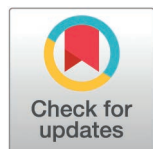


RESEARCH ARTICLE

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

Thomas D. Stanton<sup>1,2</sup>, Shaun P. Keegan<sup>3</sup>, Jabir A. Abdulahi<sup>4</sup>, Anne V. Amulele<sup>5</sup>, Matthew Bates<sup>6</sup>, Eva Heinz<sup>7,8</sup>, Yogesh Hooda<sup>9</sup>, Weiming Hu<sup>10</sup>, Kajal Jain<sup>11</sup>, Samiah Kanwar<sup>12</sup>, Rindidzani Magobo<sup>13,14</sup>, Courtney P. Olwage<sup>15</sup>, John M. Tembo<sup>16</sup>, Tolbert Sonda<sup>17</sup>, Jonathan Stryko<sup>18,19</sup>, Caroline C. Tigoi<sup>5</sup>, Sameen Ahmad Amin<sup>20</sup>, Kyle Bittinger<sup>10,21,22</sup>, Jennifer Cornick<sup>23,24</sup>, Ebenezer Foster-Nyarko<sup>5</sup>, Wilson Gumbi<sup>5</sup>, Aneeta Hotwani<sup>12</sup>, Naveed Iqbal<sup>12,25</sup>, Steven M. Jones<sup>10</sup>, Furqan Kabir<sup>26</sup>, Waqasuddin Khan<sup>27</sup>, Chileshe L. Musyani<sup>28,29</sup>, Carolyn M. McGann<sup>21,30</sup>, Varsha Mittal<sup>11</sup>, Ahmed M. Moustafa<sup>10,21,22</sup>, Patrick Musicha<sup>24,31</sup>, James C.L. Mwansa<sup>32</sup>, Moreka L. Ndumba<sup>5</sup>, Erkison E. Odih<sup>3</sup>, Donwilliams O. Omuoyo<sup>5</sup>, Oliver Pearse<sup>24,31</sup>, Laura T. Phillips<sup>3</sup>, Paul J. Planet<sup>22,33</sup>, Anika Abdul Rasool<sup>12</sup>, Charlene M. C. Rodrigues<sup>3,34,35</sup>, Kirsty Sands<sup>36,37</sup>, Arif M. Tanmoy<sup>9</sup>, Erin Theiller<sup>22</sup>, Allan M. Zuzva<sup>24,38</sup>, Sulagna Basu<sup>39</sup>, Grace J. Chan<sup>21,40,41</sup>, Kenneth C. Iregbu<sup>42,43</sup>, Jean-Baptiste Mazarati<sup>44,45</sup>, Semaria Solomon Alemayehu<sup>46</sup>, Timothy R. Walsh<sup>36,47</sup>, Rabaab Zahra<sup>48,49</sup>, Angela Dramowski<sup>50</sup>, Sombo Fwoloshi<sup>51,52</sup>, Appiah-Korang Labi<sup>53</sup>, Lola Madrid<sup>4,54</sup>, Noah Obeng-Nkrumah<sup>55</sup>, David Ojok<sup>56</sup>, Boaz D. Wadugu<sup>57</sup>, Andrew C. Whitelaw<sup>58,59</sup>, Adhisivam Bethou<sup>60</sup>, Anudita Bhargava<sup>61</sup>, Atul Jindal<sup>62</sup>, Ruchi N. Nanavati<sup>63</sup>, Priyanka S. Prasad<sup>64</sup>, Apurba Sastry<sup>65</sup>, Joveria Q. Farooqi<sup>66</sup>, Najia Ghanchi<sup>66</sup>, Fyezah Jehan<sup>12</sup>, Erum Khan<sup>66</sup>, Ramesh K. Agarwal<sup>11</sup>, Alexander M. Aiken<sup>4</sup>, James A. Berkley<sup>67,68</sup>, Susan E. Coffin<sup>21,40</sup>, Nicholas A. Feasey<sup>24,31,38</sup>, Nelesh P. Govender<sup>69,70,71</sup>, Davidson H. Hamer<sup>72,73</sup>, Shabir A. Madhi<sup>15,71</sup>, Muhammad Imran Nisar<sup>12</sup>, Samir K. Saha<sup>9</sup>, Senjuti Saha<sup>9</sup>, Mari Jeeva Sankar<sup>11</sup>, Kelly L. Wyres<sup>1,3,\*</sup>, Kathryn E. Holt<sup>1,3,\*</sup>



## OPEN ACCESS

**Citation:** Stanton TD, Keegan SP, Abdulahi JA, Amulele AV, Bates M, Heinz E, et al. (2026) Distribution of capsule and O types in *Klebsiella pneumoniae* causing neonatal sepsis in Africa and South Asia: A meta-analysis of genome-predicted serotype prevalence to inform potential vaccine coverage. PLoS Med 23(1): e1004879. <https://doi.org/10.1371/journal.pmed.1004879>

**Academic Editor:** Rebecca F. Grais, Pasteur Network, FRANCE

**Received:** June 27, 2025

**Accepted:** December 16, 2025

**Published:** January 12, 2026

**Copyright:** © 2026 Stanton et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** Sample-level data, including K/O calls, accessions, and source information are given in [S3 Table](#). Raw and adjusted counts data used for modelling, and

**1** Department of Infectious Diseases, School of Translational Medicine, Monash University, Melbourne, Australia, **2** Centre to Impact AMR, Monash University, Clayton, Victoria, Australia, **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 Department, 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, United Kingdom, **8** Parasites and Microbes Program, Wellcome Sanger Institute, Hinxton, United Kingdom, **9** Child Health Research Foundation, Dhaka, Bangladesh, **10** Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, United States of America, **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, **15** South African Medical Research Council Vaccines and Infectious Diseases Analytics Research Unit, University of the Witwatersrand, Johannesburg, South Africa, **16** Department of 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, United States of America, **20** CITRIC Center for Bioinformatics and Computational Biology, Department of Pediatrics and Child Health, The Aga Khan University, Karachi, Pakistan, **21** Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania,

model code and outputs, are available at <https://github.com/klebgenomics/KlebNNSsro> (DOI: <https://doi.org/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 public databases, specifically the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) under BioProjects PRJEB33565 (BARNARDS), PRJEB46513 (SPINZ); PRJEB42462 (MLW), PRJEB90555 (CHRF), PRJEB70311 (DH and NeoOBS-India); and the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) under BioProjects PRJNA1274034 (MBIRA), PRJNA1265413 (NeoBAC), PRJNA1261643 (AKU), PRJNA796486 (BabyGERMS), PRJNA1282934 (BabyGERMS), PRJNA1175467 (GBS-COP), PRJNA1265413 (Kilifi), and PRJNA1168993 (NIMBI-plus).

