Distribution of capsule and O types in *Klebsiella pneumoniae* causing neonatal sepsis in Africa and South Asia: meta-analysis of genome-predicted serotype prevalence and potential vaccine coverage

Short title: *Klebsiella pneumoniae* K and O serotypes from neonatal sepsis in Africa and South Asia

Thomas D Stanton[1,2,‡], Shaun P Keegan[3,‡], Jabir A Abdulahi[4], Anne V Amulele[5], Matthew Bates[6], Eva Heinz[7,8], Yogesh Hooda[9], Weiming Hu[10], Kajal Jain[11], Samiah Kanwar[12], Rindidzani Magobo[13,14], Courtney P Olwagen[15], John M Tembo[16], Tolbert Sonda[17], Jonathan Strysko[18,19], Caroline C Tigoi[5], Sameen Ahmad Amin[20], Kyle Bittinger[10], Jennifer Cornick[21,22], Ebenezer Foster-Nyarko[3], Wilson Gumbi[5], Aneeta Hotwani[12], Naveed Iqbal[12,23], Steven M Jones[10], Furqan Kabir[24], Waqasuddin Khan[25], Chileshe L Musyani[26,27], Carolyn M McGann[28,29], Varsha Mittal[11], Ahmed M Moustafa[10,29,30], Patrick Musicha[22,31], James CL Mwansa[32], Moreka L Ndumba[5], Erkison E Odih[3], Donwilliams O Omuoyo[5], Oliver Pearse[22,31], Laura T Phillips[3], Paul J Planet[30,33], Aniqa Abdul Rasool[12], Charlene MC Rodrigues[3,34,35], Kirsty Sands[36,37], Arif M Tanmoy[9], Erin Theiller[30], Allan M Zuza[22], Sulagna Basu[38], Grace J Chan[39], Kenneth C Iregbu[40,41], Jean-Baptiste Mazarati[42,43], Semaria Solomon Alemayehu[44], Timothy R Walsh[36,45], Rabaab Zahra[46,47], Angela Dramowski[48], Sombo Fwoloshi[49,50], Appiah-Korang Labi[51], Lola Madrid[4,52], Noah Obeng-Nkrumah[53], David Ojok[54], Boaz D Wadugu[55], Andrew C Whitelaw[56,57], Adhisivam Bethou[58], Anudita Bhargava[59], Atul Jindal[60], Ruchi N Nanavati[61], Priyanka S Prasad[62], Apurba Sastry[63], Joveria Q Farooqi[20], Najia Ghanchi[20], Fyezah Jehan[12], Erum Khan[20], Ramesh K Agarwal[11], Alexander M Aiken[4], James A Berkley[64,65], Susan E Coffin[29,66], Nicholas A Feasey[67,22], Nelesh P Govender[68,69], Davidson H Hamer[70,71], Shabir A Madhi[15,72], M Imran Nisar[12], Samir K Saha[9], Senjuti Saha[9], M Jeeva Sankar[11], Kelly L Wyres[1,3,#,*], Kathryn E Holt[1,3,#,*]

[1] Department of Infectious Diseases, School of Translational Medicine, Monash University, Melbourne, Australia
[2] Centre to Impact AMR, Monash University, Clayton, Victoria, Australia, 3800
[3] Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[4] Department of Infectious Disease Epidemiology and International Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
[5] Bioscience, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
[6] School of Natural Sciences, University of Lincoln, Lincoln, United Kingdom
[7] Strathclyde Institute of Pharmacy & Biomedical Sciences, University of Strathclyde, Glasgow, UK
[8] Parasites and Microbes Program, Wellcome Sanger Institute, Hinxton, UK CB10 1SA
[9] Child Health Research Foundation, Dhaka, Bangladesh
[10] Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, USA
[11] Pediatrics, All India Institute of Medical Sciences, New Delhi, India
[12] Department of Paediatrics and Child Health, The Aga Khan University, Karachi, Pakistan
[13] Centre for Healthcare Associated Infections, Antimicrobial Resistance and Mycoses, National Institute for Communicable Diseases, Johannesburg, South Africa

[14] Division of Public Health, Societies and Belonging, Human Sciences Research Council, Pretoria, South Africa, 0001
[15] South African Medical Research Council Vaccines and Infectious Diseases Analytics Research Unit, University of the Witwatersrand, Johannesburg, South Africa
[16] Infectious Diseases, HerpeZ, Lusaka, Zambia
[17] Bioinformatics Unit, Kilimanjaro Clinical Research Institute, Moshi, Tanzania
[18] Botswana-University of Pennsylvania Partnership, University of Pennsylvania, Gaborone, Botswana
[19] Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA, 19104
[20] Department of Pathology and Laboratory Medicine, The Aga Khan University, Karachi, Pakistan
[21] Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom
[22] Malawi Liverpool Wellcome Programme, Blantyre, Malawi
[23] Department of Biotechnology and informatics, Faculty of Life Sciences & Informatics, BUITEMS, Quetta, Pakistan
[24] Infectious Diseases Research laboratory, Department of Paediatrics & Child Health, Aga Khan University, Karachi, Pakistan
[25] CITRIC Center for Bioinformatics and Computational Biology, Department of Pediatrics and Child Health, The Aga Khan University, Karachi, Pakistan
[26] Antimicrobial Resistance Unit, Zambia National Public Health Institute, Lusaka, Zambia
[27] Levy Mwanawasa Medical University
[28] Pediatrics, Children's Hospital of Philadelphia, Philadelphia, United States
[29] Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
[30] Center for Microbial Medicine, Children’s Hospital of Philadelphia; Philadelphia, USA, 19104
[31] Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
[32] Directorate of Research and Post Graduate Studies, Lusaka Apex Medical University, Lusaka, Zambia
[33] Pediatrics, University of Pennsylvania, Philadelphia, USA
[34] Department of Paediatrics, Imperial College Healthcare NHS Trust, London, United Kingdom, W21NY
[35] Pathogen Genomics Programme, UK Health Security Agency, London, United Kingdom, NW9 5EQ
[36] Ineos Oxford Institute for Antimicrobial Research, University of Oxford, Oxford, United Kingdom
[37] Department of Medical Microbiology, Division of Infection and Immunity, Cardiff University, Cardiff, UK
[38] Division of Bacteriology, ICMR-National Institute for Research in Bacterial Infections (formerly ICMR-NICED), Kolkata, India
[39] Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, USA
[40] Medical Microbiology, University of Abuja, Abuja, Nigeria
[41] National Hospital Abuja, Abuja, Nigeria
[42] Health Sciences, INES-RUHENGERI, Musanze, Rwanda

[43] INES Centre for Genomic Biology, Musanze, Rwanda
[44] Department of Microbiology, Immunology and Parasitology, St. Paul's Hospital millennium Medical College, Addis Ababa, Ethiopia
[45] International Centre for Antimicrobial Resistance Solutions, Copenhagen, Denmark (ICARS)
[46] Microbiology, Quaid-i-Azam University, Islamabad, Pakistan
[47] Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, St James's Hospital, Dublin 8, Ireland D08 RX0X
[48] Department of Paediatrics and Child Health, Stellenbosch University, Cape Town, South Africa
[49] University Teaching Hospital, University of Zambia, School of Medicine, Lusaka, Zambia
[50] Infectious Diseases Unit, Department of Internal Medicine, UTH, Lusaka Zambia
[51] Department of Medical Microbiology, University of Ghana Medical School, Accra, Ghana
[52] College of Health and Medical Sciences, Haramaya University, Harar, Ethiopia
[53] Department of Medical Laboratory Sciences, School of Biomedical and Allied Health Sciences, College of Health Sciences, University of Ghana, Accra, Ghana
[54] Central Laboratory, Centre for Infectious Disease Research in Zambia, Lusaka, Zambia
[55] Biotechnology Laboratory, Kilimanjaro Clinical Research Institute, Moshi, Tanzania
[56] Medical Microbiology and Immunology, Stellenbosch University, Faculty of Medicine and Health Sciences, Cape Town, South Africa
[57] National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa, 7505
[58] Department of Neonatology, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Puducherry, India
[59] Microbiology, All India Institute of Medical Sciences, Raipur, India
[60] Pediatrics, All India Institute of Medical Sciences, Raipur, India
[61] Department of Neonatology, Seth GS Medical College and KEM Hospital, Mumbai, India
[62] Department of Microbiology, Seth GS Medical College & KEM Hospital, Mumbai, India
[63] Microbiology, JIPMER, Puducherry, India
[64] Nuffield Department of Medicine, University of Oxford, Oxford, UK
[65] Clinical Research Deprtment, KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya, 80108
[66] Children's Hospital of Philadelphia
[67] The School of Medicine, University of St Andrews, St Andrews, UK
[68] Wits Mycology Division, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[69] National Institute for Communicable Diseases, a Division of the National Health Laboratory Service, Johannesburg, South Africa
[70] Department of Global Health, Boston University School of Public Health, Boston, USA
[71] Section of Infectious Diseases, Boston University Chobanian & Avedisian School of Medicine
[72] Wits Infectious Diseases and Oncology Research Institute, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

* Corresponding author: kat.holt@lshtm.ac.uk, kelly.wyres@monash.edu

[‡]These authors contributed equally to this work
[#]These authors also contributed equally to this work

Abstract

Background: *Klebsiella pneumoniae* causes ~20% of sepsis in neonates, with ~40% crude mortality. A vaccine administered to pregnant women, protecting against ≥70% of *K. pneumoniae* infections, could avert ~400,000 cases and ~80,000 deaths annually, mostly in Africa and South Asia. Vaccine formulations targeting the capsular polysaccharide (K) or lipopolysaccharide (O) antigens are in development. Global *K. pneumoniae* populations display extensive K and O diversity, necessitating a polyvalent vaccine targeted to the serotypes associated with neonatal disease in relevant geographical regions. We investigated the prevalence of K and O types associated with neonatal sepsis in Africa and South Asia to inform maternal vaccine design.

Methods and Findings: We analysed n=1930 *K. pneumoniae* neonate blood isolates from 13 surveillance studies across 35 sites in 13 countries. We used pathogen sequencing to predict K and O serotypes and correct for local transmission clusters, and Bayesian hierarchical meta-analysis to estimate K and O prevalence. Eighty-seven K loci were identified. KL2, KL102, KL25, KL15 and KL62 accounted for 49% of isolates. We estimate that 20 K loci, combining the eight most prevalent per region, could cover 72.9% of all infections [95% credible interval, 69.4–76.5%] and ≥70% in each of Eastern, Western and Southern Africa and South Asia. Preliminary findings from three sites suggested sufficient temporal stability of K loci to maintain 20-valent K vaccine coverage over 5-10 years, but more longitudinal data are needed to support this prediction. O types were far less diverse (n=14 types). We estimate the top-5 (O1αβ,2α, O1αβ,2β, O2α, O2β and O4) would cover 86% [82.6–89.9%] of total infections (76–92% per region), while the top-10 would cover ~99% of infections in all four regions.

Conclusions: Neonatal sepsis is associated with diverse K and O types, with substantial geographic and temporal variation even after adjusting for localised transmission clusters. Despite this, a single 20-valent K vaccine could theoretically cover ≥70% of infections in all target regions. Locally-targeted vaccines could achieve higher coverage with lower valency, but are less feasible. In principle, very high coverage could be achieved with lower valency O-based vaccines, however protective efficacy of antibodies targeting the O antigen remains uncertain. Further research is needed on cross-reactivity, antigen exposure and stability of antigens over time, to better inform vaccine development.

Introduction

An estimated 670,000 newborns die annually from sepsis, with the greatest burden in Africa (~350,000) and South Asia (~220,000) [1]. The most common causative agents are the bacterial pathogens *Staphylococcus aureus, Klebsiella pneumoniae* and *Escherichia coli*, with *K. pneumoniae* accounting for >20% of cases [1–4]. The majority of *K. pneumoniae* cultured from neonates are resistant to the antimicrobials recommended by the World Health Organization (WHO) for the treatment of neonatal sepsis [3,5], ampicillin/benzylpenicillin and gentamicin, and it is estimated that nearly 100,000 neonatal sepsis deaths annually are associated with antimicrobial resistant (AMR) *K. pneumoniae* [1]. The Child Health and Mortality Prevention Surveillance (CHAMPS) study, which has sites in seven countries in Africa and Bangladesh in South Asia, identified *K. pneumoniae* as a contributor to 18% of neonatal deaths (and 21% of deaths amongst children under 5 years), with most cultured isolates resistant to ceftriaxone (84%) or gentamicin (80%) [6] the two most commonly used antibiotics.

Given the high burden of neonatal sepsis caused by AMR *K. pneumoniae*, there is an urgent need for preventive measures such as improved infection prevention and control in neonatal healthcare facilities and maternal vaccination [7–9]. It is estimated that a *K. pneumoniae* maternal vaccine given in pregnancy with 70% efficacy administered with coverage equivalent to that of the maternal tetanus vaccine (~70%) could avert ~400,000 neonatal sepsis cases and ~80,000 associated deaths annually worldwide, with the greatest impact in Africa and South Asia [10]. The WHO has identified maternal *K. pneumoniae* vaccine as a 'critical' priority, with potential to halve the antibiotic use associated with treatment of *K. pneumoniae* neonatal sepsis globally and save >$US 270 million annually in hospital costs [11].

