.76-81.2004.

32. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog*. 2007;3(11):1628-1636. doi:10.1371/journal.ppat.0030168.

33. Croucher NJ, Harris SR, Fraser C, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011;331(6016):430-434. doi:10.1126/science.1198545.

34. 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.

doi:10.1038/ncomms5544.

35. Neemuchwala A, Teatero S, Athey TBT, McGeer A, Fittipaldi N. Capsular switching and other large-Scale recombination events in invasive sequence type 1 Group B streptococcus. *Emerg Infect Dis*. 2016;22(11):1941-1944. doi:10.3201//eid2211.152064.

36. Teatero S, Ramoutar E, McGeer A, et al. Clonal complex 17 Group B streptococcus strains causing invasive disease in neonates and adults originate from the same genetic pool. *Sci Rep*. 2016;6. doi:10.1038/srep20047.

37. Teatero S, McGeer A, Li A, et al. Population structure and antimicrobial resistance of invasive serotype IV Group b streptococcus, Toronto, Ontario, Canada. *Emerg Infect Dis*. 2015;21(4):585-591. doi:10.3201/eid2104.140759.

38. Metcalf BJ, Chochua S, Gertz RE, et al. Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive *Streptococcus agalactiae* recovered in the USA. *Clin Microbiol Infect*. 2017;23(8):574.e7-574.e14. doi:10.1016/j.cmi.2017.02.021.

39. Levine OS, O'Brien KL, Deloria-Knoll M, et al. The pneumonia etiology research for child health project: A 21st century childhood pneumonia etiology study. *Clin Infect Dis*. 2012;54(SUPPL. 2). doi:10.1093/cid/cir1052.

40. Dale JB, Penfound TA, Chiang EY, Walton WJ. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine*. 2011;29(46):8175-8178. doi:10.1016/j.vaccine.2011.09.005.

41. McNeil SA, Halperin SA, Langley JM, et al. Safety and immunogenicity of 26-valent Group A streptococcus vaccine in healthy adult volunteers. *Clin Infect Dis*. 2005;41(8):1114-1122. doi:10.1086/444458.

42. O'Loughlin RE, Roberson A, Cieslak PR, et al. The epidemiology of invasive Group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin Infect Dis*. 2007;45(7):853-862. doi:10.1086/521264.

43. Tartof SY, Reis JN, Andrade AN, Ramos RT, Reis MG, Riley LW. Factors associated with Group A streptococcus *emm* type diversification in a large urban setting in Brazil: A cross-sectional study. *BMC Infect Dis*. 2010;10. doi:10.1186/1471-2334-10-327.

44. Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I. Human papillomavirus molecular biology and disease association. *Rev Med Virol*. 2015;25(S1):2-23. doi:10.1002/rmv.1822.

45. Weinberger DM, Trzciński K, Lu Y-J, et al. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog*. 2009;5(6):e1000476. doi:10.1371/journal.ppat.1000476.

46. Lewnard JA, Givon-Lavi N, Tähtinen PA, Dagan R. Pneumococcal phenotype and interaction with nontypeable *Haemophilus influenzae* as determinants of otitis media progression. *Infect Immun*. 2018;86(6). doi:10.1128/IAI.00727-17.

Table 1: Included studies indicating serotype association with invasive potential.

Study	Reference (Supplement)	Setting	Basis for pairing	Isolates obtained			
				Carriage	Neonatal disease	EOD	LOD
Ko et al., 2015	47	Australia	Geography, date	408	108	60	48
Seale et al., 2016	48	Kenya	Mother-infant dyads	915	70	30	40
van Elzakker et al., 2009	49	Taiwan	Geography, date	58	43	– –	– –
van Elzakker et al., 2009	49	Netherlands	Geography, date	92	44	– –	– –
Madzivhandila et al., 2011	50	South Africa	Mother-infant dyads	541	245	137	108
Martins et al., 2007	51	Portugal	Geography, date	269	64	42	22
Bisharat et al., 2005	52	Israel	Geography, date	104	50	– –	– –
Berg et al., 2000	53	Sweden	Geography, date	114	128	40	10
Moraleda et al., 2018	54	Morocco	Geography, date	67	5	– –	– –
Lopez et al., 2018	55	Spain	Facility, date	242	79	– –	– –
Figueira-Coelho et al., 2004	56	Portugal	Facility, date	108	17	– –	– –
Davies et al., 2001	57	Canada	Geography, date	118	– –	75	– –
Teatero et al., 2014	58	Canada	Facility, date	102	93	34	59
Martinez et al., 2004	59	Chile	Geography, date	65	10	– –	– –
Hayes et al., 2017	60	Ireland	Geography, date	200	7	– –	– –

Table 2: Included studies indicating serotype association with disease manifestation and fatal outcome.