**Funding:** This study was supported by the Gates Foundation (<https://www.gatesfoundation.org/>) (grants INV049364, INV025280, and INV077266 to K.E.H.; INV049641 to K.L.W.; INV003519 to A.M.K.; INV005180 to N.A.F.; INV041685 to J.A.B.; INV008112 to N.P.G.; INV005567 to C.C.T.; INV005691 to D.H.H.; INV023821 to SenS, INV005773 to S.A.M., INV046521 and INV080845 to M.I.N., INV065400 to A.M.M., INV065400 to S.E.C.), the Fleming Fund (<https://www.flemingfund.org/>) (grant FF25-286 for SeqAfrica to support sequencing capacity at KCRI), the Center for AIDS Research (<https://www.niaid.nih.gov/research/centers-aids-research>) (core support grant to S.E.C.), the Thrasher Research Fund (<https://www.thrasherresearch.org/>) (grant #12036 to D.H.H.), the Wellcome Trust (<https://wellcome.org/>) (grant 217303/Z/19/Z to E.H., core support grants 206194 to Wellcome Sanger Institute, 206454 to M.L.W., 203077/Z/16/Z to K.W.T.R.P.), and the National Health and Medical Research Council of Australia (<https://www.nhmrc.gov.au/>) (APP1176192 to K.L.W.). Pfizer (<https://www.pfizer.com/>) sponsored the sequencing of a subset of isolates from the GBS-COP

Philadelphia, Pennsylvania, United States of America, **22** Center for Microbial Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **23** Clinical Infection, Microbiology and Immunology, University of Liverpool, Liverpool, United Kingdom, **24** Malawi Liverpool Wellcome Programme, Kamuzu University of Health Sciences, Blantyre, Malawi, **25** Department of Biotechnology and Informatics, Faculty of Life Sciences & Informatics, BUIITEMS, Quetta, Pakistan, **26** Infectious Diseases Research laboratory, Department of Paediatrics & Child Health, Aga Khan University, Karachi, Pakistan, **27** CITRIC Center for Bioinformatics and Computational Biology, Department of Pediatrics and Child Health, The Aga Khan University, Karachi, Pakistan, **28** Antimicrobial Resistance Unit, Zambia National Public Health Institute, Lusaka, Zambia, **29** Institute of Basic and Biomedical Sciences, Levy Mwanawasa Medical University, Lusaka, Zambia, **30** Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **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, Pennsylvania, United States of America, **34** Department of Paediatrics, Imperial College Healthcare NHS Trust, London, United Kingdom, **35** Pathogen Genomics Programme, UK Health Security Agency, London, United Kingdom, **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, United Kingdom, **38** The School of Medicine, University of St Andrews, St Andrews, United Kingdom, **39** Division of Bacteriology, ICMR-National Institute for Research in Bacterial Infections (formerly ICMR-NICED), Kolkata, India, **40** Division of Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **41** Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **42** Department of Medical Microbiology, College of Health Sciences, University of Abuja, Abuja, Nigeria, **43** Department of Medical Microbiology, National Hospital Abuja, Abuja, Nigeria, **44** Health Sciences, INES-RUHENGRI, Musanze, Rwanda, **45** INES Centre for Genomic Biology, Musanze, Rwanda, **46** Department of Microbiology, Immunology and Parasitology, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia, **47** International Centre for Antimicrobial Resistance Solutions, Copenhagen, Denmark, **48** Department of Microbiology, Quaid-i-Azam University, Islamabad, Pakistan, **49** Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, St James's Hospital, Dublin, Ireland, **50** Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, **51** School of Medicine, University Teaching Hospital, University of Zambia, Lusaka, Zambia, **52** Infectious Diseases Unit, Department of Internal Medicine, UTH, Lusaka, Zambia, **53** Department of Medical Microbiology, University of Ghana Medical School, Accra, Ghana, **54** College of Health and Medical Sciences, Haramaya University, Harar, Ethiopia, **55** Department of Medical Laboratory Sciences, School of Biomedical and Allied Health Sciences, College of Health Sciences, University of Ghana, Accra, Ghana, **56** Central Laboratory, Centre for Infectious Disease Research in Zambia, Lusaka, Zambia, **57** Biotechnology Laboratory, Kilimanjaro Clinical Research Institute, Moshi, Tanzania, **58** Division of Medical Microbiology and Immunology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, **59** National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa, **60** Department of Neonatology, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Puducherry, India, **61** Department of Microbiology, All India Institute of Medical Sciences, Raipur, India, **62** Pediatrics, All India Institute of Medical Sciences, Raipur, India, **63** Department of Neonatology, Seth GS Medical College and KEM Hospital, Mumbai, India, **64** Department of Microbiology, Seth GS Medical College & KEM Hospital, Mumbai, India, **65** Microbiology, JIPMER, Puducherry, India, **66** Department of Pathology and Laboratory Medicine, The Aga Khan University, Karachi, Pakistan, **67** Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **68** Clinical Research Department, KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya, **69** Wits Mycology Division, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, **70** Division of the National Health Laboratory Service, National Institute for Communicable Diseases, Johannesburg, South Africa, **71** Wits Infectious Diseases and Oncology Research Institute, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, **72** Department of Global Health, Boston University School of Public Health, Boston, Massachusetts, United States of America, **73** Section of Infectious Diseases, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, United States of America

☞ These authors contributed equally to this work.

‡ KLW and KEH also contributed equally to this work.

\* [kelly.wyres@monash.edu](mailto:kelly.wyres@monash.edu) (KLW); [kat.holt@lshtm.ac.uk](mailto:kat.holt@lshtm.ac.uk) (KEH)

study for WITS-Vida (S.A.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AKUH, Aga Khan University Hospital; AMR, antimicrobial resistant; CHAMPS, Child Health and Mortality Prevention Surveillance; CI, credible interval; CP, carbapenemase-producing; CSF, cerebrospinal fluid; ESBL, extended-spectrum beta-lactamase; KL, K locus; LMIC, low- and middle-income countries; SNV, single nucleotide variant; STs, sequence types; WHO, World Health Organization.

## Abstract

### Background

*Klebsiella pneumoniae* causes ~20% of sepsis in neonates, with ~40% crude mortality. A vaccine administered to pregnant women, protecting against  $\geq 70\%$  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 1,930 *K. pneumoniae* neonate blood isolates from 13 surveillance studies across 35 sites in 13 countries. We used pathogen whole-genome sequencing to predict K and O serotypes and adjust for local transmission clusters, and Bayesian hierarchical meta-analysis to estimate K and O prevalence overall and per region, treating site as a random effect. 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  $\geq 70\%$  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 (O1 $\alpha\beta$ ,2 $\alpha$ , O1 $\alpha\beta$ ,2 $\beta$ , O2 $\alpha$ , O2 $\beta$ , and O4) would cover 86.2% [82.6, 89.9%] of total infections (76%–92% per region), while the top-10 would cover ~99% of infections in all four regions. The main limitations of our study are the reliance on genome sequences to predict K and O serotypes (as serological typing is not available) and a lack of longitudinal data to explore stability of antigen prevalence over time.

### 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  $\geq 70\%$  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, the protective efficacy against disease 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.

## Author summary

### Why Was This Study Done?

- *Klebsiella pneumoniae* is a leading cause of neonatal sepsis in Africa and South Asia, and it is estimated that a *K. pneumoniae* vaccine administered to pregnant women could potentially avert 80,000 neonatal deaths annually.
- The most promising targets for such a vaccine are capsular (K) and outer-lipopolysaccharide (O) antigens, as similar vaccine types have proven safe and effective against other bacterial pathogens and can elicit protection after a single dose.
- There are many different K and O antigen types in the *K. pneumoniae* population. For a future vaccine to be effective in Africa and South Asia, it must include the antigen types that are most prevalent amongst neonatal sepsis in these regions.