Polysaccharide antigens, specifically the capsular polysaccharide (K) and lipopolysaccharide (O) antigens, are considered promising candidates for a maternal *K. pneumoniae* vaccine [7]. There are several licensed capsular vaccines against other invasive bacterial pathogens, which can induce protection following a single dose [12–14]. A 6-valent capsular vaccine against group B *Streptococcus* has been developed, and its potential for use in pregnant women to prevent neonatal infection is under investigation [15]. Extensive serotype diversity (>80 K and >10 O) poses a key challenge for *K. pneumoniae* vaccine design. O-based vaccines are attractive due to the smaller number of *K. pneumoniae* O serotypes, are safe and immunogenic [16], but there are currently no such vaccines licensed. Furthermore, some studies suggest O-based vaccines may not be effective at preventing *K. pneumoniae* disease, at least in the presence of certain common K serotypes which have been proposed to exert a masking effect [16,17]. Multi-valent K-based vaccines, or some combination of K and O, may therefore be required for *K. pneumoniae*. Licensed *Streptococcus pneumoniae* glycoconjugate vaccines target up to 20 capsule types; however, the distribution of vaccine-targeted capsule types varies markedly across geographies, reducing vaccine impact [18–20]. Similar to *S. pneumoniae*, understanding the distribution of *K. pneumoniae* serotypes will be critical to inform the design of vaccines aiming to protect neonates in Africa and Asia.

Serological typing for *K. pneumoniae* is not available outside a few high-income country centres, and the primary tool for profiling K and O serotypes is currently whole-genome sequencing. Serotype predictions can be made based on knowledge of how genetic variation in the K and O antigen biosynthesis loci map to variation in serotype, implemented in the genome-analysis software Kaptive [21]. The genetic determinants underpinning O antigen variation are well understood [22] and include genes located within and outside of the O locus (also detected in Kaptive [23,24]). K antigen biosynthesis is encoded in the K locus, and >160 K loci have been defined on the basis of unique gene content, which maps to unique sugar composition and linkages [22]. Half of these K loci correspond to known serologically distinct reference K serotypes [25]; the others have yet to be fully characterised, but available data support that distinct K loci encode synthesis of K antigens with distinct sugar structures [26–28]. K loci are labelled KL1, KL2, etc, with the numbers corresponding to the antigen structures they encode, which are labelled K1, K2, etc. Comparison of serological data for isolates carrying K loci matching to known reference K serotypes indicates 85% concordance [29].

Available genomic surveillance data suggest that *K. pneumoniae* serotypes associated with bloodstream infection differ geographically and temporally [2,30–32]. The limited data available specifically for neonates suggest marked differences in serotype prevalence between sites [2,30], skewed by localised clustering that likely reflects local outbreaks within the neonatal units as occurs frequently in Africa and Asia [4,30,33].

Here, we analyse *K. pneumoniae* isolated from the blood of neonates with suspected sepsis in African and South Asian countries sourced from 13 different prospective surveillance studies over the last decade, aiming to: (1) estimate the prevalence of genomically-predicted K and O serotypes amongst *K. pneumoniae* associated with neonatal sepsis in low- and middle-income countries (LMIC) in Asia and Africa during this time period; and (2) estimate the cumulative coverage of neonatal sepsis cases offered by different sets of K or O antigens, including subgroup analyses assessing coverage within (i) geographical regions, (ii) neonates who died during follow-up, and (iii) infections caused by extended-spectrum beta-lactamase (ESBL) producing or carbapenemase producing (CP) strains. We use whole-genome sequencing to predict K and O serotypes and correct for local transmission clusters, and Bayesian hierarchical meta-analysis to estimate total and regional serotype prevalence and theoretical vaccine coverage. Our findings will inform the design and deployment of maternal vaccines against *K. pneumoniae* neonatal sepsis.

Methods

Ethical considerations

Each contributing study obtained local ethical approval, listed in **Table S1**. The Baby GERMS-SA [34] and MLW [30] studies were granted consent waivers from local ethics committees for the use of routine diagnostic specimens/isolates and clinical data for the research; all other studies obtained informed consent from the parents of participating neonates. Approval for the meta-analysis presented here was granted by the Observational / Interventions Research Ethics Committee of the London School of Hygiene and Tropical Medicine (ref #29931). Anonymised data from each primary study were shared for analysis,

including date of specimen collection, hospital site identifier (where the study included more than one site), and mortality outcome (where available) for each participant, along with pathogen genome data for the corresponding bacterial isolate.

Bacterial isolates and sequence analysis

Whole-genome sequences of *K. pneumoniae* isolated from neonates in low- and middle-income countries in Asia and Africa were sourced from prospective clinical studies of neonatal sepsis (BARNARDS [2], SPINZ [35], Baby GERMS-SA [34], MBIRA [36], DH [37], GBS-COP [38], NIMBI-plus, NeoOBS-India [3], NeoBAC), long-term prospective surveillance of bloodstream infection (MLW [30], KWTRP, CHRF [39]), and referral-based surveillance of blood cultures in Pakistan (AKU) (**Table S2**).

For the prospective studies (BARNARDS, SPINZ, Baby GERMS-SA, MBIRA, DH, GBS-COP, NIMBI-plus, NeoOBS-India, NeoBAC), blood culture was performed for all neonates with clinically suspected sepsis, and all blood culture isolates identified as *Klebsiella* that could be later revived for DNA extraction were included in sequencing and this meta-analysis. Baby GERMS-SA, GBS-COP, NeoOBS, KWTRP and MLW also included *Klebsiella* cultured from cerebrospinal fluid (CSF) of participants.

For the long-term prospective surveillance, isolates from all positive blood cultures (CHRF, KWTRP, MLW) or CSF cultures (KWTRP, MLW) were stored for future research, and all those identified as *Klebsiella* isolated from neonates that could be later revived for DNA extraction were included in sequencing and this meta-analysis. At KWTRP, all neonates had blood culture on admission and again if their clinical condition deteriorated. At MLW, cultures were performed for all neonates with clinical signs of sepsis or meningitis, temperature >37.5 °C, or other signs of clinical deterioration. At CHRF, blood cultures were performed as part of routine diagnostics whenever clinically indicated.

For the AKU referral-based study, all positive blood cultures obtained from infant inpatients (including emergency) at Aga Khan University Hospital (AKUH), as well as all those obtained from the AKUH Laboratory Network (comprising laboratories across Pakistan), as part of routine diagnostics were stored for analysis. All those identified as *Klebsiella* isolated from neonates during the study period, that could be later revived for DNA extraction, were included in sequencing and this meta-analysis.

Details of microbiology, sequencing and bioinformatics methods used in each study, including quality-control criteria applied prior to sharing genome sequences for this meta-analysis, are summarised in **Table S2**. All studies utilised Illumina platforms and assembled genomes using SKESA v2.3.0 (MBIRA) or SPAdes (all other studies). Genome assemblies were analysed using Kleborate v3.0 [40] to confirm species, identify multi-locus sequence types (STs), antimicrobial resistance (AMR) and hypervirulence determinants, and to identify K and O types via Kaptive v3.0 [21]. For the K antigen, we used the K locus call as the unit of analysis. For the O antigen we used O type as the unit of analysis, using the latest O subtype nomenclature [22], inferred from the combination of O locus and O type reported by Kaptive v3.0 (full genetic definitions in **Appendix S1**). Genomes identified by Kaptive as untypeable (<1%) were included in the denominators for all analyses.

Inclusion criteria for meta-analysis

Inclusion criteria for individual samples from these studies were: *K. pneumoniae* isolated from the blood or CSF of a neonate (defined as 0-30 days post birth, or 0-28 days for Baby GERMS-SA, DH, MBIRA, MLW, and SPINZ), between 2013-2023. CSF isolates were included from four studies only (NeoOBS-India, n=9; KWTRP surveillance, n=5; GBS-COP, n=2; MLW, n=31). Repeat isolates of *K. pneumoniae* from the same participant with the same KL were excluded, such that each genome represents a unique infection from an individual neonate. Genome sequences identified as species other than *K. pneumoniae*, or failing to meet the meta-analysis quality control criteria for assemblies (≤1000 contigs and genome size 5-6.2 Mbp, as reported by Kleborate v3), were excluded. Individual study sites with fewer than 10 high-quality *K. pneumoniae* genomes passing the inclusion criteria were excluded (81 genomes from 21 sites). A flow diagram for sample inclusion is shown in **Figure S1**, including available information on the number of samples that were not stored, were unable to be revived for DNA extraction, failed sequencing, or failed genome quality control for each contributing study.

Subgroup definitions

African countries were assigned to regions using the United Nations Statistics Division standard M49 [41] (intermediate region level). Bangladesh, India and Pakistan were assigned to South Asia.

Isolates were assigned as ESBL if they had a Kleborate v3 resistance score of ≥1 (indicating detection of at least one known ESBL gene in the genome), and as CP if they had a Kleborate v3 resistance score of ≥2 (indicating detection of at least one known carbapenemase gene in the genome [40]).

Fatalities were defined as all neonates with blood-culture isolates of *K. pneumoniae* who were recorded as dying from any cause, either during the hospital admission in which the blood culture was taken or during a defined follow-up period, depending on the study. Some studies recorded in-hospital deaths only (Baby GERMS-SA, KWTRP, NeoBAC, SPINZ), others also undertook follow-up for a defined period (MBIRA: 30 days post-enrolment, NeoOBS-India: 28 days post-enrolment, DH: 28 days of life, BARNARDS: 60 days of life). Mortality data were not available from the MLW, CHRF, NIMBI-plus and AKU surveillance studies.

Identifying clusters

To ensure a simple and reproducible analysis, we used pairwise single nucleotide variant (SNV) distances downloaded from Pathogenwatch (which we refer to as ‘PW distances’). These are calculated across the core gene library defined by Pathogenwatch for *K. pneumoniae* (1,972 genes, 2,172,367 bp) [42]. There is emerging consensus across multiple studies that a threshold of 21-25 genome-wide SNVs (estimated using the more common method of mapping reads to a complete *K. pneumoniae* reference genome) is suitable for identifying nosocomial transmission clusters of *K. pneumoniae* [43–45]. We recently showed that a genome-wide threshold of 25 SNVs is equivalent to a PW distance of 10 SNVs in *K. pneumoniae* [46], hence we used a threshold of PW distance ≤10 SNVs, between genomes isolated from the same study site and within 4 weeks (28 days), to define putative nosocomial transmission clusters for our primary analysis. We explored sensitivity of our

analyses to clustering thresholds by varying the genetic distance (1, 2, 3, 5, 10, 15, 20, 25 or 50 PW SNVs) and temporal distance (0, 1, 2, 4, 8, 52 weeks) (reported in **Appendix S2**). Clustering was done in R using the *igraph* package (v2.0.3) [47], by first creating an (undirected) graph from a list of edges connecting isolates from the same site and meeting the genetic and temporal distance thresholds, then extracting clusters (groups of linked components, i.e. single-linkage clustering).

Statistical analysis

Crude prevalence for a given K locus (or O type) was calculated as the proportion of genome counts (crude prevalence), or proportion of unique clusters (cluster-adjusted crude prevalence), carrying that locus/type.

Global (i.e. across all included samples) and regional prevalences were estimated using a Bayesian framework to fit a generalized linear mixed model with a binomial family, with events being either the crude genome counts (to model prevalence) or cluster counts (to model cluster-adjusted prevalence). Models were fitted using the *brms* package in R (v2.22.0) [48], which provides an interface to the probabilistic programming language Stan (v2.32.2) [49].

The model formulation is as follows:

$$event_i \sim Binomial(n_i, \pi_i)$$

where:

- $event_i$ is the number of events observed for the i-th trial (i.e. for locus/type i)
- n_i is the number of trials for the i-th observation
- π_i is the probability of an event occurring for the i-th trial, modeled as:

$$\pi_i = logit^{-1}\left(\sum_{k=1}^{K} \beta_{locus,k} I(locus_i = k) + \sum_{j=1}^{J} \beta_{region,j} I(region_i = j) + \sum_{m=1}^{M} \gamma_m I(site_i = m) + \sum_{k=1}^{K} \sum_{j=1}^{J} \gamma_{k,j} I(locus_i = k \wedge region_i = j) \right)$$

where:

- $logit^{-1}(x) = \frac{1}{1+e^{-x}}$ is the inverse logit function that transforms the linear predictor into a probability between 0 and 1.
- The indicator functions, *I*, return observations if $locus_i = k$ and $region_i = j$.
- $\beta_{locus,k}$ and $\beta_{region,j}$ are the fixed effects for locus and region, respectively.
- $\gamma_m I(site_i = m)$ is the random effect intercept for site.
- $\gamma_{k,j} I(locus_i = k \wedge region_i = j)$ is the random intercept for each *locus* and *region* combination.