Study	Reference (Supplement)	Setting	Isolates obtained		
			Meningitis	Non-meningitis	Fatal outcome
Davies et al., 2001	57	Canada	17	74	91
Ekelund et al., 2004	61	Denmark	107	405	– –
Gray et al., 2007	62	Malawi	21	25	– –
Zhao et al., 2008	63	Australia, New Zealand	20	73	– –
Martins et al., 2007	51	Portugal	4	55	– –
Trijbels-Smeulders et al., 2006	64	Netherlands	63	135	– –
Fluegge et al., 2005	65	Germany	67	225	– –
Chang et al., 2014	66	Japan	37	21	– –
Liu et al., 2015	67	Chiu	3	5	– –
Morozumi et al., 2014	68	Japan	81	48	– –
Six et al., 2016	69	France	4	12	– –
Teatero et al., 2014	58	Canada	9	80	– –
Yoon et al., 2015	70	Republic of Korea	25	24	– –
Siguaque et al., 2018	71	Mozambique	14	21	– –
Dangor et al., 2016	72	South Africa	172	424	596
Phares et al., 2008	73	United States	156	841	997

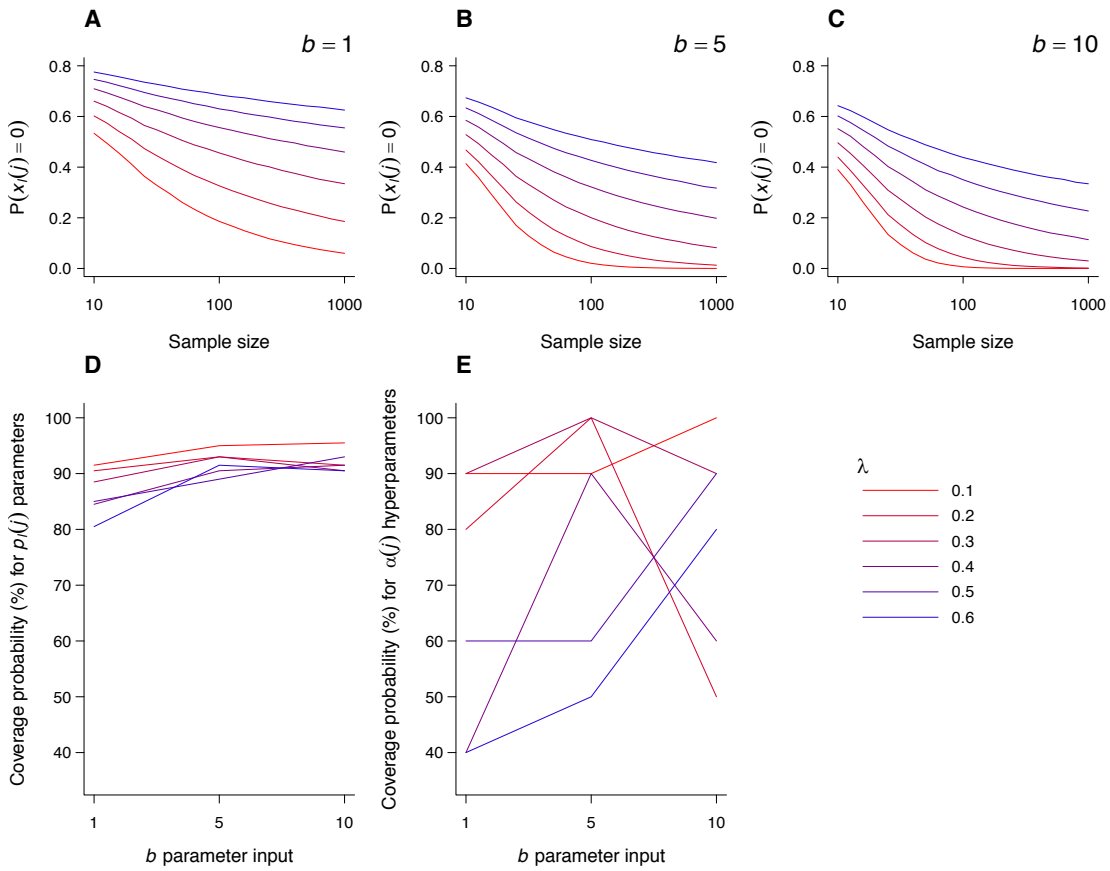


Figure 1: Assessment of parameter identifiability in simulated data. A-C) To illustrate the range of epidemiologic situations covered by our simulation study, we illustrate the probability for serotype-specific counts in individual studies ($x_\ell(j)$ for count x of