### What Did the Researchers Do and Find?

- We combined genome sequence data from 1,930 *K. pneumoniae* blood isolates collected from neonates across 35 hospitals in 13 countries in Africa and South Asia.
- These genome sequences were analysed to infer the K and O antigen types expressed by the isolated bacteria.
- We modelled the prevalence of each K and O type overall and per region, and found that the five most common K types accounted for 49% of all isolates.
- We explored different combinations of K types, and found that a set of 20 could cover 72.9% of all infections, including at least 70% of the isolates in each of the four target regions (Eastern, Western, and Southern Africa and South Asia).
- We estimated that the five most common O types could cover more than 85% of infections (range: 76%–92% per region).

### What Do These Findings Mean?

- *Klebsiella pneumoniae* associated with neonatal sepsis are genetically diverse, and different strains vary in the antigens they express, making vaccine development challenging.
- A single vaccine against 20 capsular (K) antigens could theoretically cover  $\geq 70\%$  of *K. pneumoniae* neonatal sepsis cases in Africa and South Asia. While complex and costly to manufacture, this is comparable to the existing 20-valent pneumococcal conjugate vaccine.
- Region-specific capsular vaccines could provide higher coverage with fewer antigens, but may be challenging to manufacture and regulate.
- There are some uncertainties about whether O-based vaccines will provide adequate protection from disease, but if so, a single *K. pneumoniae* vaccine with 5–10 O antigens could potentially achieve high (85%–99%) global coverage.
- Further research is needed on antigen cross-reactivity, natural antigen exposure, and stability of antigens within the *K. pneumoniae* population over time, to better guide vaccine development.
- The main limitations of our study are the reliance on genome sequences to predict antigen types, rather than directly typing them (as the required laboratory assays are not broadly available), and the lack of long-term data to see how antigen patterns change over time and how frequently vaccine formulations may need to be updated.



## 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], and a WHO Technical Advisory Group has been established to develop a WHO Research Roadmap for vaccines against *K. pneumoniae*.

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 by blocking access of antibodies to the shorter O polysaccharide [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 [S1 Table](#). 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 written informed consent from the parents or guardians 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 LMIC in Asia and Africa were sourced from prospective clinical studies of neonatal sepsis (BARNARDS [2], SPINZ [35,36], Baby GERMS-SA [34], MBIRA [37], DH [38], GBS-COP [39], NIMBI-plus [40], NeoOBS-India [3], and NeoBAC [41]), long-term prospective surveillance of bloodstream infection (MLW [30], KWTRP [41], and CHRF [42]), and referral-based surveillance of blood cultures in Pakistan (AKU) ([S2 Table](#)).

For the prospective studies (BARNARDS, SPINZ, Baby GERMS-SA, MBIRA, DH, GBS-COP, NIMBI-plus, NeoOBS-India, and 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, and MLW) or CSF cultures (KWTRP and 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  $>37.5^{\circ}\text{C}$ , 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 [S2 Table](#). 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 [43] to confirm species, identify multi-locus sequence types (STs), 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 [S1 Appendix](#)). 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 and 2023. CSF isolates were included from four studies only (NeoOBS-India,  $n=9$ ; KWTRP surveillance,  $n=5$ ; GBS-COP,  $n=2$ ; and 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 ( $\leq 1,000$  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 [S1 Fig](#), 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 [44] (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  $\geq 1$  (indicating detection of at least one known ESBL gene in the genome), and as CP if they had a Kleborate v3 resistance score of  $\geq 2$  (indicating detection of at least one known carbapenemase gene in the genome [43]).

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, and 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) [45]. 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* [46–48]. We recently showed that a genome-wide threshold of 25 SNVs is equivalent to a PW distance of 10 SNVs in *K. pneumoniae* [49]; hence, we used a threshold of PW distance  $\leq 10$  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, and 52 weeks) (reported in [S2 Appendix](#)). Clustering was done in R using the *igraph* package (v2.0.3) [50], 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 generalised hierarchical linear mixed model with a binomial family, with events being the count (raw or cluster-adjusted) of isolates with each locus observed at each site, with fixed effects for locus and region and random effects for site and locus:region. Specifically, the data comprises  $N$  sites and  $K$  loci resulting in  $K \times N$  observations. One model each was fitted across all locus types and sites, for each combination of raw and cluster-adjusted K and O counts (4 models in total). Models were fitted using the *brms* package in R (v2.22.0) [51], which provides an interface to the probabilistic programming language Stan (v2.32.2) [52].

Model specification is shown below. Each grouping  $i$  represents a unique combination of locus, site, and region.

$$y_i \sim \text{Binomial}(n_i, \pi_i)$$

$$\text{logit}(\pi_i) = \beta_x[\text{locus}_i] + \beta_r[\text{region}_i] + \gamma_s[\text{site}_i] + \gamma_{xr}[\text{locus}_i][\text{region}_i]$$

The priors for the fixed effects (*site* and *region*) were a Uniform distribution from  $-\infty$  to  $\infty$  (default *brms* priors). For the standard deviations of the random effects (for *site* and *locus:region*), priors were specified as student- $t$  distributions with 1 degree of freedom, location 0, and scale 10. Each model was fit with 4 chains, each running for 6,000 iterations, with 3,000 burn-in iterations. Control parameters were set to ensure stable sampling, with `adapt_delta=0.999999` and `max_treedepth=55`. Model diagnostics (Rhat, effective sample size ratio, and divergent transitions) were inspected for all runs (see [S3 Appendix](#)). In cases where initial model diagnostics warranted, models were run again with 30,000 iterations with 15,000 burn-in iterations.

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 theoretical vaccine analyses, for each set of K loci (or O types), we obtained posterior distributions by summing the relevant locumed/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 versus 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) [53] with packages *dplyr* (v1.1.4), *tidyverse* (v2.0.0), *ggplot2* (v3.5.1), *ggridges* (v0.5.6), and *patchwork* (v1.3.0).

## Results

A total of 1,930 *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, Figs 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 and 99.5% O), and a total of 87 distinct K loci and 14 O types were identified across the studies.

The raw distribution of K loci is shown in S1.1 Appendix. Four K loci were common, with crude prevalence exceeding 5% each (KL102, KL15, KL2, and 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 12 K loci found in all four regions sampled (KL2, KL3, KL10, KL15, KL17, KL19, KL25, KL39, KL48, KL62, KL102, and 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., ST in S1.1 Appendix). 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 (S1.2 Appendix), consistent with local transmission in the hospital or the wider community. For example, isolates from Ethiopia were dominated by ST35-KL108/O1 $\alpha\beta$ ,2 $\beta$  (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/O2 $\beta$ , 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/O1 $\alpha\beta$ ,2 $\beta$ ; 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 support that observed K (and O) locus prevalence can be heavily influenced by localised nosocomial outbreaks, as noted above.

### 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 1,930 *K. pneumoniae* genomes included from 13 studies in four regions (Fig 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 ( $\leq 10$  SNVs and  $\leq 28$  days) within a site (i.e., the same neonatal unit) is treated as a single observation (see Methods). Five loci (KL2, KL102, KL25, KL15, and KL62) had mean prevalence estimates exceeding 4% each (with lower 95% CI each  $\geq 3.2\%$ , see Fig 2a). A further seven K loci (KL30, KL10, KL17, KL23, KL51, KL112, and KL39) had mean estimates exceeding 2% each (lower CI,  $\geq 1.3\%$ ). In total, 27 K loci had mean prevalence estimates exceeding 1% (see Fig 2a).