The priors for the standard deviations of the random effects (for *site* and *locus:region*) were specified as student-t distributions with 3 degrees of freedom, location 0, and scale 5. Each model was fit with 4 chains, each running for 6000 iterations, with 3000 burn-in iterations.

Control parameters were set to ensure stable sampling, with adapt_delta = 0.999999 and max_treedepth = 55. In cases where initial model diagnostics warranted, models were run again with 30,000 iterations with 15,000 burn-in iterations (**Appendix S3**).

Region was included as a fixed effect in all models, so that each region contributed equally to the global prevalence estimates despite heterogeneity in population sizes and the number of sites and samples available from each region.

To estimate global and regional prevalence for subgroups (fatalities, ESBL, or CP), the primary data were filtered to include only cases belonging to that subgroup, then any sites with fewer than 10 included cases were excluded from the analysis, before fitting the same model specified above.

For each model we used the *posterior_epred* function from the *brms* package to generate a posterior distribution of predicted values from the fitted model, for each *locus* (i.e. the *locus* effect, reflecting global estimate) and for each combination of *locus* and *region* (i.e. the *locus + region + locus:region* effects, reflecting the regional estimates for each locus). These were summarised to obtain the mean value and 2.5% and 97.5% quantiles, which we interpret as the point estimate and 95% credible interval (CI) for the global and regional prevalences of each locus/type.

To estimate cumulative coverage, for each set of loci we obtained posterior distributions by summing the relevant locus/type estimates for each draw, then calculated the mean value and 2.5% and 97.5% quantiles, which we interpret as the point estimate and 95% CI for cumulative sets of loci/types.

Leave-one-out analysis was undertaken to explore robustness of global cluster-adjusted prevalence estimates, and particularly to understand how global prevalence estimates are influenced by individual studies. The above Bayesian models were each run repeatedly using the same methods outlined above (using cluster-adjusted counts), but excluding one study each time, and loci/types were ranked according to the resulting prevalence estimate.

To assess whether the prevalence of common K loci were associated with year or region of sampling, for each K locus in the overall top-30 we used the *logistf* R package (v1.26.0) to fit a logistic regression using Firth's bias reduction method, with the response variable being 1 or 0 (to indicate the presence of this locus vs any other) and predictors being year and region. A p-value <0.05 was used to assess significance.

All data processing, analyses and data visualisations were conducted using R Statistical Software (v4.4.1) [50] with packages *dplyr* (v1.1.4), *tidyverse* (v2.0.0), *ggplot2* (v3.5.1), *ggridges* (v0.5.6), and *patchwork* (v1.3.0).

Code availability

Model diagnostics are shown in **Appendix 3.** Model input, output and code, and all other analysis and plotting code required to produce all figures and tables, including supporting items and appendices, are available at github.com/klebgenomics/KlebNNSsero (DOI: 10.5281/zenodo.15756796).

Data availability

Sample-level data, including K/O calls, accessions and source information are given in **Table S3**. Raw and adjusted counts data used for modelling, and model code and outputs, are available at github.com/klebgenomics/KlebNNSsero (DOI: 10.5281/zenodo.15756796). An interactive web application, implemented in R shiny, is also available to (i) reproduce and explore the modelled estimates for prevalence and coverage data included in this paper, and (ii) undertake additional analyses of the raw data (e.g. to explore crude pooled estimates of prevalence and coverage for different subsets of loci and/or different subsets of samples). The web application is available at https://klebsiella.shinyapps.io/neonatal/. Raw whole-genome sequence data were deposited by the originating study teams in INSDC databases, under the following BioProjects: BARNARDS, PRJEB33565; SPINZ, PRJEB46513; MLW, PRJEB42462; NIMBIplus, PRJNA1168993; DH, PRJEB70311; BabyGERMS, PRJNA796486 and PRJNA1282934; CHRF, PRJEB90555; NeoOBS-India, PRJEB70311; GBS-COP, PRJNA1175467; Kilifi, PRJNA1265413; MBIRA: PRJNA1274034, NeoBAC, PRJNA1265413; AKU, PRJNA126164.

Results

A total of 1930 *K. pneumoniae* genomes, each representing a unique neonatal infection sampled in the last decade, were included from 35 sites in 13 countries from four regions in Africa and Asia (**Table 1, Figures 1 and S1**). Four studies included CSF isolates, accounting for n=47 infections (2.4%). Nearly all genomes yielded typeable K and O biosynthesis loci (99.1% K, 99.5% O), and a total of 87 distinct K loci and fourteen O types were identified across the studies.

The raw distribution of K loci is shown in **Appendix S1.1**. Four K loci were common, with crude prevalence exceeding 5% each (KL102, KL15, KL2, KL25), and together accounting for 45% of isolates. The distribution of K loci varied across countries, with no K loci found in all countries (KL2 and KL25 seen in n=11 of 13 countries,), and only twelve K loci found in all four regions sampled (KL2, KL3, KL10, KL15, KL17, KL19, KL25, KL39, KL48, KL62, KL102, KL122; accounting for 59% of isolates). Most K loci were observed in combination with one or two O types, and these patterns were associated with the clonal population structure of *K. pneumoniae* and spatiotemporal clustering (represented by lineage, i.e. sequence type (ST) in **Appendix S1.1**). In some countries the samples were dominated by specific KL/O combinations; these were mostly associated with a single lineage or ST, and identified in temporal clusters at a single study site (**Appendix S1.2**), consistent with local transmission in the hospital or the wider community. For example, isolates from Ethiopia were dominated by ST35-KL108/O1αβ,2β (30%) and ST37-KL15/O4 (22%), but these were all isolated from the same hospital in 2017 and these clones were not detected in a later study at a different site in Ethiopia. Similarly, isolates from Zambia were dominated by ST307-KL102/O2β, all isolated from one hospital in 2015-2017, but this clone was not identified in a later study from a different hospital in Zambia. In Tanzania, all data came from a single study site in 2021, and 90% were ST1741-KL104/O1αβ,2β; however this clone was rare elsewhere (n=5, 5.6% in India; n=5, 2.1% in Kenya; not detected elsewhere). The genome data therefore supports that observed K (and O) locus prevalence can be heavily influenced by localised nosocomial outbreaks, as noted above.

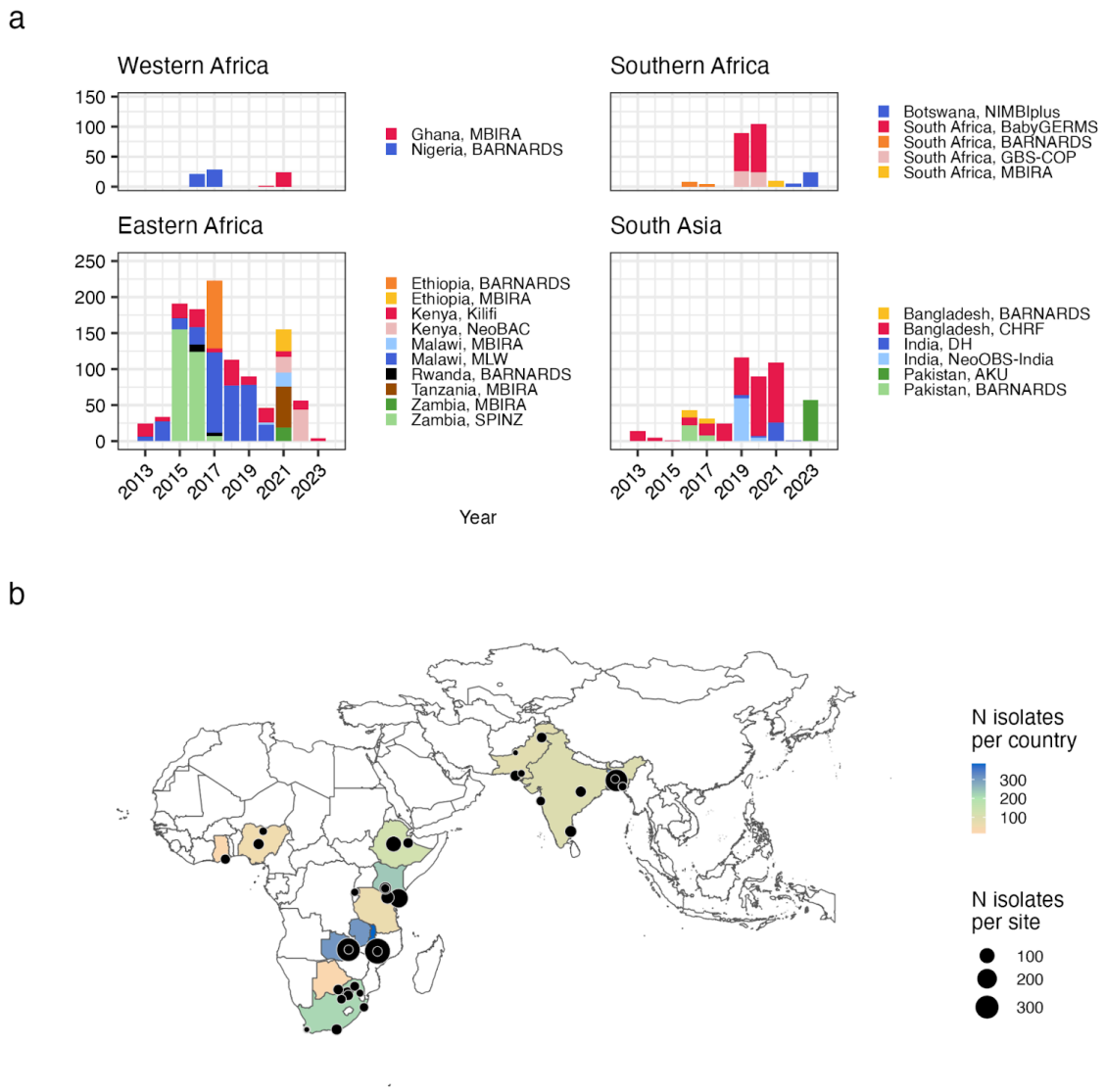

Figure 1. Geotemporal distribution of isolates included in the meta-analysis.
(a) Barplots indicate the number of samples included per year, from each study site (labelled as "Country, Study"), stratified by region. (b) Map shows location of study sites. Each point represents a study site, sized to indicate the number of samples included from that site, as per inset legend. Countries are coloured to indicate the total number of isolates included from that country, as per inset legend.

Table 1. Details of studies and sites included in the meta-analysis.

Region	Country	Study	Unique Sites	Unique Isolates	Period	Unique K loci	Unique O types
Eastern Africa	*All*	-	11	1119	-	60	11
	Ethiopia	BARNARDS	1	94	2017-2017	11	4
		MBIRA	1	30	2021-2021	10	6
	Kenya	Kilifi	1	168	2013-2023	43	10
		NeoBAC	2	66	2021-2022	5	3
	Malawi	MBIRA	1	22	2020-2021	11	4
		MLW	1	362	2013-2020	35	10
	Rwanda	BARNARDS	1	15	2016-2017	6	3
	Tanzania	MBIRA	1	57	2021-2021	5	3
	Zambia	MBIRA	1	19	2021-2021	10	6
		SPINZ	1	286	2015-2017	13	7
Southern Africa	*All*	-	10	244	-	33	8
	Botswana	NIMBIplus	1	29	2022-2023	8	7
	South Africa	BARNARDS	1	12	2016-2017	8	4
		Baby GERMS-SA	6	143	2019-2020	24	8
		GBS-COP	1	50	2019-2020	12	6
		MBIRA	1	10	2021-2021	6	3
Western Africa	*All*	-	3	76	-	27	9
	Ghana	MBIRA	1	26	2020-2021	14	7
	Nigeria	BARNARDS	2	50	2016-2017	18	8
South Asia	*All*	-	11	491	-	63	12
	Bangladesh	BARNARDS	1	17	2016-2017	12	8
		CHRF	3	289	2013-2021	41	10
	India	DH	1	34	2019-2022	8	8
		NeoOBS	2	64	2019-2020	26	9
	Pakistan	AKU	3	57	2023-2023	26	9
		BARNARDS	1	30	2016-2017	9	6
Total	-	-	*35*	*1930*	-	*87*	*14*

Modelled K locus prevalence estimates

We used Bayesian modelling to estimate global and regional prevalence for each K locus, based on meta-analysis of the 1930 *K. pneumoniae* genomes included from 13 studies in four regions (**Figure 2**). Given the influence of localised outbreak events on observed K locus counts at each site, we used cluster-adjusted counts whereby each genetic-temporal cluster (≤10 SNVs and ≤28 days) within a site (i.e. the same neonatal unit) is treated as a single observation (see **Methods**). Five loci (KL2, KL102, KL25, KL15, KL62) had mean prevalence estimates exceeding 4% each (with lower 95% credibility interval (CI) each ≥3.2%, see **Figure 2a**). A further seven K loci (KL30, KL10, KL17, KL23, KL51, KL112, KL39) had mean estimates exceeding 2% each (lower CI, ≥1.3%). In total, 27 K loci had mean prevalence estimates exceeding 1% (see **Figure 2a**).