serotype j in study ℓ) to take on values of zero under differing sample sizes (x-axis) and values of the parameters b and λ. Higher values of b correspond to lower variance across settings in the prevalence of a given serotype prevalence among GBS isolates. Higher values of λ correspond to a scenario where a given serotype (or serotypes) occupies a greater proportion of all isolates, i.e. reduced Simpson diversity among serotypes. D) We identify 80% or greater coverage (probability for 95% credible intervals to contain the true value) for estimates of $\mathbf{p}_\ell$, the setting-level prevalence of serotypes among all isolates, under the conditions explored. Estimates approach 95% at increasing values of b and decreasing values of λ. E) We identify poorer coverage for estimates of the hyperparameter $\boldsymbol{\alpha}$, which also tend improve under the conditions that lead to more accurate estimation of $\mathbf{p}_\ell$; our method does not apply the hyperparameter toward estimates of epidemiologic association.

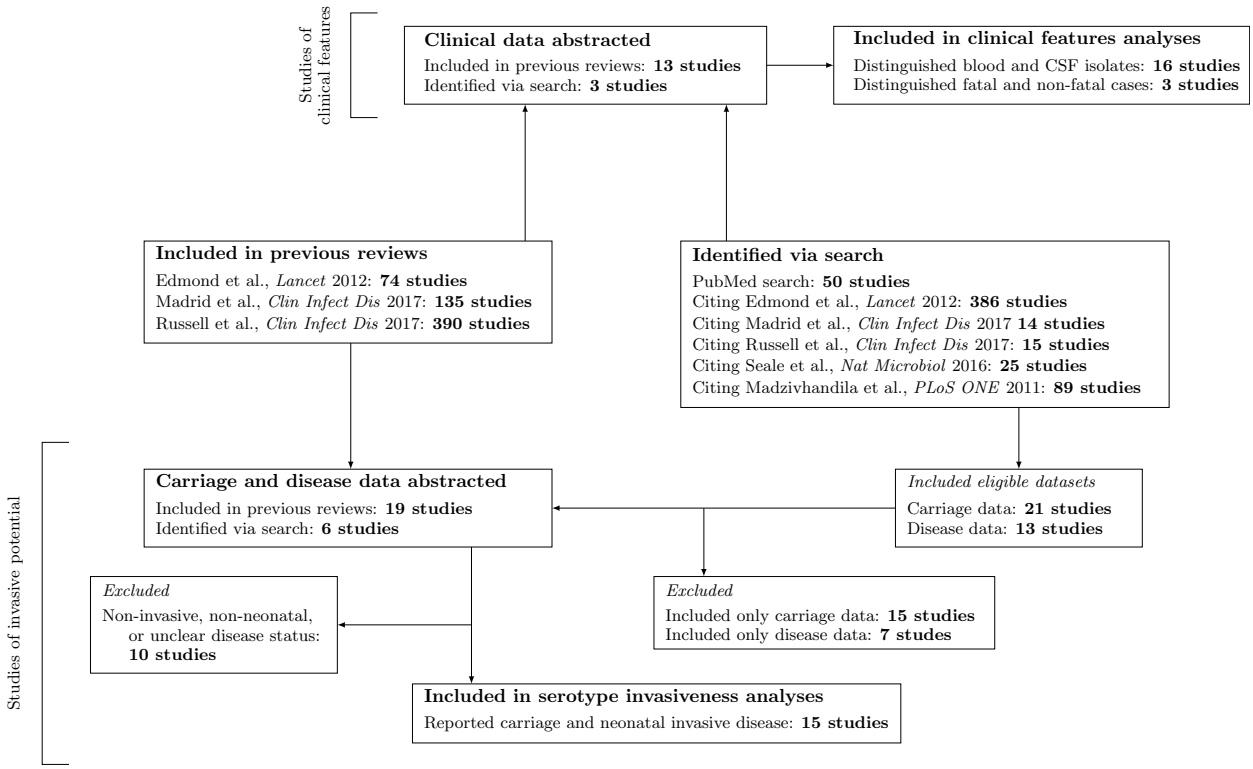
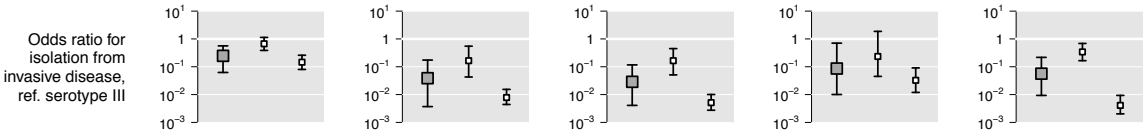