**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	1,119	—	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	NIMBI-plus	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	1,930	—	87	14

<https://doi.org/10.1371/journal.pmed.1004879.t001>

K locus prevalence estimates varied substantially across regions (Figs 2b, S2, and S3). Some of the global top-30 loci had significantly higher prevalence in a single region (S3 Fig), e.g., KL149 in Southern Africa (8.0% [95% CI 4.3%, 12.7%], versus  $\leq 0.84\%$  in other regions), and KL15 and KL51 in South Asia (9.1% [5.9%, 13.0%] versus  $\leq 6\%$  elsewhere, and 7.0% [4.2%, 10.6%] versus  $\leq 0.7\%$ , respectively) (Figs 2b and S3). Outside the top-20, KL81 was common in South Asia (3.6%, [1.6%–6.3%]) but very rare elsewhere ( $\leq 0.5\%$ ), 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 S2 Fig). These regional differences could not be explained by differences in the timing of samples from different regions (S1.3 Appendix). 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 15K loci ( $^{\wedge}$  in S3 Fig), but year was not significant for any K loci.

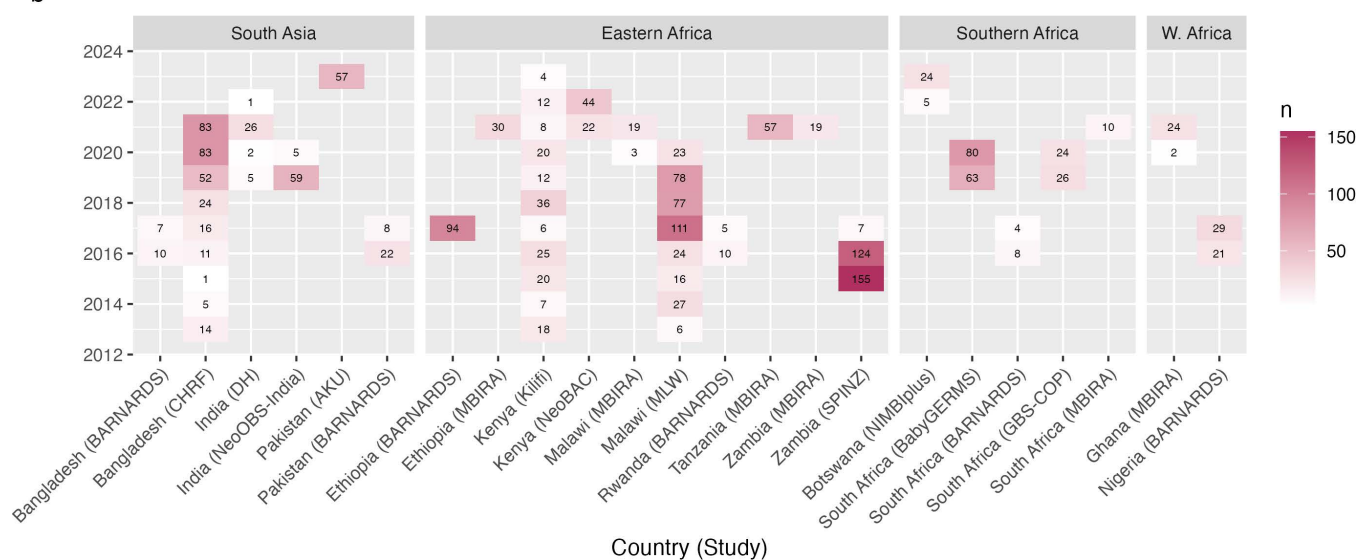
### Theoretical K antigen coverage

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 (Fig 3).

a



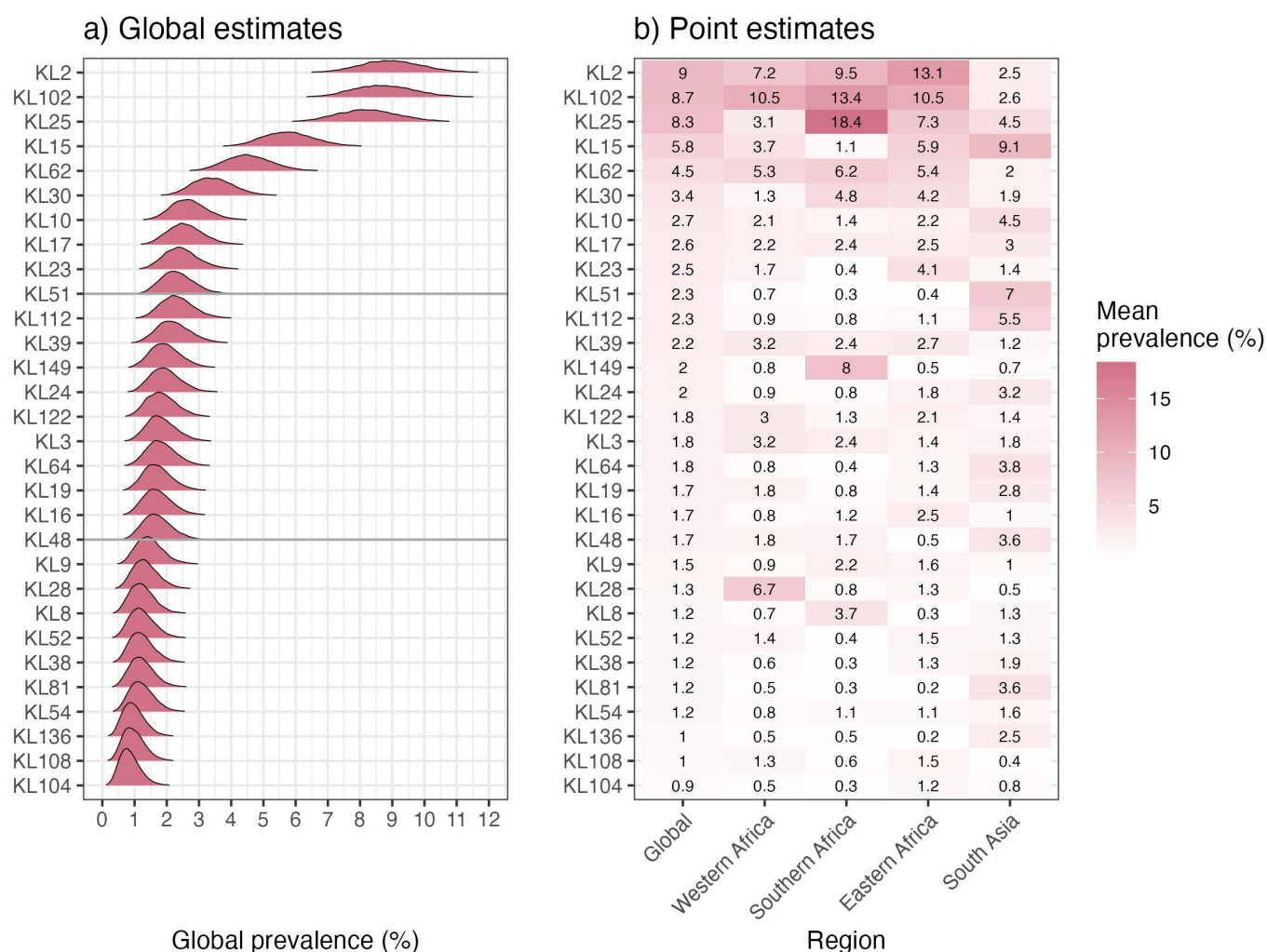
b



**Fig 1. Geotemporal distribution of isolates included in the meta-analysis.** The map shows the 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. Base map source: Natural Earth, <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/>, accessed using rnaturalearth R package v1.10, terms of use: <https://www.naturalearthdata.com/about/terms-of-use/>. (a) Heatmaps indicate the number of samples included per year, from each study site (labelled as “Country, Study”), stratified by region.