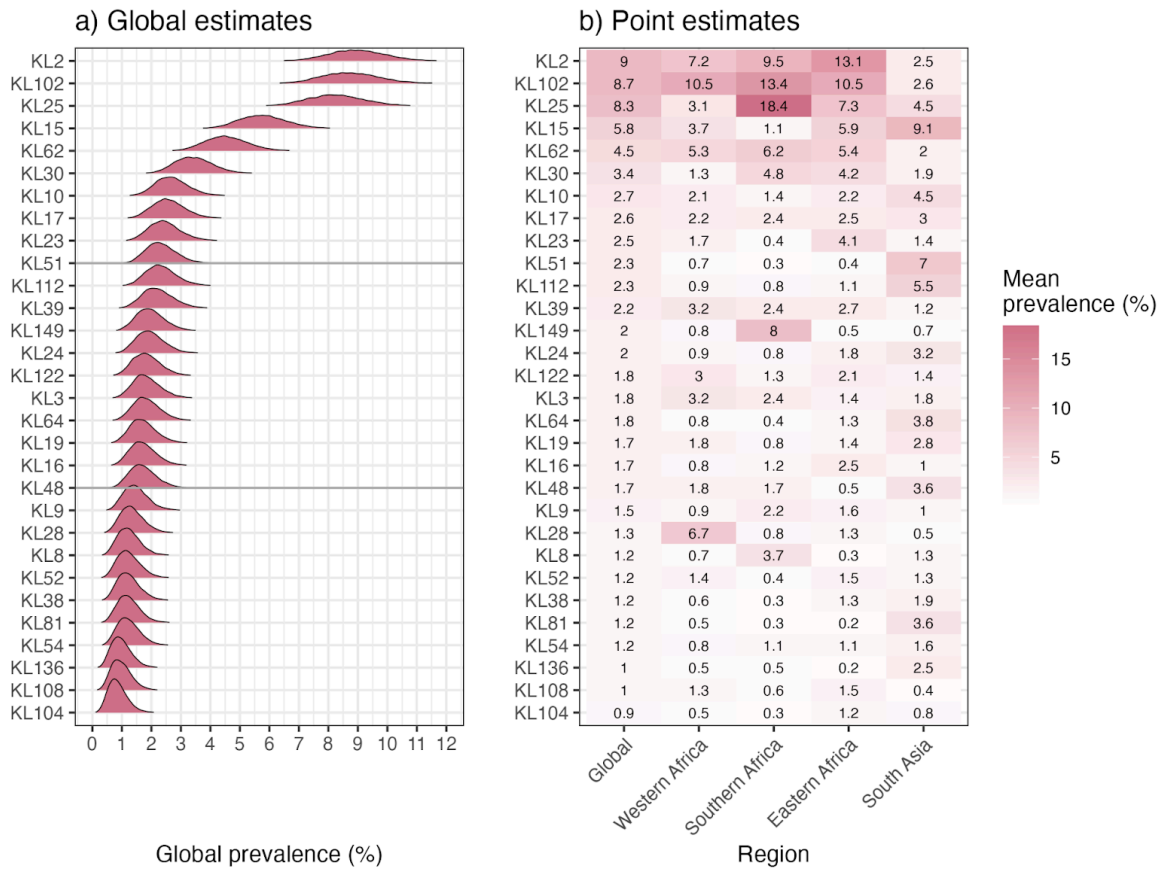

Figure 2. Modelled estimates of cluster-adjusted global and regional prevalence of K loci. (a) Posterior density distribution for global prevalence estimates (i.e. across all sites), modelled using cluster-adjusted counts per site, for top-30 K loci ordered by the mean point estimate. Horizontal lines indicate the top 10 and 20 loci. (b) Regional prevalence estimates. Cells are coloured and labelled to indicate the mean cluster-adjusted regional prevalence estimates for each K locus in each region, according to the inset legend. The underlying posterior distributions for K locus prevalence per region are shown in Figure S5.

K locus prevalence estimates varied substantially across regions (**Figures 2b, S2, S3**). Some of the global top-30 loci had significantly higher prevalence in a single region (**Figure S3**), e.g. KL149 in Southern Africa (8.0% [95% CI, 4.3–12.7%], vs ≤0.84% in other regions), and KL15 and KL51 in South Asia (9.1% [5.9–13.0%] vs ≤6% elsewhere and 7.0% [4.2-10.6%] vs ≤0.7%, respectively) (**Figures 2b, S3**). Outside the top-20, KL81 was common in South Asia (3.6%, [1.6–6.3%]) but very rare elsewhere (≤0.5%), and KL28, KL116 and KL53 were common in Western Africa (6.7% [2.0–14.0%], 5.6% [1.4–12.9%] and 4.0% [0.7–10.0%], respectively) but were rarely detected elsewhere (0.3–1.4%; see **Figures S2**). These regional differences could not be explained by differences in the timing of samples from different regions (**Appendix S1.3**). In a simple logistic regression for each K locus in the top-30, with year and region as linear predictors, region was significantly associated with 15 K loci (^ in **Figure S3**) but year was not significant for any K loci.

Theoretical K antigen coverage

To estimate the proportion of neonatal sepsis cases that could theoretically have been covered by a vaccine comprising subsets of K antigens, we modelled cumulative coverage based on prevalence estimated from raw (observed) counts (**Figure 3**). We estimate a vaccine with the global top-20 K-loci would cover 72.8% of all cases [95% CI, 69.3-76.4%], with higher coverage in Eastern Africa (75.7% [71.1–80.6%]) and Southern Africa (76.4% [66.6–86.8%]) and lower coverage in Western Africa (50.5% [36.9–66.4%]) and South Asia (67.8% [61.1–75.0%]) (**Table S4**). The marginal gain in coverage associated with adding more loci (in order of global rank) reduced substantially beyond the first 5 K-loci; these would provide theoretical global coverage of 48.8% of cases [46.0–51.7%], but we estimated the next five loci (global ranks 6-10) would provide additional global coverage of just 9.8% [9.6–10.2%], a further five (global ranks 11-15) 8.2% [8.0–8.4%], and a further five (global ranks 16-20) 5.8% [5.7–6.0%] (**Table S4**, **Figure 3**).

The global top-20 K loci provided very similar estimated coverage of neonatal fatalities (outcome data available for n=1063, 55.0% of genomes; of which n=394, 37.1% died), and coverage of infections with ESBL-producing strains (n=1747, 90.5% of genomes overall, 78.3–96.7% per region) (see **Table S4**, **Figure 3**). Infections with CP strains were very common in South Asia (87% in Bangladesh, 78% in India, 87% in Pakistan); common in Nigeria (46%), Ghana (27%) and South Africa (31%); and rare elsewhere (<3%). Estimated coverage of these CP infections was similar to the overall coverage (top-20 K loci covering 71.7% [64.8–79.0%]). However, four of the global top-5 K loci (KL2, KL102, KL25, KL62) were rare amongst CP isolates, and so the cumulative coverage of the top-5 was low for CP infections (40.9%, compared with 48.8% of total infections and 50.8% of ESBL infections).

As coverage estimates using the global top-20 K loci varied substantially by region (**Table S4, Figure 3**), we explored alternative strategies to select K loci in order to achieve better regional coverage. For each region, a locally tailored set of K loci including the top 20 most prevalent in that region could provide ≥80% [lower CI, ≥62%] coverage in the target region (**Figure S4**). However, any such formulations would provide insufficient coverage of non-target regions (**Figure S4**). We estimate an alternative set of 20 K-loci, comprising all those that ranked in the top-8 for any region, would cover ≥69.8% of cases in each region (range 69.8-82.2% per region, 72.9% overall), with similar coverage of ESBL, CP and fatalities (74.9%, 81.5%, 70.6%, respectively; see **Figure 3, Table 2**).

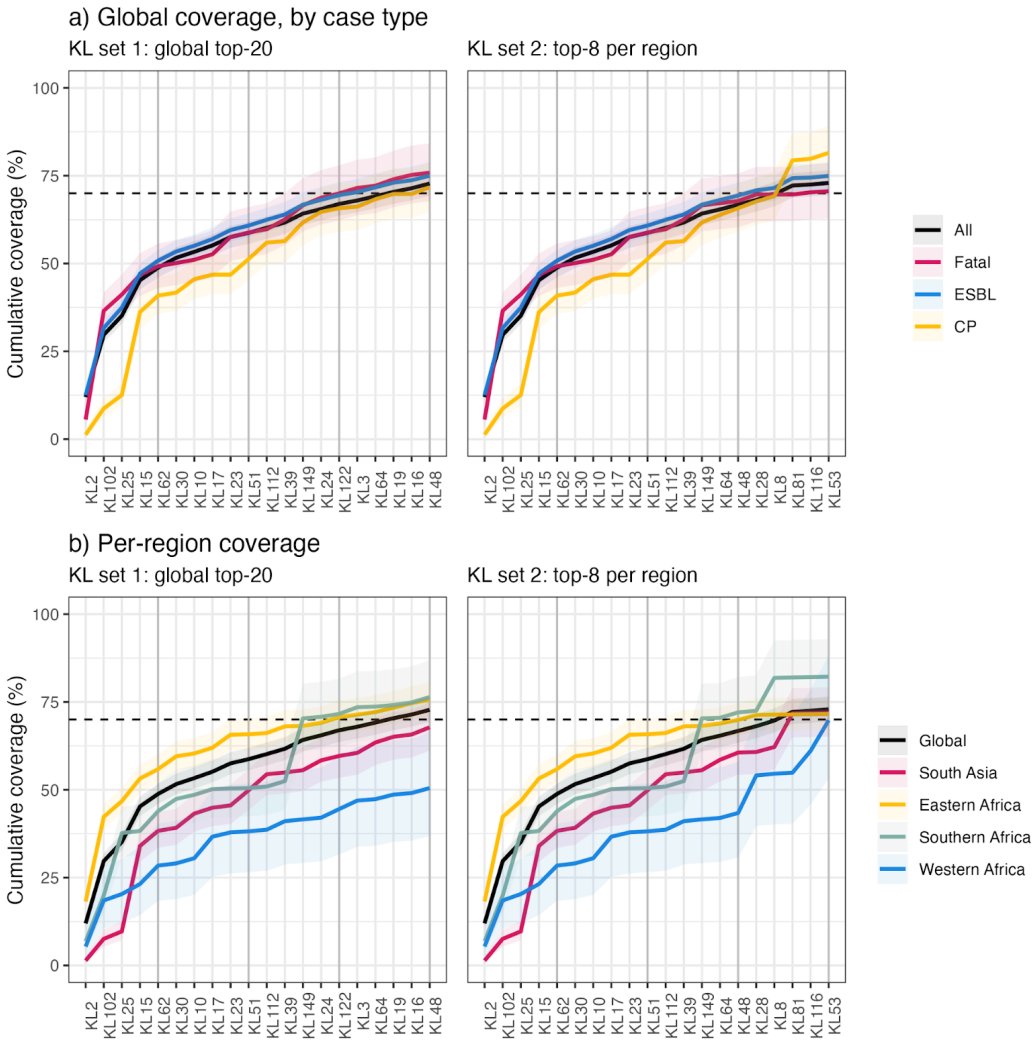

Figure 3. Cumulative coverage estimates for sets of K loci. (a) Global (overall) coverage of neonatal sepsis and subgroups of sepsis cases (fatal, ESBL, CP); and (b) per-region coverage of neonatal sepsis; selecting either the top-20 K loci ordered by cluster-adjusted global prevalence estimate (KL set 1) or selecting the top-8 K loci within each region (top-7 in Southern Africa) (KL set 2, total 20 K loci). KL, K locus; ESBL, extended-spectrum beta-lactamase; CP, carbapenemase-producing.

Table 2. Coverage estimates for 20 loci including the top-8 per region.