Figure 2: Systematic review flowchart. We detail the identification and inclusion/exclusion of articles from previous systematic reviews and articles as well as their citations, and a PubMed search for articles published after January 1, 2016. Tabulations do not exclude duplicate results.

Study	Serotype Ia				Serotype Ib				Serotype II				Serotype IV				Serotype V			
	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD
Hayes et al., 2017	59	1			20	0			31	1			8	0			23	1		
Martinez et al., 2004	20	3			3	0			14	0			0	0			1	0		
Teatero et al., 2017	24	13	5	8	12	6	5	1	13	4	3	1	6	1	0	1	20	8	5	3
Davies et al., 2001	24		20		10		4		10		6		0		0		33		13	
Figueira–Coelho et al., 2004	21	5			5	0			14	1			1	2			26	3		
Lopez et al., 2018	36	18			20	9			55	10			21	2			26	9		
Moraleda et al., 2018	6	0			0	0			17	2			5	1			24	1		
Bergal et al., 2000	15	19	5	0	15	4	1	1	13	5	1	0	3	3	0	1	25	18	9	2
Misharat et al., 2005	12	10			9	2			23	4			0	0			18	9		
Martins et al., 2000	42	18	13	5	14	2	1	1	46	8	8	0	6	2	1	1	59	7	7	0
Madzivhandila et al., 2011	163	46	31	15	36	7	7	0	61	7	7	0	20	5	5	0	55	10	8	2
van Elzakker et al., 2009	24	6			6	0			12	0			7	0			14	4		
van Elzakker et al., 2009	13	5			5	2			2	0			0	0			15	2		
Seale et al., 2016	194	10	7	3	114	4	3	1	80	1	1	0	18	1	1	0	156	0	0	0
Ko et al., 2015	107	26	18	8	31	6	4	2	40	5	4	1	5	2	2	0	68	18	12	6



Study	Serotype VI				Serotype VII				Serotype VIII				Serotype IX/NT				Serotype III (ref.)			
	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD	Car.	Dis.	EOD	LOD
Hayes et al., 2017	2	1			4	0			1	0			1	0			51	3		
Martinez et al., 2004		0				0				0				0			24	7		
Teatero et al., 2017	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	26	59	15	44
Davies et al., 2001	0				0				0				23				18		28	
Figueira–Coelho et al., 2004	0	0			1	0			1	0			11	0			28	6		
Lopez et al., 2018	0	0			2	0			0	0			0	0			82	31		
Moraleda et al., 2018	0	0			0	0			0	0			3	0			12	1		
Bergal et al., 2000	0	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0	36	78	24	6
Misharat et al., 2005	0	1			0	0			0	0			16	5			26	19		
Martins et al., 2000	0	0	0	0	5	0	0	0	0	0	0	0	38	1	0	1	59	26	12	14
Madzivhandila et al., 2011	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	202	170	79	91
van Elzakker et al., 2009	2	0			0				0				2				20	33		
van Elzakker et al., 2009	2	0			0	0			0	0			2	0			19	34		
Seale et al., 2016	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	350	54	18	36
Ko et al., 2015	3	4	4	0	3	1	1	0	0	0	0	0	9	0	0	0	142	46	15	31