<https://doi.org/10.1371/journal.pmed.1004879.g001>

We estimate a vaccine with the global top-20K 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



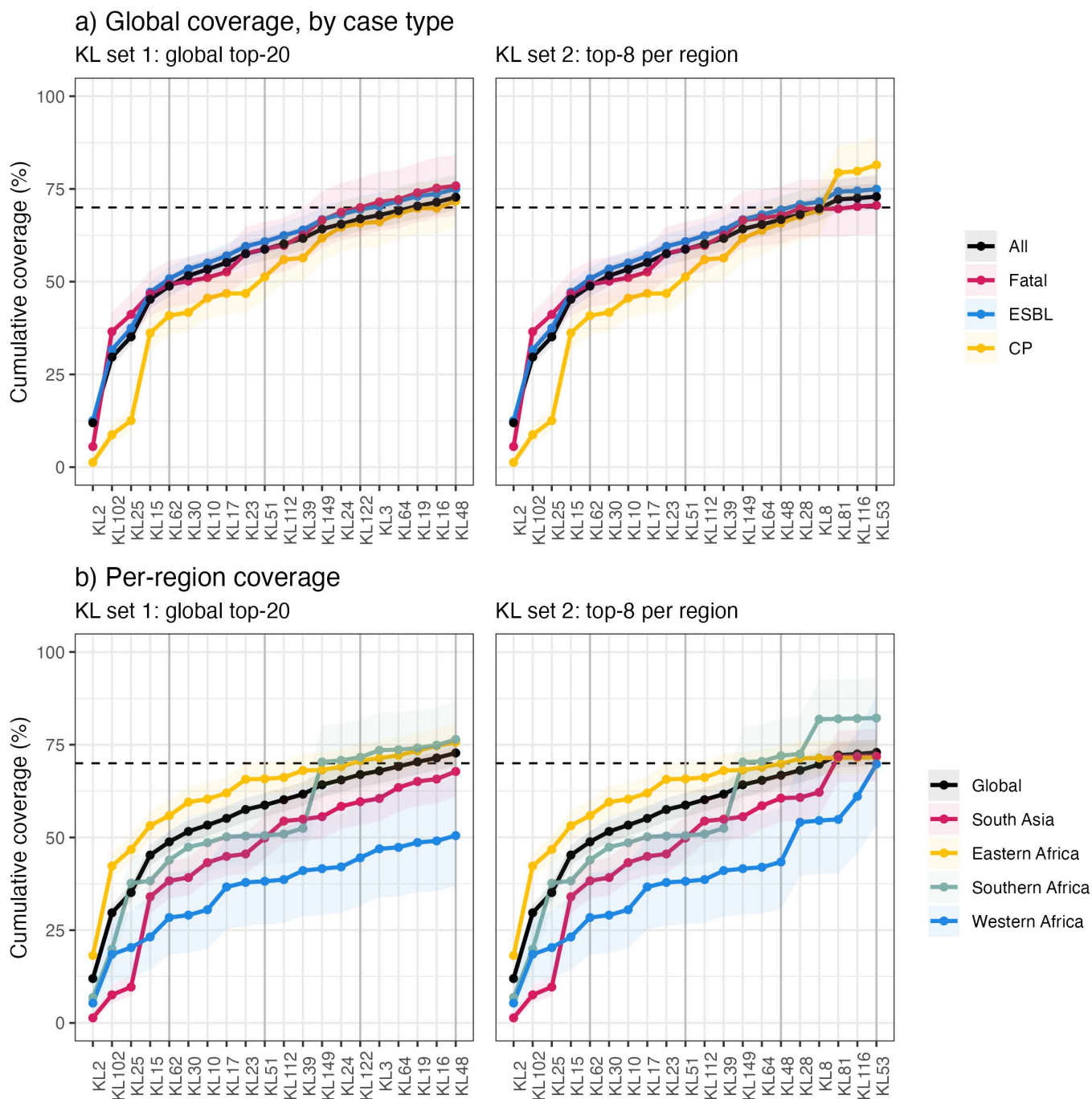
**Fig 2. Modelled estimates of cluster-adjusted global and regional prevalence of K loci.** (a) Posterior density distribution for global prevalence estimates (i.e., across all sites), modelled using cluster-adjusted counts per site, for top-30 K loci ordered by the mean point estimate. Horizontal lines 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 S5 Fig.

<https://doi.org/10.1371/journal.pmed.1004879.g002>

Africa (50.5% [36.9%, 66.4%]) and South Asia (67.8% [61.1%, 75.0%]) (S4 Table). The marginal gain in coverage associated with adding more loci (in order of global rank) reduced substantially beyond the first five 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%] (S4 Table, Fig 3).

The global top-20 K loci provided very similar estimated coverage of neonatal fatalities (outcome data available for  $n=1,063$ , 55.0% of genomes; of which  $n=394$ , 37.1% died during follow-up), and coverage of infections with ESBL-producing strains ( $n=1,747$ , 90.5% of genomes overall, 78.3%–96.7% per region) (see S4 Table, Fig 3). Infections with CP strains were very common in South Asia (87% in Bangladesh, 78% in India, and 87% in Pakistan); common in Nigeria (46%), Ghana (27%), and South Africa (31%); and rare elsewhere (<3%). Estimated coverage of these CP infections





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

<https://doi.org/10.1371/journal.pmed.1004879.g003>

was similar to the overall coverage (top-20K loci covering 71.7% [64.8%–79.0%]). However, four of the global top-5K loci (KL2, KL102, KL25, and KL62) were rare amongst CP isolates, and so the cumulative coverage of the top-5 was lower 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 ([S4 Table](#), [Fig 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  $\geq 80\%$  [lower CI,  $\geq 62\%$ ] coverage in the target region ([S4 Fig](#)). However, any such formulations would provide insufficient coverage of nontarget regions ([S4 Fig](#)). We estimate an alternative set of 20 K loci, comprising all those that ranked in the top-8 for any region, would cover  $\geq 69.8\%$  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 [Fig 3](#), [Table 2](#)).

## O type prevalence and coverage

Global prevalence modelling of predicted O types ([Fig 4a](#)) confirmed the population was dominated by two O1 subtypes, O1 $\alpha\beta$ ,2 $\alpha$  and O1 $\alpha\beta$ ,2 $\beta$ . The median prevalence estimates for these were 26.7% [95% CI 23.6%, 29.9%] for O1 $\alpha\beta$ ,2 $\alpha$  and 21.8% [19.1%, 24.8%] for O1 $\alpha\beta$ ,2 $\beta$ . Next most common ( $\geq 9\%$  each) were O2 $\beta$  (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 O2 $\alpha$  (6.2% [4.6%, 7.9%]); two O3 subtypes (O3 $\gamma$ , 6.0% [4.5%, 7.8%] and O3 $\alpha$ /O3 $\beta$ , 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 [Fig 4a](#)).