	Counts	N	5 KL	10 KL	15 KL	20 KL
Global	raw	1930	48.8 (46.0-51.7)	58.7 (55.6-61.9)	66.8 (63.4-70.2)	72.9 (69.4-76.5)
	adj	805	36.3 (32.4-40.3)	49.8 (45.3-54.6)	59.9 (54.9-65.1)	64.8 (59.6-70.2)
Fatal infections	raw	394	49.2 (43.1-55.7)	58.8 (51.8-66.1)	67.8 (60.3-75.6)	70.6 (62.8-78.5)
	adj	214	40.1 (31.5-49.7)	50.3 (40.5-61.2)	61.1 (50.1-73.2)	65.0 (53.5-77.2)
ESBL infections	raw	1747	50.8 (47.8-53.9)	60.8 (57.5-64.2)	69.4 (65.8-73.0)	74.9 (71.2-78.8)
	adj	683	38.2 (33.9-42.7)	52.3 (47.3-57.6)	63.5 (57.8-69.5)	68.3 (62.5-74.5)
CP infections	raw	525	40.9 (35.8-46.2)	51.3 (45.5-57.4)	65.8 (59.2-72.6)	81.5 (74.1-89.1)
	adj	228	24.0 (18.0-30.7)	41.5 (33.6-50.3)	59.5 (49.7-70.0)	71.0 (60.4-82.4)
Eastern Africa	raw	1119	55.9 (52.0-60.0)	65.8 (61.5-70.3)	69.9 (65.5-74.6)	71.5 (67.0-76.3)
	adj	367	42.1 (36.2-48.6)	55.5 (48.5-62.9)	61.6 (54.3-69.4)	63.9 (56.4-71.9)
Fatal infections	raw	237	56.4 (48.7-64.6)	64.5 (55.9-73.6)	68.4 (59.5-77.7)	68.8 (59.8-78.3)
	adj	100	45.5 (32.9-59.8)	54.4 (40.5-69.9)	62.7 (47.8-79.5)	64.3 (49.3-81.5)
ESBL infections	raw	1019	57.1 (52.9-61.3)	67.3 (62.7-72.0)	71.6 (66.8-76.4)	73.3 (68.5-78.2)
	adj	306	45.0 (38.2-52.2)	59.2 (51.2-67.7)	65.7 (57.4-74.6)	68.1 (59.6-77.1)
CP infections	raw	10	-	-	-	-
	adj	5	-	-	-	-
Southern Africa	raw	244	43.9 (36.5-51.9)	50.5 (42.4-59.0)	72.0 (62.7-82.1)	82.2 (72.1-92.8)
	adj	151	48.7 (39.1-59.4)	58.1 (47.6-69.5)	71.3 (59.5-84.2)	76.6 (64.4-90.0)
Fatal infections	raw	65	51.3 (31.9-74.1)	52.9 (33.1-75.9)	87.6 (62.9-100.0)	88.2 (63.4-100.0)
	adj	46	55.8 (31.4-85.8)	60.2 (34.8-90.7)	75.1 (46.0-100.0)	76.9 (47.4-100.0)
ESBL infections	raw	191	51.9 (42.6-61.8)	54.6 (45.0-64.7)	82.1 (70.3-94.2)	85.1 (73.2-97.3)
	adj	117	57.3 (44.7-71.0)	62.1 (48.9-76.7)	78.0 (62.9-94.1)	81.1 (65.6-97.7)
CP infections	raw	66	47.8 (30.9-67.4)	48.0 (31.1-67.6)	96.7 (74.3-100.0)	97.1 (74.5-100.0)
	adj	33	68.8 (38.4-100.0)	69.4 (38.8-100.0)	91.1 (55.7-100.0)	92.1 (56.5-100.0)
Western Africa	raw	76	28.4 (18.5-40.3)	38.2 (26.6-51.7)	43.4 (30.8-58.0)	69.8 (53.3-88.2)
	adj	55	29.8 (18.4-44.1)	37.8 (24.7-53.5)	45.2 (30.6-62.8)	62.8 (45.6-83.1)
Fatal infections	raw	36	19.6 (7.9-36.5)	40.4 (22.3-62.6)	50.4 (29.5-75.2)	74.1 (48.3-100.0)
	adj	28	27.8 (12.0-48.5)	38.5 (19.3-62.8)	51.2 (28.6-79.4)	65.9 (39.6-98.0)
ESBL infections	raw	62	32.8 (20.2-47.9)	44.8 (29.5-62.6)	50.7 (34.4-69.6)	82.2 (61.4-100.0)
	adj	43	34.9 (19.7-54.0)	43.5 (26.6-64.6)	52.1 (33.4-74.9)	72.3 (49.6-99.5)
CP infections	raw	30	5.4 (0.2-18.0)	5.7 (0.3-18.7)	6.4 (0.4-20.0)	94.0 (62.9-100.0)
	adj	17	9.4 (0.4-30.5)	10.3 (0.6-32.2)	11.5 (0.8-34.5)	89.0 (49.3-100.0)
South Asia	raw	491	38.3 (33.5-43.6)	49.9 (44.3-55.8)	60.6 (54.3-67.3)	71.9 (65.1-79.1)
	adj	232	20.8 (15.7-26.5)	38.5 (31.5-46.3)	53.3 (45.0-62.5)	59.3 (50.5-68.9)
Fatal infections	raw	56	35.8 (22.5-51.1)	50.7 (34.7-69.4)	61.9 (43.7-82.9)	63.0 (44.6-84.1)
	adj	40	25.7 (12.5-43.1)	42.6 (24.8-64.3)	55.3 (34.9-80.0)	57.8 (36.6-83.5)
ESBL infections	raw	475	39.0 (34.1-44.1)	51.0 (45.2-56.9)	62.2 (55.9-68.9)	73.9 (66.9-81.3)
	adj	217	20.7 (15.3-26.7)	39.8 (32.4-47.9)	55.9 (46.9-65.8)	62.3 (52.6-72.6)
CP infections	raw	419	41.9 (36.4-47.7)	54.0 (47.6-60.8)	65.4 (58.2-72.9)	79.1 (71.2-87.4)
	adj	173	20.3 (14.3-27.3)	40.8 (32.1-50.4)	59.6 (48.9-71.1)	67.5 (56.0-79.9)

Cumulative coverage estimates for KL set 2, comprising the top-8 K loci within each region, ranked according to cluster-adjusted global prevalence estimate. For each geographic location, the breakdown of isolate counts and corresponding coverage estimates are shown for subgroups defined by types of isolates (ESBL, extended-spectrum beta-lactamase; CP, carbapenemase-producing) or outcome of infection (fatal infections only). Coverage estimated from raw and cluster-adjusted counts ('adj') are shown.

O type prevalence and coverage

Global prevalence modelling of predicted O types (**Figure 4a**) confirmed the population was dominated by two O1 subtypes, O1αβ,2α and O1αβ,2β. The median prevalence estimates for these were 26.7% [95% CI, 23.6–29.9%]) for O1αβ,2α, and 21.8% [19.1–24.8%]) for O1αβ,2β. Next most common (≥9% each) were O2β (15.8% [13.5–18.4%]) and O4 (9.1% [7.2–11.1%). Combined, these top-4 O types were estimated to cover 81.1% [77.6–84.6%] of all included cases. Five other O types were detected with >3.0% prevalence, including O2α (6.2% [4.6–7.9%]); two O3 subtypes (O3γ, 6.0% [4.5–7.8%] and O3α/O3β, 2.8% [1.8–4.1%]), O5 (5.4% [3.9–7.0%]) and O13 (4.9% [3.5–6.5%]). Loci associated with other O types were very rare (mean estimates <0.4%, upper 95% CIs <0.9%; see **Figure 4a**).

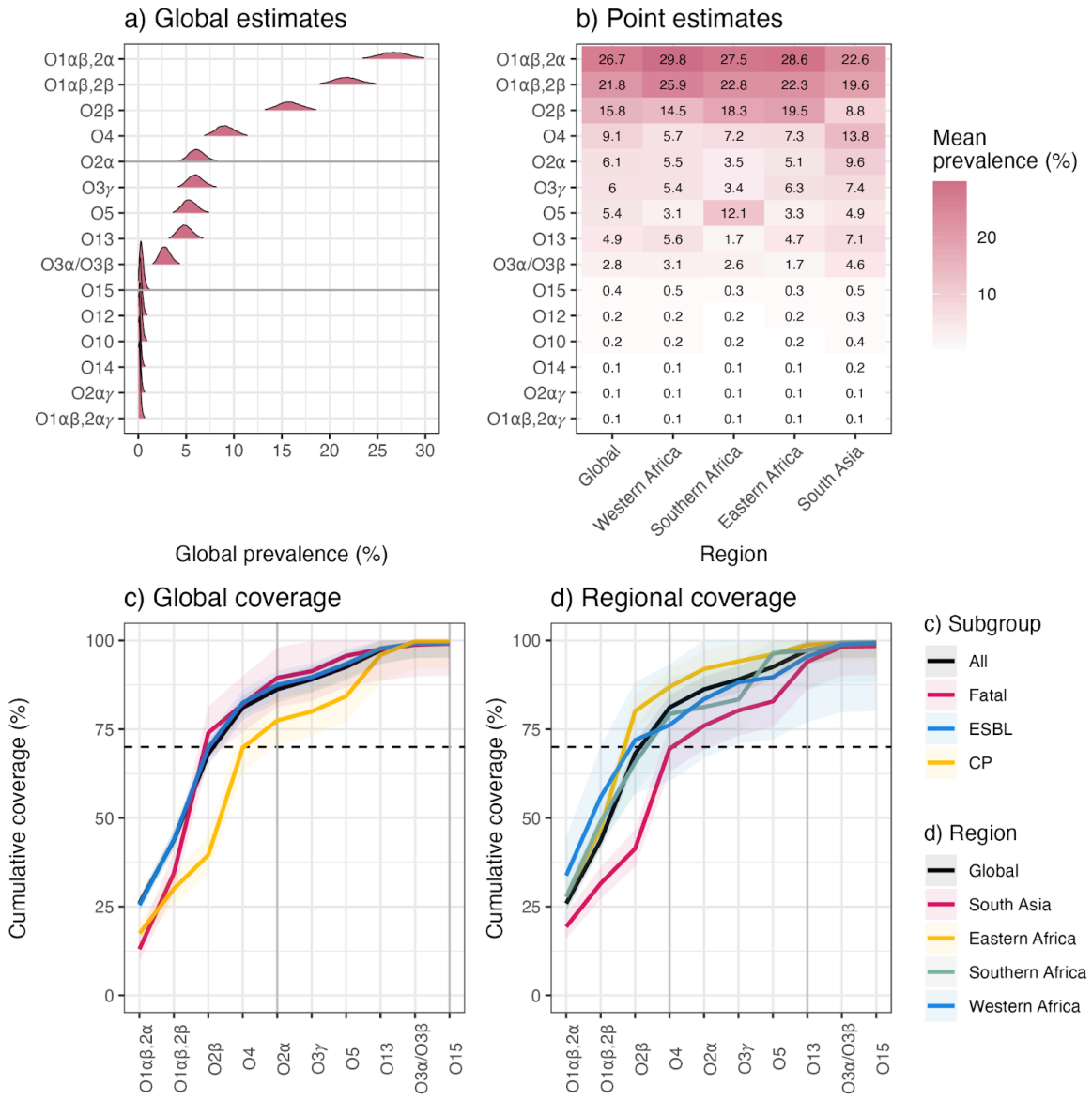

Figure 4. Modelled estimates of cluster-adjusted global and regional prevalence and coverage of O types. (a) Posterior density distribution for global (overall) prevalence estimates, modelled using cluster-adjusted counts per site, for O types ordered by the mean point estimate. Horizontal lines indicate the top 5 and 10 types. (b) Regional mean prevalence estimates. Cells are coloured and labelled to indicate the cluster-adjusted regional prevalence estimates for each O type in each region, according to the inset legend. (c) Global (overall) coverage of neonatal sepsis and subgroups of sepsis cases (fatalities, ESBL, CP), and (d) per-region coverage of neonatal sepsis, selecting the top-10 O types ordered by cluster-adjusted global prevalence.

There was some regional variation in O types (**Figures 4b and S5, Table S5**). Whilst O1 variants dominated in all regions, O2β was consistently the third most common in African regions (14.5–19.5%) but was less common in South Asia (8.8%), which had higher prevalence of O2α (9.6%, vs 3.5–5.5% in African regions) and O4 (13.8%, vs 5.7–7.3% in African regions). Compared with other regions, Southern Africa had higher prevalence of O5 (12.1%, vs 3.1–5.0% elsewhere), and lower prevalence of O3γ (3.3%, vs 5.4–7.4% elsewhere) and O13 (1.7%, vs 4.7–7.1% elsewhere). As with K locus prevalence, there was no evidence that regional differences were driven by differences in time of sampling (i.e. the calendar year of sampling for different sites); e.g. in South Asia, the prevalence of O2β was persistently lower than elsewhere, and O4 and O2α persistently higher, across the combined sampling period (**Appendix S1.4**). The differences in O types were also not explained by localised clonal outbreaks; e.g. in South Asia O4 was associated with 15 different ST-KL combinations (the most common being ST11-KL15 found in three sites in Bangladesh and two in India) and O2α with nine ST-KL combinations (**Appendix S1.5**). O5 in Southern Africa was primarily associated with a single clone ST17-KL25/O5, but this was distributed across seven sites in South Africa and one in Botswana, consistent with endemic spread in the region.

The estimated coverage of infections per subgroup and region are shown in **Figure 4c-d, Table S5**. We estimate the top-5 O types (O1αβ,2α, O1αβ,2β, O2β, O2α, O4) collectively cover 86.2% [82.6–89.9%] of total cases, 89.5% [81.3–98.0%] of fatalities, 87.5% [83.6–91.4%] of ESBL and 77.5% of CP [70.6–84.5%]; and ≥76% of cases in each region (range 76.0-92.0% per region, lower CI ≥69%) (**Table S5**).