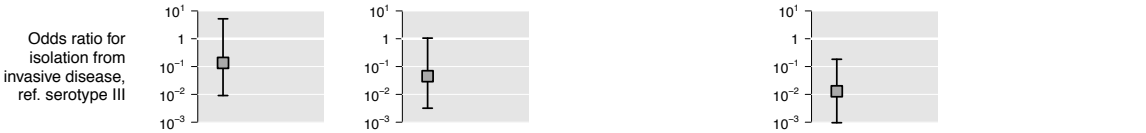
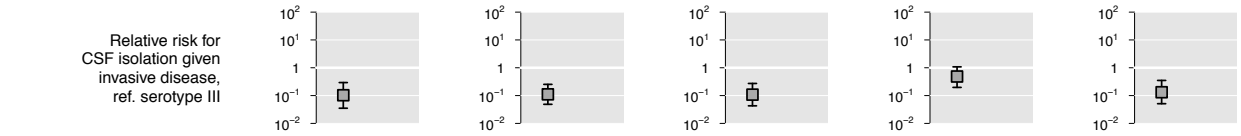


Figure 3: Association of serotype with neonatal invasive potential. For each study, we indicate serotype frequencies in intrapartum recto-vaginal colonization ("Car.") and neonatal invasive disease ("Dis."), as well as EOD and LOD, as permitted by reporting within original studies. We plot pooled estimates of serotype-specific relative invasive potential beneath serotype frequencies for all disease (grey points) as well as EOD and LOD (white points), measured from the odds ratio that an isolate came from invasive disease versus colonization, relative to serotype III (lower right corner). Our derivation of this measure is provided in the **Supporting information**. Pooled estimates include data from studies reporting only partial selections of serotypes, where total numbers of included isolates provide a basis for sampling from proportions attributable to each serotype. Bars signify 95% credible intervals around pooled estimates.

Study	Serotype Ia						Serotype Ib						Serotype II						Serotype IV						Serotype V					
	All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis		
	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD
Phares et al., 2008	268	157	111	41	12	29	68	37	31	9	2	7	78	69	9	3	1	2					0		158	93	65	13	3	10
Dangor et al., 2016	172	123	49	34	8	26	21	16	5	1	0	1	21	19	2	1	1	0	5	5	0	0	0	0	47	39	8	2	1	1
Sigauque et al., 2018	1			0			0			0			0			0			0			0			1			0		
Yoon et al., 2015	6			4			6			3			0			0			0			0			13			5		
Teatero et al., 2014	12	5	7	2	0	2	6	5	1	0	0	0	4	3	1	0	0	0	1	0	1	0	0	0	6	3	3	0	0	0
Six et al., 2016	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Morozumi et al., 2014	29	8	21	12	3	9	12	4	8	7	1	6	5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Liu et al., 2015	2	1	1	0			3	1	2	1			0	0	0	0			0	0	0	0			1	0	1	1		
Chang et al., 2014	18	4	14	12	3	9	5	1	4	4	1	3	2	1	1	1	1	0	2	1	1	1	0	1	1	1	0	0	0	0
Fluegge et al., 2005	43	29	14	5	1	4	13	8	5	1	0	1	16	13	3	1	0	1	2	2	0	0	0	0	22	15	7	4	1	3
Trijbels–Smeulders et al., 2006	37	30	7	10	6	4	6	5	1	0	0	0	16	15	1	0	0	0					0		14	13	1	2	1	1
Martins et al., 2007	18			1			2			0			8			1			2			0			7			0		
Zhao et al., 2008			25			2			14			3			1			0			1			0			3			1
Gray et al., 2007	10	7	3	4	2	2	3	2	1	1	1	0	2	2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Ekelund et al., 2004	95	70	25	20	10	10	37	29	8	5	2	3	29	26	3	0	0	0	5	3	2	2	1	1	21	19	2	3	3	0
Davies et al., 2001	14	13	1	2	2	0	4	2	2	0	0	0	2	2	0	0	0	0			0	0	0	0	13	9	4	1	1	0