There was some regional variation in O types ([Figs 4b](#) and [S5](#), [S5 Table](#)). Whilst O1 variants dominated in all regions, O2 $\beta$  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 O2 $\alpha$  (9.6%, versus 3.5%–5.5% in African regions) and O4 (13.8%, versus 5.7%–7.3% in African regions). Compared with other regions, Southern Africa had higher prevalence of O5 (12.1%, versus 3.1%–5.0% elsewhere), and lower prevalence of O3 $\gamma$  (3.3%, versus 5.4%–7.4% elsewhere) and O13 (1.7%, versus 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 O2 $\beta$  was persistently lower than elsewhere, and O4 and O2 $\alpha$  persistently higher, across the combined sampling period ([S1.4 Appendix](#)). 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 O2 $\alpha$  with nine ST-KL combinations ([S1.5 Appendix](#)). 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 is shown in [Fig 4c](#) and [4d](#), [S5 Table](#). We estimate the top-5 O types (O1 $\alpha\beta$ ,2 $\alpha$ , O1 $\alpha\beta$ ,2 $\beta$ , O2 $\beta$ , O2 $\alpha$ , and O4) collectively cover 86.2% [95% CI 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  $\geq 76\%$  of cases in each region (range 76.0%–92.0% per region, lower CI  $\geq 69\%$ ) ([S5 Table](#)).

## Sensitivity analysis

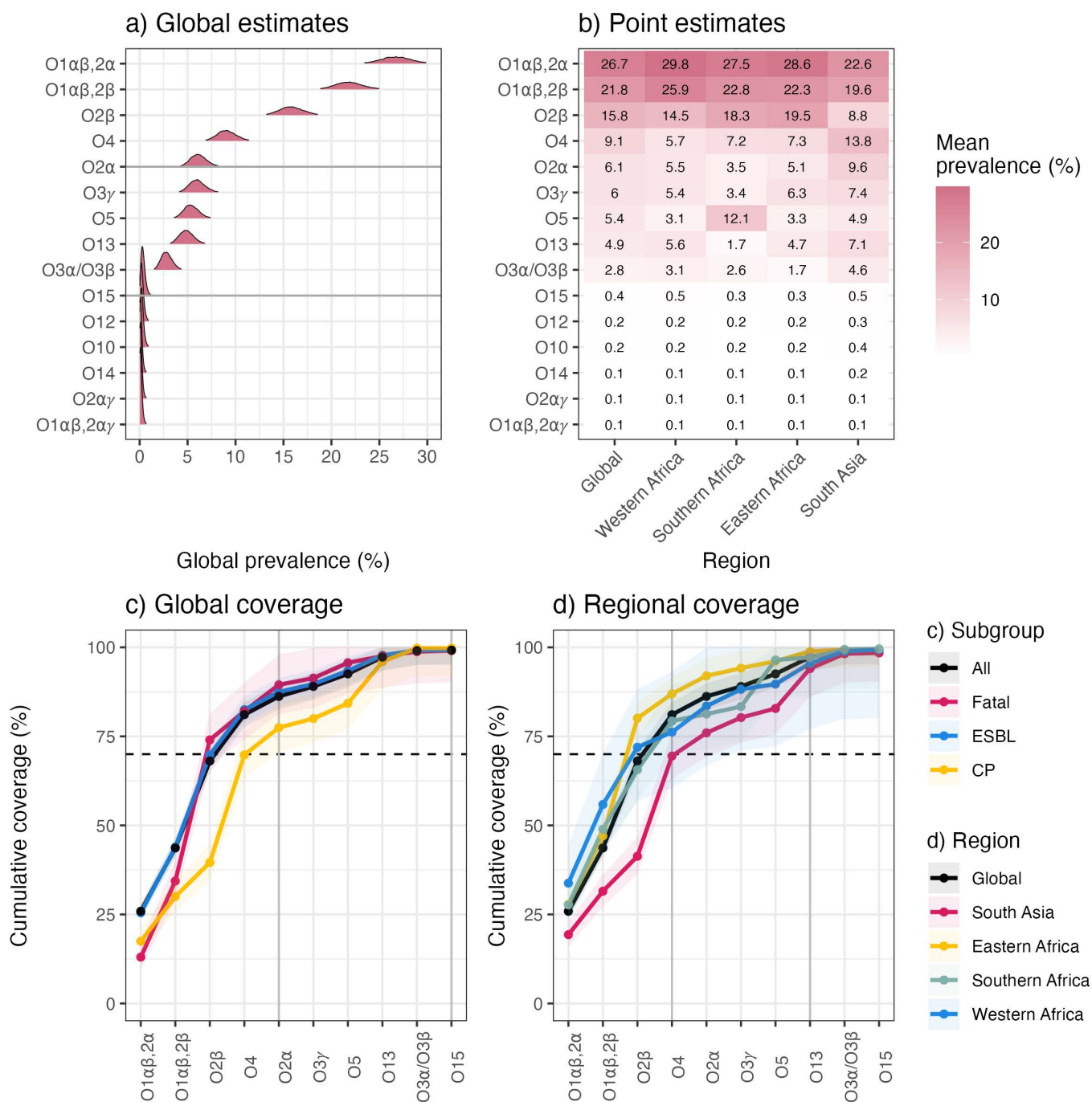
To explore the sensitivity of our analyses to which primary studies were included, we repeated the meta-analysis leaving out each of the  $n = 13$  studies in turn ([S2.1](#), [S2.2 Appendix](#)). Global modelled ranks for K loci were robust to study inclusion, with the same top-5 in all analyses and 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 [S2.1 Appendix](#)). 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 [S2.2 Appendix](#)).

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

	Counts	N	5 KL	10 KL	15 KL	20 KL
<b>Global</b>	raw	1,930	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	1,747	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)
<b>Eastern Africa</b>	raw	1,119	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	1,019	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	–	–	–	–
<b>Southern Africa</b>	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)
<b>Western Africa</b>	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)
<b>South Asia</b>	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 K locus (KL) set 2, comprising the top-8K 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 is 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') is shown as percentages, with 95% credible intervals in brackets.

<https://doi.org/10.1371/journal.pmed.1004879.t002>



**Fig 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; extended-spectrum beta-lactamase, CP; carbapenemase-producing), and (d) per region coverage of neonatal sepsis, selecting the top-10 O types ordered by cluster-adjusted global prevalence.

<https://doi.org/10.1371/journal.pmed.1004879.g004>



The impact of cluster-adjustment on the modelled and simple weighted global prevalence estimates for K loci is shown in [S2.3](#), [S2.4 Appendix](#). 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 ( $\geq 3.0\%$  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 [S2.4 Appendix](#)). 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, and 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 ([S2.5–S2.7 Appendix](#)).

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 [S2.8–S2.10 Appendix](#) 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 [Fig 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 Kenya) 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 ([S6 Fig](#)).

## 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 ([Figs 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 ([Figs 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 ([Fig 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 ([Fig 3](#), [Table 2](#)). This variation was partly driven by differences in geographic distribution of clones carrying distinct K loci ([S1 Appendix](#)). 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-focussed antigen set would provide sub-optimal coverage in other regions ([S4 Fig](#)). 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 ([Fig 2](#), [S4 Table](#)).