Sensitivity analysis

To explore the sensitivity of our analyses to which primary studies were included, we repeated the meta-analysis leaving out each of the n=13 studies in turn (**Appendix S2.1-2.2**). Global modelled ranks for K loci were robust to study inclusion, with the same top-5 in all analyses and the top-20 also quite stable (only 5 loci ranking top-20 in any leave-one-out analysis were not included in the overall top-20). The cumulative coverage estimated for the final global top-20 K loci was also very robust to study inclusion (with the exception of regional coverage in West Africa, for which we had just two contributing studies, see **Appendix S2.1**). O type rankings were also very robust to study inclusion, with the same top-4 for all analyses, and cumulative coverage estimates for the top-4 were stable across analyses (see **Appendix S2.2**).

The impact of cluster-adjustment on the modelled and simple weighted global prevalence estimates for K loci are shown in **Appendix S2.3–2.4**. Note that cluster-adjustment is equivalent to counting only the index case for each cluster in each site, removing the random effect of nosocomial transmission events that may be driven by local infection control issues rather than specific bacteriological properties. Cluster-adjustment had a substantial effect (0.5–9.0% reduction in mean global prevalence estimate) for eight K loci, due to raw counts of these K loci each being inflated by a few large clusters (1-3 clusters per K locus accounting for 54-100% of total observations for that K locus). The largest (≥3.0% prevalence) reductions were observed for the three loci with highest crude count (KL2, KL102 and KL15), although these remained in the top-4 even after cluster-adjustment (see

Appendix S2.4). The next most affected by cluster-adjustment were KL104, KL157, KL81, KL108, and KL149 (0.5–2.5% reduction); these were ranked 6, 16, 9, 11, 8 based on raw counts but only KL149 was in the top-20 following cluster-adjustment (the other four being ranked 30, 73, 26 and 29, respectively). We also explored sensitivity to the choice of genetic and temporal thresholds used for clustering, and found only minor impact on per-locus prevalence estimates and ranks (**Appendix S2.5–2.7**).

The primary estimates of cumulative coverage reported above were based on raw counts (i.e. observed infections. This is because we do not expect that a vaccine would prevent nosocomial transmission clusters in neonatal units, as these most often represent point source outbreaks due to contaminated equipment or hospital environment, rather than neonate-to-neonate in a transmission chain where prevention of an index case would prevent onward transmission to others. For comparison, estimates of K or O coverage based on cluster-adjusted prevalence are shown in **Appendix S2.8-2.10 and Tables 2, S4, S5**. Overall the cluster-adjusted coverage estimates were a few percentage points lower, as clusters were often associated with common serotypes, although this varied by region.

Our data set is not suitable to assess sensitivity to time of sampling (i.e. the calendar year of sampling for different studies), or to investigate changes in locus prevalence over time within sites, since the contributing studies were conducted in different time periods (see **Figure 1**) and year of isolation is confounded with both study and geographical location. However, to motivate future longitudinal studies addressing this issue, we explored the data from three sites (in Bangladesh, Malawi, and Kilifi) where blood culture isolates had been routinely collected over extended periods. Ranking K loci and O types based on simple pooled prevalence estimates across these three sites from isolates collected up to 2017 (n=512 isolates), the top-20 K loci account for 69.9% (66.9–73.6% per site) of neonatal sepsis isolates collected after 2017 (n=512), and the top-5 O types for 80.3% (73.98–90.4% per site) collected after 2017 (**Figure S6**).

Discussion

In this study we estimated the prevalence of K loci and O types among *K. pneumoniae* causing neonatal sepsis in Africa and South Asia, overall and for each of the four included regions (**Figures 2, 4**). KL2, KL102 and KL25 were the most common K loci, with cluster-adjusted prevalence of 8.3–8.9% each, and together covering 35% of neonatal sepsis isolates (**Figures 2, 3**). Two other K loci, KL15 and KL62, each exceeded 4.5% prevalence, and together with the top three loci accounted for 49% of isolates (**Table 2**).

Our analyses revealed substantial variation in antigen prevalence by region (**Figure 2**), with the global top-5 K loci estimated to cover 56% of neonatal sepsis isolates in Eastern Africa and 44% in Southern Africa, but only 38% and 28% in South Asia and Western Africa, respectively (**Figure 3, Table 2**). This variation was partly driven by differences in geographic distribution of clones carrying distinct K loci (**Appendix S1**). Notably, the global top-20 K loci collectively covered only 67.8% and 50.5% of neonatal sepsis isolates from South Asia and Western Africa, respectively, lower than the 70% threshold used to model the potential impact of a maternal vaccine [10] and proposed as a target in the *K. pneumoniae* vaccine value profile [7]. While the estimated population coverage for any single region could be substantially improved by selecting the top-20 K loci in that region,

any region-focused antigen set would provide sub-optimal coverage in other regions (**Figure S4)**. However, the minimum coverage estimate of 70% could be reached for all regions using a single set of 20 K loci, by selecting the eight highest ranking K loci in each region (**Figure 2, Table S4**).

We report primary coverage estimates based on raw counts, rather than cluster-adjusted counts, as it is not anticipated that a maternal vaccine would reduce *K. pneumoniae* colonisation or nosocomial transmission, rather the expected effect is through protection against invasive disease following colonisation of the neonate. We estimated similar cumulative coverage for neonates who died during followup (**Figure 3, Table 2**), although we were not able to discern how many of these fatalities were a direct result of *K. pneumoniae* sepsis. We were also unable to assess differences in case fatality rate by K locus or O type, due to sample size limitations and the lack of direct assessment of cause of death. Future estimation of KL- or O-specific fatality rates, and invasiveness (i.e. the probability of causing an infection given colonisation), could provide useful context to prioritise antigen targets, as has been shown for *S. pneumoniae* [51]. While capsule overexpression and hypermucoidy in *K. pneumoniae* are known to promote increased virulence [52,53], very few neonatal sepsis isolates in this study carried the associated *rmpADC* hypermucoidy locus (n=15, 0.8%) [54].

The vast majority of isolates included in our analyses were predicted to produce ESBLs and there was little difference in the cumulative coverage of the top-20 K loci among ESBL producing isolates compared to total isolates. Nor was there a substantial coverage difference for the subset of isolates predicted as carbapenemase producers (**Table 2**). We therefore recommend that genomic surveillance efforts to inform vaccine design should aim to capture the largest possible sample size by including all neonatal sepsis cases regardless of AMR. Subgroup analyses for AMR phenotypes of interest can be conducted post-hoc to check for variation in coverage, as we have shown here.

O types are far less diverse than K types in *K. pneumoniae*, and here we estimated the top-5 O types (two dominant O1 subtypes, plus O2α, O2β and O4) collectively covered 86% of total isolates (76–92% per region). Together with the next five most common O types (O3 variants, O5, O13 and O15), we estimate that ~99% of isolates could be covered in all regions (**Figure 4, Table S5**), i.e. a 10-valent O-based vaccine could theoretically protect against ~99% of infections. However, there remains uncertainty about the protective efficacy of antibodies targeting the O antigen, due to the possibility of masking by the capsule [16,17,55,56]. There is evidence that immunisation with an O1 containing vaccine does not protect mice against K2/O1 *K. pneumoniae* (irrespective of hypermucoidy) [16,57], which is concerning given that our data showed KL2, encoding K2 biosynthesis, was the most common K locus. Notably, the recent phase III trial of the ExPEC9V anti-O vaccine targeting extraintestinal pathogenic *Escherichia coli* – a close relative of *K. pneumoniae* that produces similar K and O antigens – was aborted because it was found to be insufficiently effective at preventing invasive *E. coli* disease in older adults in high-income countries (clinical trial reference: NCT04899336).

We caution against broad generalisation of our findings to other target populations or disease manifestations, e.g. healthcare-associated infections in adults in low or high resource settings. Our data show that geographic variation in circulating clones (each with dominant K locus associations) results in geographic variation of K locus prevalence, and it

is well established that similar geographic clone variation exists among adult disease-causing populations [32,44,58]. While our global top-5 K loci were all reported among the top 10 causes of adult bloodstream infections in Oxford (UK) [59], only KL2 and KL62 were among the top 10 causes of bloodstream infections reported from South and Southeast Asia [32]. Additional comparisons are limited because most studies of adult *K. pneumoniae* infections are skewed towards outbreak clusters and/or CP strains, highlighting a key limitation for the design of vaccines to prevent adult disease.

Study limitations

Our K and O prevalence estimates were derived using a robust Bayesian statistical framework applied to cluster-adjusted antigen counts, to limit the impact of stochastic outbreak clusters. Importantly our sensitivity analyses suggest the findings are generally robust to clustering thresholds and study inclusion in the meta-analysis. Nonetheless, there are a number of limitations that warrant careful consideration.

Firstly, the currently available data have several limitations: i) The total sample size (n=1930) is relatively small given the total possible K loci and O types. ii) Our data captured relatively few countries per region (≤3 countries each, except Eastern Africa with n=6 countries), and with unequal representation per region (**Table 1**). As more data are collected from additional countries we expect that the relative ranks of individual K loci may change. We anticipate that the composition of the overall top-5 will remain fairly stable, because the credible intervals for these loci did not substantially overlap those for lower ranked loci (**Figure 2a**), they were common across all regions (**Figure 2b**) and remained stable in our leave-one-out sensitivity analysis (**Appendix S2**). However the other K loci showed substantial geographic variation, and had broadly overlapping credible intervals for estimated prevalence (**Figure 2a**). As a consequence, the composition of the overall top-20 is likely to change as more countries are considered, especially those from Western Africa which is most underrepresented in the current data. iii) The data were collated from different studies and surveillance programs, each using different protocols and likely influenced by local differences in diagnostic standards, stewardship and prophylactic antimicrobial use. Together these differences may have contributed to differential ascertainment bias, e.g. the samples at some sites may be relatively skewed towards more severe cases which could influence antigen profiles if a subset of antigens are associated with more severe disease (although there is currently no evidence to suggest this). iv) Only a subset of studies collected outcome data, and those that did focussed on all-cause mortality within a broad time range (up to 60 days of life or 30 days post enrollment), which means we were unable to directly compare antigen profiles among fatal and non-fatal neonatal sepsis cases, nor to adjust for variation in gestation, birthweight, comorbidities or other risk factors. v) Most sites did not have longitudinal data, and there was confounding of country with time period, hence we could not directly explore temporal fluctuations in KL or O type prevalence. However, we note that the available data indicate that regional differences could not be explained by differences in the calendar timing of samples from different regions (**Appendix 1**), and those sites with longitudinal data showed reasonable stability within individual sites over time (**Figure S6**).

Secondly, our analyses have relied upon genomic antigen predictions rather than direct serological data, which is very challenging to obtain. K typing data for 731 isolates, including a subset of those from the BARNARDS, GBS-COP and MLW studies, indicates a high level of concordance between genomic predictions and serological phenotypes, in particular for isolates carrying one of the global top-20 K loci (84.5% concordance overall, 91.4% for top-20 K loci); however, there is still room for improvement [29]. Additionally, only half of the known K loci correspond to serologically defined K types and/or published glycan structures. While the available data indicate that unique loci correspond to unique polysaccharide structures, it remains unclear if these structures are serologically distinct, meaning the genomic predictions may overestimate the true antigen diversity (and thus underestimate vaccine coverage). Conversely, genetic variation within K loci, i.e. allelic variation within copies of the same K locus type, may create differences in antibody binding, meaning the genomic predictions may underestimate antigen diversity (and thus overestimate vaccine coverage). Ongoing work from our team and collaborators aims to better understand the link between K genotype and phenotype (serological, structural, and immunological), to support continued improvements for genomic predictions. This will include the generation of additional serological phenotype data. However, it is not feasible to serotype all isolates because the process is labour intensive and currently only available at a single reference laboratory at the Statens Serum Institut (Denmark), which poses resource constraints and accessibility issues e.g. due to shipping and export permit requirements for Nagoya protocol compliance. Similarly, there is currently no broadly accessible O antigen typing protocol, and although the genetic determinants of O serotypes are considered comparatively very well understood [22], there are no data available with which to perform a formal assessment of the predictive accuracy of our genomic approaches. There is therefore a clear need for low cost and broadly accessible K and O serotyping approaches to support sero-epidemiology analysis, that could inform preliminary vaccine design and ongoing surveillance to monitor for population shifts following vaccine introduction.

Thirdly, our analyses have not attempted to capture information about potentially cross-reactive serotypes. For example, there is converging evidence that antibodies raised against individual O1 or O2 subtypes may be cross-protective against other O1 and O2 subtypes, respectively [16,17]. In that regard, the total number of O antigens required to reach maximal population coverage may be lower than those reported here which are based on individual subtypes. Similarly, there is emerging evidence that antibodies raised against serologically distinct K types may have cross-binding and cross-protective activity [60]. More work is needed to fully assess these cross-reactions and understand their relevance for protection against *K. pneumoniae* disease. However, if cross-protection is confirmed for one or more of the top-ranked K loci, the population coverage we have reported would be considered underestimates.