Study	Serotype VI						Serotype VII						Serotype VIII						Serotype IX/NT						Serotype III (ref.)					
	All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis			All inv. dis.			Meningitis		
	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD	All	EOD	LOD
Phares et al., 2008					0						0						0						0		385	146	239	85	17	68
Dangor et al., 2016	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	329	139	190	133	20	113
Sigauque et al., 2018	0			0			0			0			0			0			0			0			33			14		
Yoon et al., 2015	1			0			0			0			0			0			0			0			23			13		
Teatero et al., 2014	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59	15	44	7	1	6
Six et al., 2016	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	3	9	4	1	3
Morozumi et al., 2014	4	3	1	2	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	5	73	60	4	56
Liu et al., 2015	0	0	0	0			0	0	0	0			0	0	0	0			0	0	0	0			2	0	2	1		
Chang et al., 2014	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	28	5	23	19	3	16
Fluegge et al., 2005			0	0	0	0			0	0	0	0			0	0	0	0			0	0	0	0	193	98	95	56	9	47
Trijbels–Smeulders et al., 2006					0						0						0		14	10	4	1	0	1	109	68	41	49	22	27
Martins et al., 2007	0			0			0			0			0			0			1			0			21			2		
Zhao et al., 2008			0			0			0			0			0			0			0			0			49			14
Gray et al., 2007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	30	13	17	14	5	9
Ekelund et al., 2004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	28	3	0	0	0	294	184	110	77	38	39
Davies et al., 2001			0	0	0	0			0	0	0	0			0	0	0	0			1	0	0	0	55	20	35	14	5	9

Figure 4: Association of serotype with meningeal invasion. For each study, we indicate serotype frequencies in all neonatal disease and in meningitis (signified by isolation from CSF), including all neonatal disease as well as EOD and LOD, as permitted by reporting within original studies. We plot pooled estimates of serotype-specific relative risk of meningitis beneath serotype frequencies for all disease (grey points), measured from the relative risk that an isolate came from CSF, measured against serotype III (lower right corner), for all infections. Our derivation of this measure is provided in the **Supporting information**. Pooled estimates include data from studies reporting only partial selections of serotypes, where total numbers of included isolates provide a basis for sampling from proportions attributable to each serotype. Bars signify 95% credible intervals around pooled estimates.

	Davies et al. (2001)		Dangor et al. (2016)		Phares et al. (2008)		Relative risk of fatal outcome, ref. serotype Ia
	All	Fatal	All	Fatal	All	Fatal	
Serotype Ia (ref.)							
All dis.	14	6	172	31	268	18	
EOD	13	1	123	21	157	10	
LOD	1	0	49	10	111	8	
Serotype Ib							
All dis.	4	3	21	2	68	3	
EOD	2	1	16	1	37	2	
LOD	2	0	5	1	31	1	
Serotype II							
All dis.	2	1	21	4	78	3	
EOD	2	0	19	3	69	3	
LOD	0	0	2	1	9	0	
Serotype III							
All dis.	55	5	329	56	385	11	
EOD	20	2	139	28	146	6	
LOD	35	0	190	28	239	5	
Serotype IV							
All dis.	--	0	5	1	--	--	
EOD	--	0	5	1	--	--	
LOD	0	0	0	0	--	0	
Serotype V							
All dis.	13	4	47	10	158	13	
EOD	9	1	39	9	93	8	
LOD	4	1	8	1	65	5	
Serotype VI							
All dis.	--	0	1	1	--	--	
EOD	--	0	0	0	--	--	
LOD	0	0	1	1	--	0	
Serotype VII							
All dis.	--	0	0	0	--	--	
EOD	--	0	0	0	--	--	
LOD	0	0	0	0	--	0	
Serotype VIII							
All dis.	--	0	0	0	--	--	
EOD	--	0	0	0	--	--	
LOD	0	0	0	0	--	0	
Serotype IX/NT							
All dis.	--	0	0	0	--	--	
EOD	--	0	0	0	--	--	
LOD	1	0	0	0	--	0	

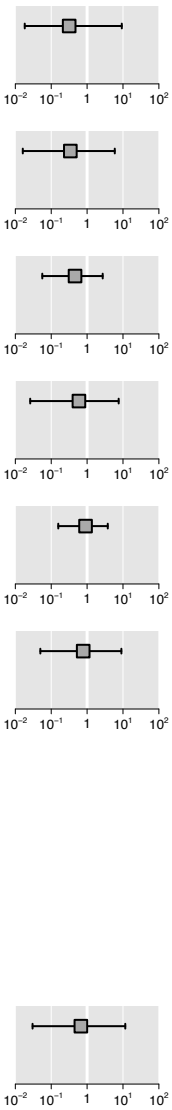


Figure 5: Association of serotype with risk of fatal outcome. For each study, we indicate serotype frequencies in all neonatal invasive disease and in fatal cases, including totals as well as EOD and LOD. We plot pooled estimates of serotype-specific relative risk of death given invasive infection, measured against serotype Ia, for all infections. Pooled estimates include data from studies reporting only partial selections of serotypes, where total numbers of included isolates provide a basis for sampling from proportions attributable to each serotype. Bars signify 95% credible intervals around pooled estimates.