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. Coverage estimates using cluster-adjusted counts are also presented in [Tables 2](#), [S4](#), and [S5](#), and are generally a few percentage points lower than the primary estimates, but are likely much lower than the true expected coverage given the high frequency of nosocomial transmission in LMIC neonatal care settings. We estimated similar cumulative coverage for neonates who died during follow-up ([Fig 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* [54]. While capsule overexpression and hypermucoidy in *K. pneumoniae* are known to promote increased virulence [55,56], very few neonatal sepsis isolates in this study carried the associated *ompADC* hypermucoidy locus ( $n = 15$ , 0.8%) [57].

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-20K 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 O2 $\alpha$ , O2 $\beta$ , 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 ([Fig 4](#), [S5 Table](#)), 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 [16,17,58,59]. There is evidence that immunisation with an O1 containing vaccine does not protect mice against K2/O1 *K. pneumoniae* (irrespective of hypermucoidy) [16,60], 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 [32,47,61]. While our global top-5K loci were all reported among the top-10 in adult bloodstream infections in Oxford (UK) [62], only KL2 and KL62 were amongst the top-10 in 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.

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 ( $n = 1,930$ ) is relatively small given the total possible K loci and O types. ii) Our data captured relatively few countries per region ( $\leq 3$  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 CIs for these loci did not substantially overlap those for lower ranked loci ([Fig 2a](#)), they were common across all regions ([Fig 2b](#)), and remained stable in our leave-one-out sensitivity analysis ([S2 Appendix](#)). However, the other K loci showed substantial geographic variation, and had broadly overlapping CIs for estimated prevalence ([Fig 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 and 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 is 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-enrolment), which means we were unable to directly compare antigen profiles among fatal and nonfatal 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 ([S1 Appendix](#)), and those sites with longitudinal data showed reasonable stability within individual sites over time ([S6 Fig](#)).

Secondly, our analyses have relied upon genomic antigen predictions rather than direct serological data, which is very challenging to obtain. K serotyping 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 [63]. More work is needed to fully assess these cross-reactions and understand their relevance for protection against invasive *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 [59], and single or bivalent anti-K vaccines can be effective in preventing invasive *K. pneumoniae* infection in murine models [64,65]. While protein-based vaccines are also in development [66], 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].

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 [67,68]. 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-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.

## Supporting information

**S1 Table. Ethics committees approving individual studies.** 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), and covers inclusion of data from the studies whose primary ethical approvals are listed in this table.

(PDF)

**S2 Table. Studies included in the meta-analysis.** Details of type of study, number of isolates; laboratory methods for identification, storage, DNA extraction, sequencing; informatics methods for quality control and assembly; BioProject accession and citation/s for primary study.

(TSV)

**S3 Table. Isolates and sequence data included in the study.**

(TSV)

**S4 Table. Coverage estimates for KL set 1 (global top-20).** Cumulative coverage estimates for the top 5, 10, 15, or 20K loci (KL) ordered by cluster-adjusted global prevalence estimate. For each geographic location, the breakdown



of isolate counts and corresponding coverage estimates is 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') is shown as percentages, with 95% credible intervals in brackets.  
(PDF)

**S5 Table. Coverage estimates for global top-10 O genotypes.** Cumulative coverage estimates for the 10 O genotypes, ordered by cluster-adjusted global prevalence estimate. For each geographic location, the breakdown of isolate counts and corresponding coverage estimates is 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') is shown as percentages, with 95% credible intervals in brackets.  
(TSV)

**S1 Fig. Flow diagram for inclusion of isolates.** WGS, whole-genome sequencing; QC, quality control.  
(TIFF)

**S2 Fig. Modelled estimates of cluster-adjusted global and regional prevalence of all K loci.** (a) Posterior density distribution for global (overall) prevalence estimates, modelled using cluster-adjusted counts per site, ordered by the mean point estimate. Horizontal lines indicate groups of 10 loci. (b) Regional prevalence estimates. Cells are coloured and labelled to indicate the cluster-adjusted regional mean prevalence estimates for each K locus in each region, according to the inset legend.  
(TIFF)

**S3 Fig. Posterior distributions of K locus prevalence per region.** \*Indicates K loci for which the median estimate per region differs by >2%, for at least one pair of regions. ^Indicates K loci for which region was a significant linear predictor in a logistic regression model. Horizontal lines indicate the top 10 and 20 global ranked loci.  
(TIFF)

**S4 Fig. Regional coverage estimates for top-20 K loci selected on a region-focussed basis.** Each plot shows the modelled coverage in each region, using K loci ranked by cluster-adjusted counts in one region (indicated in the title of the plot).  
(TIFF)

**S5 Fig. Posterior distributions of O type prevalence per region.** \*Indicates O types for which the median estimate per region differs by >5%. ^Indicates O types for which region was a significant linear predictor in a logistic regression model.  
(TIFF)

**S6 Fig. Longitudinal data.** Three sites collected routine blood culture isolates for extended periods, providing an opportunity to assess whether a set of loci selected based on prevalence during one surveillance period (here,  $n=512$  samples obtained up to 2017 inclusive, labelled "Training") could provide coverage of infections occurring in a later time period (here,  $n=512$  data after 2017, labelled "Post-training"). Note the coverage estimates shown reflect the crude proportion per surveillance site (or pooled across sites, labelled "All"), as opposed to Bayesian modelled estimates.  
(TIFF)

**S1 Appendix. Genomic diversity.**  
(PDF)

**S2 Appendix. Sensitivity analyses.**  
(PDF)

### S3 Appendix. Model Diagnostics.

(PDF)

**S1 STROBE Statement.** STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) Statement – checklist of items that should be included in reports of observational studies, available at <https://www.strobe-statement.org/>, licenced under CC BY 4.0.

(PDF)

### S1 Protocol. Study protocol.

(PDF)

## 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), Neuberger 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.

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.

## Author contributions

**Conceptualization:** Kelly L. Wyres, Kathryn E. Holt.

**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 Stryko, Caroline C. Tigoi.

**Formal analysis:** Thomas D. Stanton, Shaun P. Keegan, Kelly L. Wyres, Kathryn E. Holt.

**Funding acquisition:** Ramesh K. Agarwal, Alexander M. Aiken, James A. Berkley, Susan E. Coffin, Nicholas A. Feasey, Nelesh P. Govender, Davidson H. Hamer, Shabir A. Madhi, Muhammad Imran Nisar, Samir K. Saha, Senjuti Saha, Mari Jeeva Sankar, Kelly L. Wyres, Kathryn E. Holt.

**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 C.L. Mwansa, Moreka L. Ndumba, Erkison E. Odih, Donwilliams O. Omuoyo, Oliver Pearse, Laura T. Phillips, Paul J. Planet, Anika Abdul Rasool, Charlene M.C. Rodrigues, Kirsty Sands, Arif M. Tanmoy, Erin Theiller, Allan M. Zuza.

**Methodology:** Thomas D. Stanton, Shaun P. Keegan, Kelly L. Wyres, Kathryn E. Holt.

**Resources:** Kajal Jain, Sulagna Basu, Grace J. Chan, Kenneth C. Iregbu, Jean-Baptiste Mazarati, Semaria Solomon Alemayehu, Timothy R. Walsh, Rabaab Zahra, 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, Ramesh K. Agarwal, Alexander M. Aiken, James A. Berkley, Susan E. Coffin, Nicholas A. Feasey, Nelesh P. Govender, Davidson H. Hamer, Shabir A. Madhi, Muhammad Imran Nisar, Samir K. Saha, Senjuti Saha, Mari Jeeva Sankar.