Finally, while this work has focussed solely on polysaccharides, we acknowledge that protein antigens are likely to be much less variable across the population. However, polysaccharide vaccines are currently more popular vaccine targets, as polyvalent *K. pneumoniae* K vaccines have been shown to be safe and immunogenic in humans [56], and single or bivalent anti-K vaccines can be effective in preventing invasive *K. pneumoniae* infection in murine models [61,62]. While protein-based vaccines are also in development [63], these are unlikely to be suitable for a maternal vaccine to be deployed in low-income countries as they typically require more than one dose to induce protection, and the timeframe for

immunisation in pregnancy is limited, especially given that up to half of neonates who develop sepsis are born preterm [3].

Concluding remarks

Our data suggest that it could be possible to achieve a target 70% population coverage of circulating K serotypes in all key geographic regions with a single 20-valent anti-K vaccine, particularly if there turns out to be cross-reactivity between antigens produced by distinct K loci. Greater coverage would likely be achieved, with lower valency, using locally-targeted formulations. However, due to the complex nature of glycoconjugate production, development of multiple high-valency formulations would generate additional manufacturing and regulatory complexities with important cost and logistical considerations. On the other hand, this scenario may represent an important opportunity to support and expand local manufacturing capabilities and processes, of the kind demonstrated to great effect during the COVID-19 pandemic [64,65]. In either scenario, additional K type data will be required to support antigen prioritisation and ensure adequate population coverage per region, e.g. including a greater number of countries and sites per region, particularly for those regions that were underrepresented in our meta-analysis. Longitudinal sampling will also be essential to understand the extent of temporal fluctuations and the impact on theoretical vaccine coverage. Our preliminary analysis suggests a level of stability sufficient to maintain 20-valent vaccine coverage >70% over a 5-10 year timeframe, which is very encouraging but should be considered with caution until further data are available. An O antigen based vaccine could theoretically provide much greater population coverage than a K based vaccine and at lower valency. However, much more data is needed to understand the potential efficacy of such a vaccine, given growing evidence about capsular masking. A combination K and O vaccine may offer an alternative solution, for example if it is found that some key K antigens are more or less likely than others to mask O antigens.

Acknowledgments

We thank the participants and their families in all the contributing studies, and the clinical and laboratory staff involved in collection and processing of relevant samples and isolates. This work was supported by Monash eResearch capabilities, including M3. We would also like to acknowledge the expert support of the sequencing and computational genomics teams at the CHOP Microbiome Core (Philadelphia), CHRF (Dhaka), Infectious Diseases Research Laboratory (Karachi), International Livestock Research Institute (Nairobi), Kilimanjaro Clinical Research Institute (Moshi), Neuberg Center for Genomic Medicine (Ahmedabad), NICD (Johannesburg), Quadram Institute (Norwich), Wellcome Centre for Human Genetics (Oxford), and Wellcome Sanger Institute (Cambridge); and the BARNARDS microbiology and genomics team at Cardiff University. We

References

1. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404: 1199–1226. doi:10.1016/S0140-6736(24)01867-1

2. Sands K, Carvalho MJ, Portal E, Thomson K, Dyer C, Akpulu C, et al. Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. Nat Microbiol. 2021;6: 512–523. doi:10.1038/s41564-021-00870-7

3. Russell NJ, Stöhr W, Plakkal N, Cook A, Berkley JA, Adhisivam B, et al. Patterns of antibiotic use, pathogens, and prediction of mortality in hospitalized neonates and young infants with sepsis: A global neonatal sepsis observational cohort study (NeoOBS). PLOS Med. 2023;20: e1004179. doi:10.1371/journal.pmed.1004179

4. Okomo U, Akpalu ENK, Doare KL, Roca A, Cousens S, Jarde A, et al. Aetiology of invasive bacterial infection and antimicrobial resistance in neonates in sub-Saharan Africa: a systematic review and meta-analysis in line with the STROBE-NI reporting guidelines. Lancet Infect Dis. 2019;19: 1219–1234. doi:10.1016/S1473-3099(19)30414-1

5. Thomson KM, Dyer C, Liu F, Sands K, Portal E, Carvalho MJ, et al. Effects of antibiotic resistance, drug target attainment, bacterial pathogenicity and virulence, and antibiotic access and affordability on outcomes in neonatal sepsis: an international microbiology and drug evaluation prospective substudy (BARNARDS). Lancet Infect Dis. 2021;21: 1677–1688. doi:10.1016/S1473-3099(21)00050-5

6. Verani JR, Blau DM, Gurley ES, Akelo V, Assefa N, Baillie V, et al. Child deaths caused by Klebsiella pneumoniae in sub-Saharan Africa and south Asia: a secondary analysis of Child Health and Mortality Prevention Surveillance (CHAMPS) data. Lancet Microbe. 2024;5: e131–e141. doi:10.1016/S2666-5247(23)00290-2

7. Dangor Z, Benson N, Berkley JA, Bielicki J, Bijsma MW, Broad J, et al. Vaccine value profile for *Klebsiella pneumoniae*. Vaccine. 2024;42: S125–S141. doi:10.1016/j.vaccine.2024.02.072

8. Robinson ML, Johnson J, Naik S, Patil S, Kulkarni R, Kinikar A, et al. Maternal Colonization Versus Nosocomial Transmission as the Source of Drug-Resistant Bloodstream Infection in an Indian Neonatal Intensive Care Unit: A Prospective Cohort Study. Clin Infect Dis. 2023;77: S38–S45. doi:10.1093/cid/ciad282

9. Frantzis I, Levasseur S, Huebner J, Mahida M, Larussa P, James W, et al. Infection prevention and control and related practices in African neonatal units: The Pan-African neonatal care assessment study (PANCAS). Int J Hyg Environ Health. 2024;259: 114357. doi:10.1016/j.ijheh.2024.114357

10. Kumar CK, Sands K, Walsh TR, O'Brien S, Sharland M, Lewnard JA, et al. Global, regional, and national estimates of the impact of a maternal Klebsiella pneumoniae vaccine: A Bayesian modeling analysis. PLoS Med. 2023;20: e1004239.

doi:10.1371/journal.pmed.1004239

11. Estimating the impact of vaccines in reducing antimicrobial resistance and antibiotic use: technical report. World Health Organization; 2024. Available: https://www.who.int/publications/i/item/9789240098787

12. Ahmed T, Tauheed I, Hoque S, Sarower Bhuyan G, Biswas R, Tarikul Islam M, et al. A phase 3 non-inferiority trial of locally manufactured Meningococcal ACWY vaccine "Ingovax ACWY" among Bangladeshi adults. Vaccine. 2024;42: 126063. doi:10.1016/j.vaccine.2024.06.030

13. Lucinde RK, Ong'ayo G, Houlihan C, Bottomley C, Goldblatt D, Scott J a. G, et al. Pneumococcal conjugate vaccine dose-ranging studies in humans: A systematic review. Vaccine. 2021;39: 5095–5105. doi:10.1016/j.vaccine.2021.07.033

14. Gloeck NR, Leong TD, Mthethwa M, Iwu-Jaja CJ, Katoto PD, Wiysonge CS, et al. Typhoid conjugate vaccines for preventing typhoid fever (enteric fever). Cochrane Database Syst Rev. 2025;5: CD015746. doi:10.1002/14651858.CD015746.pub2

15. Male V, Jones CE. Vaccination in pregnancy to protect the newborn. Nat Rev Immunol. 2025; 1–13. doi:10.1038/s41577-025-01162-5

16. Wantuch PL, Knoot CJ, Robinson LS, Vinogradov E, Scott NE, Harding CM, et al. Heptavalent O-Antigen Bioconjugate Vaccine Exhibiting Differential Functional Antibody Responses Against Diverse Klebsiella pneumoniae Isolates. J Infect Dis. 2024;230: 578–589. doi:10.1093/infdis/jiae097

17. Hwang W, Wantuch PL, Bernshtein B, Zhiteneva JA, Slater DM, Vater KH, et al. Antibody responses in Klebsiella pneumoniae bloodstream infection: a prospective cohort study. Lancet Microbe. 2025;6: 100988. doi:10.1016/j.lanmic.2024.100988

18. Bennett JC, Knoll MD, Kagucia EW, Quesada MG, Zeger SL, Hetrich MK, et al. Global impact of ten-valent and 13-valent pneumococcal conjugate vaccines on invasive pneumococcal disease in all ages (the PSERENADE project): a global surveillance analysis. Lancet Infect Dis. 2024;25: 457–470. doi:10.1016/S1473-3099(24)00665-0

19. Ganaie FA, Beall BW, Yu J, van der Linden M, McGee L, Satzke C, et al. Update on the evolving landscape of pneumococcal capsule types: new discoveries and way forward. Clin Microbiol Rev. 2025;38: e00175-24. doi:10.1128/cmr.00175-24

20. Vestjens SMT, van Mens SP, Meek B, Lalmahomed TA, de Jong B, Goswami D, et al. Streptococcus pneumoniae serotype distribution in Bangladeshi under-fives with community-acquired pneumonia pre-10-valent pneumococcal conjugate vaccination. Pneumonia. 2024;16: 29. doi:10.1186/s41479-024-00152-w

21. Stanton TD, Hetland MAK, Löhr IH, Holt KE, Wyres KL. Fast and accurate in silico antigen typing with Kaptive 3. Microb Genomics. 2025;11: 001428. doi:10.1099/mgen.0.001428

22. Whitfield C, Kelly SD, Stanton TD, Wyres KL, Clarke BR, Forrester TJB, et al. O-antigen

polysaccharides in Klebsiella pneumoniae: structures and molecular basis for antigenic diversity. Microbiol Mol Biol Rev. 2025;89: e00090-23. doi:10.1128/mmbr.00090-23

23. Wick RR, Heinz E, Holt KE, Wyres KL. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for Klebsiella Genomes. J Clin Microbiol. 2018;56: 10.1128/jcm.00197-18. doi:10.1128/jcm.00197-18

24. Lam MMC, Wick RR, Judd LM, Holt KE, Wyres KL. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. Microb Genomics. 2022;8: 000800. doi:10.1099/mgen.0.000800

25. Pan Y-J, Lin T-L, Chen C-T, Chen Y-Y, Hsieh P-F, Hsu C-R, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of Klebsiella spp. Sci Rep. 2015;5: 15573. doi:10.1038/srep15573

26. Bellich B, Lagatolla C, Rizzo R, D'Andrea MM, Rossolini GM, Cescutti P. Determination of the capsular polysaccharide structure of the Klebsiella pneumoniae ST512 representative strain KPB-1 and assignments of the glycosyltransferases functions. Int J Biol Macromol. 2020;155: 315–323. doi:10.1016/j.ijbiomac.2020.03.196

27. Bellich B, Ravenscroft N, Rizzo R, Lagatolla C, D'Andrea MM, Rossolini GM, et al. Structure of the capsular polysaccharide of the KPC-2-producing Klebsiella pneumoniae strain KK207-2 and assignment of the glycosyltransferases functions. Int J Biol Macromol. 2019;130: 536–544. doi:10.1016/j.ijbiomac.2019.02.128

28. Kubler-Kielb J, Vinogradov E, Ng W-I, Maczynska B, Junka A, Bartoszewicz M, et al. The capsular polysaccharide and lipopolysaccharide structures of two carbapenem resistant Klebsiella pneumoniae outbreak isolates. Carbohydr Res. 2013;369: 6–9. doi:10.1016/j.carres.2012.12.018

29. Stanton T, Jacobsen M, Gymoese P, Basu S, Chan G, Dangor Z, et al. Technical Report: Accuracy of Klebsiella pneumoniae capsule serotype predictions from whole genome sequences. Zenodo; 2025 Jun. Available: https://zenodo.org/records/15742130

30. Heinz E, Pearse O, Zuza A, Bilima S, Msefula C, Musicha P, et al. Longitudinal analysis within one hospital in sub-Saharan Africa over 20 years reveals repeated replacements of dominant clones of Klebsiella pneumoniae and stresses the importance to include temporal patterns for vaccine design considerations. Genome Med. 2024;16: 67. doi:10.1186/s13073-024-01342-3

31. Shamanna V, Srinivas S, Couto N, Nagaraj G, Sajankila SP, Krishnappa HG, et al. Geographical distribution, disease association and diversity of Klebsiella pneumoniae K/L and O antigens in India: roadmap for vaccine development. Microb Genomics. 2024;10: 001271. doi:10.1099/mgen.0.001271

32. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, Dance DAB, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. Genome Med. 2020;12: 11. doi:10.1186/s13073-019-0706-y

33. Chung The H, Karkey A, Pham Thanh D, Boinett CJ, Cain AK, Ellington M, et al. A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of Klebsiella pneumoniae. EMBO Mol Med. 2015;7: 227–239. doi:10.15252/emmm.201404767

34. Mashau RC, Meiring ST, Dramowski A, Magobo RE, Quan VC, Perovic O, et al. Culture-confirmed neonatal bloodstream infections and meningitis in South Africa, 2014–19: a cross-sectional study. Lancet Glob Health. 2022;10: e1170–e1178. doi:10.1016/S2214-109X(22)00246-7

35. Egbe FN, Cowden C, Mwananyanda L, Pierre C, Mwansa J, Lukwesa Musyani C, et al. Etiology of Bacterial Sepsis and Isolate Resistance Patterns in Hospitalized Neonates in Zambia. Pediatr Infect Dis J. 2023;42: 921–926. doi:10.1097/INF.0000000000004008