SEARCH STRATEGY

We searched the following string in titles and abstracts of articles indexed in PubMed between 1 January, 2016 and 22 April, 2018 to identify relevant studies not included in previous reviews:

(("group b streptococc") OR ("streptococcus agalactiae")) AND ((coloni* OR carria* OR carrie* OR carry*) AND (diseas* OR infect* OR invas* OR mening* OR sepsis OR septic*)) AND (serotyp* OR capsul* OR cps*)*

We included studies that:

1. Were written in English, Spanish, German, or French.

2. Conducted serotyping by molecular or phenotypic assays for, at minimum, serotypes Ia, Ib, II, III, IV, and V, and reported the total number of isolates tested; this led to the exclusion of studies before 1999. We recorded the total number of isolates tested and the numbers identified as serotypes Ia, Ib, II, III, IV, V, VI, VII, VIII, and IX or NT-GBS; we pooled serotype IX and NT-GBS to accommodate data from studies undertaken before identification of serotype IX.

3. Described the population, setting, and study design: We excluded studies selecting patients based on microbiological findings post-hoc (e.g. studies limited to disease caused by a pre-defined sets clones or serotypes), and studies within which it was not possible to associate isolates with the specific colonization or disease outcomes under consideration (e.g. studies that did not specify whether patients were symptomatic or asymptomatic).

For analyses of neonatal invasive potential of serotypes, we included studies presenting “paired” sets of carriage and disease isolates, defined as isolates obtained from epidemiologically-linked individuals (i.e., mother-newborn dyads), or from geographically- and temporally-matched samples of neonatal disease cases and asymptomatically-colonized pregnant women. For analyses of serotype associations with disease manifestation and clinical outcome, we included studies presenting frequencies of individual serotypes in samples of all neonatal sepsis and meningitis cases, and in samples of all fatal and non-fatal neonatal invasive disease cases.

When studies presented data from multiple settings or populations, we abstracted data at the finest demographic and geographic resolution permitted by available metadata. We excluded articles presenting duplicate data from the same original study.

SUPPLEMENTAL REFERENCES

The references listed here comprise all articles included in the review, matching the reference numbers listed in **Table 1** and **Table 2**.

47. Ko DWH, Zurynski Y, Gilbert GL. Group B Streptococcal disease and genotypes in Australian infants. *J Paediatr Child Health*. 2015;51(8):808–814.

48. Seale AC, Koech AC, Sheppard AE, et al. Maternal colonization with *Streptococcus agalactiae* and associated stillbirth and neonatal disease in coastal Kenya. *Nat Microbio*l. 2016;1:16067.

49. Elzakker E Van, Yahiaoui R, Visser C, et al. Epidemiology of and prenatal molecular distinction between invasive and colonizing group B Streptococci in the Netherlands and Taiwan. *Eur J Clin Microbiol Infect Dis*. 2009;28(8):921–928.

50. Madzivhandila M, Adrian P V., Cutland CL, Kuwanda L, Schrag SJ, Madhi SA. Serotype distribution and invasive potential of group B Streptococcus isolates causing disease in infants and colonizing maternal-newborn dyads. *PLoS One*. 2011;6(3):e17861.

51. Martins ER, Pessanha MA, Ramirez M, et al. Analysis of group B streptococcal isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness. *J Clin Microbiol*. 2007;45(10):3224–3229.

52. Bisharat N, Jones N, Marchaim D, et al. Population structure of group B Streptococcus from a low-incidence region for invasive neonatal disease. *Microbiology*. 2005;151(6):1875–1881.

53. Berg S, Trollfors B, Lagergård T, Zackrisson G, Claesson BA. Serotypes and clinical manifestations of group B Streptococcal infections in western Sweden. *Clin Microbiol Infect*. 2000;6(1):9–13.

54. Moraleda C, Benmessaoud R, Esteban J, et al. Prevalence, antimicrobial resistance and serotype distribution of group B Streptococcus isolated among pregnant women and newborns in Rabat, Morocco. *J Med Microbiol*. 2018;doi:10.1099/jmm.0.000720.

55. López Y, Parra E, Cepas V, et al. Serotype, virulence profile, antimicrobial resistance and macrolide-resistance determinants in *Streptococcus agalactiae* isolates in pregnant women and neonates in Catalonia, Spain. *Enferm Infecc Microbiol Clin*. 2017;doi:10.1016/j.eimc.2017.08.006.

56. Figueira-Coelho J, Ramirez M, Salgado MJ, Melo-Cristino J. *Streptococcus agalactiae* in a large Portuguese teaching hospital: antimicrobial susceptibility, serotype distribution, and clonal analysis of macrolide-resistant isolates. *Microb Drug Resist*. 2004;10(1):31–36.

57. Davies HD, Adair C, McGeer A, et al. Antibodies to capsular polysaccharides of group B Streptococcus in pregnant Canadian women: relationship to colonization status and infection in the neonate. *J Infect Dis*. 2001;184(3):285–291.

58. Teatero S, McGeer A, Low DE, et al. Characterization of invasive group B Streptococcus strains from the greater Toronto area, Canada. *J Clin Microbiol*. 2014;52(5):1441–1447.

59. Martinez MA, Ovalle A, Duran C, et al. Serotipos y susceptibilidad antimicrob *Streptococcus agalactiae*. *Rev med Chile*. 2004;132(5):549–555.

60. Hayes K, Cotter L, Barry L, O'Halloran F. Emergence of the L phenotype in Group B Streptococci in the South of Ireland. *Epidemiol Infect*. 2017;145(16):3535–3542.

61. Ekelund K, Konradsen HB. Invasive group B Streptococcal disease in infants: A 19-year nationwide study. Serotype distribution, incidence and recurrent infection. *Epidemiol Infect*. 2004;132(6):1083–1090.

62. Gray KJ, Bennett SL, French N, Phiri AJ, Graham SM. Invasive group B Streptococcal infection in infants, Malawi. *Emerg Infect Dis*. 2007;13(2):223–229.

63. Zhao Z, Kong F, Zeng X, Gidding HF, Morgan J, Gilbert GL. Distribution of genotypes and antibiotic resistance genes among invasive *Streptococcus agalactiae* (group B streptococcus) isolates from Australasian patients belonging to different age groups. *Clin Microbiol Infect*. 2008;14(3):260–267.

64. Trijbels-Smeulders MAJM, Kimpen JLL, Kollée LAA, et al. Serotypes, genotypes, and antibiotic susceptibility profiles of group B Streptococci causing neonatal sepsis and meningitis before and after introduction of antibiotic prophylaxis. *Pediatr Infect Dis J.* 2006;25(10):945–948.

65. Fluegge K, Supper S, Siedler A, Berner R. Serotype distribution of invasive group B streptococcal isolates in infants: results from a nationwide active laboratory surveillance study over 2 years in Germany. *Clin Infect Dis*. 2005;40:760–763.

66. Chang B, Wada A, Hosoya M, et al. Characteristics of group B Streptococcus isolated from infants with invasive infections: A population-based study in Japan. *Jpn J Infect Dis*. 2014;67(5):356–360.

67. Liu H, Zeng H, Wang W, et al. Estimating the burden of invasive Group B Streptococcal disease in young infants in southern mainland China: An observational study. *Int J Clin Exp Med*. 2015;8(8):13699–13707.

68. Morozumi M, Wajima T, Kuwata Y, et al. Associations between capsular serotype, multilocus sequence type, and macrolide resistance in *Streptococcus agalactiae* isolates from Japanese infants with invasive infections. *Epidemiol Infect*. 2014;142(4):812–819.

69. Six A, Firon A, Plainvert C, et al. Molecular characterization of nonhemolytic and nonpigmented group b Streptococci responsible for human invasive infections. *J Clin Microbiol*. 2016;54(1):75–82.

70. Yoon IA, Jo DS, Cho EY, Choi EH, Lee HJ, Lee H. Clinical significance of serotype V among infants with invasive group B streptococcal infections in South Korea. *Int J Infect Dis*. 2015;38:136–140.

71. Sigaúque B, Kobayashi M, Vubil D, et al. Invasive bacterial disease trends and characterization of group B streptococcal isolates among young infants in southern Mozambique, 2001–2015. *PLoS* One. 2018;13:e0191193.

72. Dangor Z, Cutland CL, Izu A, et al. Temporal changes in invasive group B streptococcus serotypes: Implications for vaccine development. *PLoS One*. 2016;11(12):e0169101.

73. Phares CR. Epidemiology of invasive group B Streptococcal disease in the United States, 1999-2005. *JAMA*. 2008;299(17):2056.