36. Aiken AM, Rehman AM, Kraker MEA de, Madrid L, Kebede M, Labi A-K, et al. Mortality associated with third-generation cephalosporin resistance in Enterobacterales bloodstream infections at eight sub-Saharan African hospitals (MBIRA): a prospective cohort study. Lancet Infect Dis. 2023;23: 1280–1290. doi:10.1016/S1473-3099(23)00233-5

37. Jain K, Kumar V, Plakkal N, Chawla D, Jindal A, Bora R, et al. Multidrug-resistant sepsis in special newborn care units in five district hospitals in India: a prospective cohort study. Lancet Glob Health. 2025;13: e870–e878. doi:10.1016/S2214-109X(24)00564-3

38. Olwagen CP, Izu A, Khan S, Van der Merwe L, Dean NJ, Mabena FC, et al. Genomic relatedness of colonizing and invasive disease Klebsiella pneumoniae isolates in South African infants. Sci Rep. 2025;15: 8043. doi:10.1038/s41598-025-92517-4

39. Hooda Y, Tanmoy AM, Kanon N, Rahman H, Islam MS, Ahmed ZB, et al. Escalating carbapenem-resistant Klebsiella burden and mortality in Bangladeshi infants. medRxiv; 2025. p. 2025.06.11.25329357. doi:10.1101/2025.06.11.25329357

40. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun. 2021;12: 4188. doi:10.1038/s41467-021-24448-3

41. United Nations Statistics Division. Standard country or area codes for statistical use (M49). [cited 15 Mar 2025]. Available: https://unstats.un.org/unsd/methodology/m49/

42. Argimón S, David S, Underwood A, Abrudan M, Wheeler NE, Kekre M, et al. Rapid Genomic Characterization and Global Surveillance of Klebsiella Using Pathogenwatch. Clin Infect Dis. 2021;73: S325–S335. doi:10.1093/cid/ciab784

43. Sherry NL, Lane CR, Kwong JC, Schultz M, Sait M, Stevens K, et al. Genomics for Molecular Epidemiology and Detecting Transmission of Carbapenemase-Producing Enterobacterales in Victoria, Australia, 2012 to 2016. J Clin Microbiol. 2019;57: 10.1128/jcm.00573-19. doi:10.1128/jcm.00573-19

44. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol. 2019;4: 1919–1929. doi:10.1038/s41564-019-0492-8

45. Gorrie CL, Mirčeta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. Clin Infect Dis. 2017;65: 208–215. doi:10.1093/cid/cix270

46. Foster-Nyarko E, Cottingham H, Wick RR, Judd LM, Lam MMC, Wyres KL, et al. Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, Klebsiella pneumoniae. Microb Genomics. 2023;9: 000936. doi:10.1099/mgen.0.000936

47. Antonov M, Csárdi G, Horvát S, Müller K, Nepusz T, Noom D, et al. igraph enables fast and robust network analysis across programming languages. arXiv [Preprint]; 2023. doi:10.48550/arXiv.2311.10260

48. Bürkner P-C. Advanced Bayesian Multilevel Modeling with the R Package brms. R J. 2018;10: 395–411.

49. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. J Stat Softw. 2017;76: 1–32. doi:10.18637/jss.v076.i01

50. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available: https://www.r-project.org/

51. 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. J Infect Dis. 2004;190: 1203–1211. doi:10.1086/423820

52. Russo TA, Carlino-MacDonald U, Drayer ZJ, Davies CJ, Alvarado CL, Hutson A, et al. Deciphering the relative importance of genetic elements in hypervirulent Klebsiella pneumoniae to guide countermeasure development. eBioMedicine. 2024;107. doi:10.1016/j.ebiom.2024.105302

53. Lee H-C, Chuang Y-C, Yu W-L, Lee N-Y, Chang C-M, Ko N-Y, et al. Clinical implications of hypermucoviscosity phenotype in Klebsiella pneumoniae isolates: association with invasive syndrome in patients with community-acquired bacteraemia. J Intern Med. 2006;259: 606–614. doi:10.1111/j.1365-2796.2006.01641.x

54. Lam MMC, Salisbury SM, Treat LP, Wick RR, Judd LM, Wyres KL, et al. Genomic and functional analysis of rmp locus variants in Klebsiella pneumoniae. Genome Med. 2025;17: 36. doi:10.1186/s13073-025-01461-5

55. Held TK, Jendrike NR, Rukavina T, Podschun R, Trautmann M. Binding to and opsonophagocytic activity of O-antigen-specific monoclonal antibodies against encapsulated and nonencapsulated Klebsiella pneumoniae serotype O1 strains. Infect Immun. 2000;68: 2402–2409. doi:10.1128/IAI.68.5.2402-2409.2000

56. Miller JC, Cross AS, Tennant SM, Baliban SM. Klebsiella pneumoniae Lipopolysaccharide as a Vaccine Target and the Role of Antibodies in Protection from Disease. Vaccines. 2024;12: 1177. doi:10.3390/vaccines12101177

57. Wantuch PL, Knoot CJ, Robinson LS, Vinogradov E, Scott NE, Harding CM, et al. Capsular polysaccharide inhibits vaccine-induced O-antigen antibody binding and function across both classical and hypervirulent K2:O1 strains of Klebsiella pneumoniae. PLoS Pathog. 2023;19: e1011367. doi:10.1371/journal.ppat.1011367

58. Wang M, Earley M, Chen L, Hanson BM, Yu Y, Liu Z, et al. Clinical outcomes and bacterial characteristics of carbapenem-resistant Klebsiella pneumoniae complex among patients from different global regions (CRACKLE-2): a prospective, multicentre, cohort study. Lancet Infect Dis. 2022;22: 401–412. doi:10.1016/S1473-3099(21)00399-6

59. Lipworth S, Vihta K-D, Chau KK, Kavanagh J, Davies T, George S, et al. Ten Years of Population-Level Genomic Escherichia coli and Klebsiella pneumoniae Serotype Surveillance Informs Vaccine Development for Invasive Infections. Clin Infect Dis. 2021;73: 2276–2282. doi:10.1093/cid/ciab006

60. Nonne F, Molfetta M, Belciug GF, Carducci M, Cianchi V, Zakroff C, et al. The characterization of Klebsiella pneumoniae associated with neonatal sepsis in low- and middle-income countries to inform vaccine design. Commun Biol. 2025;8: 1–12. doi:10.1038/s42003-025-08258-7

61. Wantuch PL, Knoot CJ, Marino EC, Harding CM, Rosen DA. Klebsiella pneumoniae bioconjugate vaccine functional durability in mice. Vaccine. 2025;43: 126536. doi:10.1016/j.vaccine.2024.126536

62. Feldman MF, Mayer Bridwell AE, Scott NE, Vinogradov E, McKee SR, Chavez SM, et al. A promising bioconjugate vaccine against hypervirulent Klebsiella pneumoniae. Proc Natl Acad Sci. 2019;116: 18655–18663. doi:10.1073/pnas.1907833116

63. Choi M, Tennant SM, Simon R, and Cross AS. Progress towards the development of Klebsiella vaccines. Expert Rev Vaccines. 2019;18: 681–691. doi:10.1080/14760584.2019.1635460

64. Doua J, Ndembi N, Auerbach J, Kaseya J, Zumla A. Advancing local manufacturing capacities for vaccines within Africa – Opportunities, priorities and challenges. Vaccine. 2025;50: 126829. doi:10.1016/j.vaccine.2025.126829

65. Otu A, Osifo-Dawodu E, Atuhebwe P, Agogo E, Ebenso B. Beyond vaccine hesitancy: time for Africa to expand vaccine manufacturing capacity amidst growing COVID-19 vaccine nationalism. Lancet Microbe. 2021;2: e347–e348. doi:10.1016/S2666-5247(21)00126-9

Author Contributions

Conceptualization: Kathryn E Holt, Kelly L Wyres
Data Curation: Jabir A Abdulahi, Anne V Amulele, Matthew Bates, Eva Heinz, Yogesh Hooda, Weiming Hu, Kajal Jain, Samiah Kanwar, Rindidzani Magobo, Courtney P Olwagen, John M Tembo, Tolbert Sonda, Jonathan Strysko, Caroline C Tigoi
Formal Analysis: Shaun P Keegan, Thomas D Stanton, Kathryn E Holt, Kelly L Wyres
Funding Acquisition: Alexander M Aiken, James A Berkley, Susan E Coffin, Nicholas A Feasey, Nelesh P Govender, Davidson H Hamer, Shabir A Madhi, M Imran Nisar, Samir K Saha, Senjuti Saha, M Jeeva Sankar, Ramesh K Agarwal, Kathryn E Holt, Kelly L Wyres
Investigation: Sameen Ahmad Amin, Kyle Bittinger, Jennifer Cornick, Ebenezer Foster-Nyarko, Wilson Gumbi, Aneeta Hotwani, Naveed Iqbal, Steven M Jones, Furqan Kabir, Waqasuddin Khan, Chileshe L Musyani, Carolyn M McGann, Varsha Mittal, Ahmed M Moustafa, Patrick Musicha, James CL Mwansa, Moreka L Ndumba, Erkison E Odih, Donwilliams O Omuoyo, Oliver Pearse, Laura T Phillips, Paul J Planet, Aniqa Abdul Rasool, Charlene MC Rodrigues, Kirsty Sands, Arif M Tanmoy, Erin Theiller, Allan M Zuza
Methodology: Shaun P Keegan, Thomas D Stanton, Kathryn E Holt, Kelly L Wyres
Resources: Sulagna Basu, Grace J Chan, Kenneth C Iregbu, Jean-Baptiste Mazarati, Semaria Solomon Alemayehu, Timothy R Walsh, Rabaab Zahra, Kajal Jain, Angela Dramowski, Sombo Fwoloshi, Appiah-Korang Labi, Lola Madrid, Noah Obeng-Nkrumah, David Ojok, Boaz D Wadugu, Andrew C Whitelaw, Adhisivam Bethou, Anudita Bhargava, Atul Jindal, Ruchi N Nanavati, Priyanka S Prasad, Apurba Sastry, Joveria Q Farooqi, Najia Ghanchi, Fyezah Jehan, Erum Khan, Alexander M Aiken, James A Berkley, Susan E Coffin, Nicholas A Feasey, Nelesh P Govender, Davidson H Hamer, Shabir A Madhi, M Imran Nisar, Samir K Saha, Senjuti Saha, M Jeeva Sankar, Ramesh K Agarwal
Software: Shaun P Keegan, Thomas D Stanton, Kathryn E Holt, Kelly L Wyres
Supervision: Sulagna Basu, Grace J Chan, Kenneth C Iregbu, Jean-Baptiste Mazarati, Semaria Solomon Alemayehu, Timothy Walsh, Rabaab Zahra, Kajal Jain, Angela Dramowski, Sombo Fwoloshi, Appiah-Korang Labi, Lola Madrid, Noah Obeng-Nkrumah, David Ojok, Boaz D Wadugu, Andrew C Whitelaw, Adhisivam Bethou, Anudita Bhargava, Atul Jindal, Ruchi N Nanavati, Priyanka S Prasad, Apurba Sastry, Joveria Q Farooqi, Najia Ghanchi, Fyezah Jehan, Erum Khan, Alexander M Aiken, James A Berkley, Susan E Coffin, Nicholas A Feasey, Nelesh P Govender, Davidson H Hamer, Shabir A Madhi, M Imran Nisar, Samir K Saha, Senjuti Saha, M Jeeva Sankar, Ramesh K Agarwal
Visualization: Shaun P Keegan, Thomas D Stanton, Kathryn E Holt, Kelly L Wyres
Writing – Original Draft Preparation: Kathryn E Holt, Kelly L Wyres
Writing – Review & Editing: All authors.

Financial Disclosure Statement

This study was supported by the Gates Foundation (grants INV049364, INV025280, INV077266 to KEH; INV049641 to KLW; INV003519 to AAiken; INV005180 to NAF; INV041685 to JAB; INV008112 to NPG; INV005567 to CT; INV005691 to DHH; INV023821 to Senjutia Saha, INV005773 to SAM, INV046521 and INV080845 to MIN, INV065400 to AMM, INV065400 to SC), the Fleming Fund (grant FF25-286 to TSonda), the Center for AIDS Research (core support grant to SC), and the Wellcome Trust (grant 217303/Z/19/Z to EH, core support grants 206194 to WSI, 206454 to MLW, 203077/Z/16/Z to KWTRP). KLW is supported by a National Health and Medical Research Council of Australia Investigator Grant (APP1176192). Pfizer sponsored the sequencing of a subset of isolates from the GBS-COP study for WITS-Vida. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The conclusions and opinions expressed in this work are those of the author(s) alone and shall not be attributed to any funder. Under the grant conditions of the Gates Foundation and Wellcome Trust, a Creative Commons Attribution 4.0 License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. Please note works submitted as a preprint have not undergone a peer review process.