**Software:** Thomas D. Stanton, Shaun P. Keegan, Kelly L. Wyres, Kathryn E. Holt.

**Supervision:** Kajal Jain, Sulagna Basu, Grace J. Chan, Kenneth C. Iregbu, Jean-Baptiste Mazarati, Semaria Solomon Alemayehu, Timothy R. Walsh, Rabaab Zahra, 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, Ramesh K. Agarwal, Alexander M. Aiken, James A. Berkley, Susan E. Coffin, Nicholas A. Feasey, Nelesh P. Govender, Davidson H. Hamer, Shabir A. Madhi, Muhammad Imran Nisar, Samir K. Saha, Senjuti Saha, Mari Jeeva Sankar.

**Visualization:** Thomas D. Stanton, Shaun P. Keegan, Kelly L. Wyres, Kathryn E. Holt.

**Writing – original draft:** Kelly L. Wyres, Kathryn E. Holt.

**Writing – review & editing:** Thomas D. Stanton, Shaun P. Keegan, Jabir A. Abdulahi, Anne V. Amulele, Matthew Bates, Yogesh Hooda, Weiming Hu, Kajal Jain, Samiah Kanwar, Rindidzani Magobo, Courtney P. Olwagen, John M. Tembo, Tolbert Sonda, Jonathan Stryzko, Caroline C. Tigoi, 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 C.L. Mwansa, Moreka L. Ndumba, Erkison E. Odih, Donwilliams O. Omuoyo, Oliver Pearse, Laura T. Phillips, Paul J. Planet, Anika Abdul Rasool, Charlene M.C. Rodrigues, Kirsty Sands, Arif M. Tanmoy, Erin Theiller, Allan M. Zuza, Sulagna Basu, Grace J. Chan, Kenneth C. Iregbu, Jean-Baptiste Mazarati, Semaria Solomon Alemayehu, Timothy R. Walsh, Rabaab Zahra, Sombo Fwoloshi, Appiah-Korang Labi, Lola Madrid, Noah Obeng-Nkrumah, David Ojok, Boaz D. Wadugu, Adhisivam Bethou, Anudita Bhargava, Atul Jindal, Ruchi N. Nanavati, Priyanka S. Prasad, Apurba Sastry, Joveria Q. Farooqi, Najia Ghanchi, Fyezah Jehan, Ramesh K. Agarwal, Alexander M. Aiken, James A. Berkley, Susan E. Coffin, Nicholas A. Feasey, Nelesh P. Govender, Davidson H. Hamer, Shabir A. Madhi, Muhammad Imran Nisar, Samir K. Saha, Senjuti Saha, Mari Jeeva Sankar, Kelly L. Wyres, Kathryn E. Holt.

## 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. *Lancet*. 2024;404(10459):1199–226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1) PMID: [39299261](https://pubmed.ncbi.nlm.nih.gov/39299261/)
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(4):512–23. <https://doi.org/10.1038/s41564-021-00870-7> PMID: [33782558](https://pubmed.ncbi.nlm.nih.gov/33782558/)
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(6):e1004179. <https://doi.org/10.1371/journal.pmed.1004179> PMID: [37289666](https://pubmed.ncbi.nlm.nih.gov/37289666/)
4. Okomo U, Akpalu ENK, Le Doare K, 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(11):1219–34. [https://doi.org/10.1016/S1473-3099\(19\)30414-1](https://doi.org/10.1016/S1473-3099(19)30414-1) PMID: [31522858](https://pubmed.ncbi.nlm.nih.gov/31522858/)
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(12):1677–88. [https://doi.org/10.1016/S1473-3099\(21\)00050-5](https://doi.org/10.1016/S1473-3099(21)00050-5) PMID: [34384533](https://pubmed.ncbi.nlm.nih.gov/34384533/)
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(2):e131–41. [https://doi.org/10.1016/S2666-5247\(23\)00290-2](https://doi.org/10.1016/S2666-5247(23)00290-2) PMID: [38218193](https://pubmed.ncbi.nlm.nih.gov/38218193/)
7. Dangor Z, Benson N, Berkley JA, Bielicki J, Bijlma MW, Broad J, et al. Vaccine value profile for *Klebsiella pneumoniae*. *Vaccine*. 2024;42(19 Suppl 1):S125–41. <https://doi.org/10.1016/j.vaccine.2024.02.072> PMID: [38503661](https://pubmed.ncbi.nlm.nih.gov/38503661/)

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(Suppl 1):S38–45. <https://doi.org/10.1093/cid/ciad282> PMID: 37406039
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. <https://doi.org/10.1016/j.ijheh.2024.114357> PMID: 38564877
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(5):e1004239. <https://doi.org/10.1371/journal.pmed.1004239> PMID: 37216371
11. World Health Organization. Estimating the impact of vaccines in reducing antimicrobial resistance and antibiotic use: technical report. World Health Organization. 2024. Available from: <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(23):126063. <https://doi.org/10.1016/j.vaccine.2024.06.030> PMID: 38897895
13. Lucinde RK, Ong'ayo G, Houlihan C, Bottomley C, Goldblatt D, Scott JAG, et al. Pneumococcal conjugate vaccine dose-ranging studies in humans: a systematic review. *Vaccine*. 2021;39(36):5095–105. <https://doi.org/10.1016/j.vaccine.2021.07.033> PMID: 34340858
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(5):CD015746. <https://doi.org/10.1002/14651858.CD015746.pub2> PMID: 40326553
15. Male V, Jones CE. Vaccination in pregnancy to protect the newborn. *Nat Rev Immunol*. 2025;25(9):649–61. <https://doi.org/10.1038/s41577-025-01162-5> PMID: 40269273
16. Wantuch PL, Knoop 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(3):578–89. <https://doi.org/10.1093/infdis/jiae097> PMID: 38401891
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(4):100988. <https://doi.org/10.1016/j.lanmic.2024.100988> PMID: 39952262
18. Bennett JC, Deloria Knoll M, Kagucia EW, Garcia Quesada M, 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*. 2025;25(4):457–70. [https://doi.org/10.1016/S1473-3099\(24\)00665-0](https://doi.org/10.1016/S1473-3099(24)00665-0) PMID: 39706204
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(1):e0017524. <https://doi.org/10.1128/cmr.00175-24> PMID: 39878373
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 (Nathan)*. 2024;16(1):29. <https://doi.org/10.1186/s41479-024-00152-w> PMID: 39497193
21. Stanton TD, Hetland MAK, L  hr IH, Holt KE, Wyres KL. Fast and accurate in silico antigen typing with Kaptive 3. *Microb Genom*. 2025;11(6):001428. <https://doi.org/10.1099/mgen.0.001428> PMID: 40553506
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(2):e0009023. <https://doi.org/10.1128/mmb.00090-23> PMID: 40116577
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(6):e00197-18. <https://doi.org/10.1128/JCM.00197-18> PMID: 29618504
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 Genom*. 2022;8(3):000800. <https://doi.org/10.1099/mgen.0.000800> PMID: 35311639
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. <https://doi.org/10.1038/srep15573> PMID: 26493302
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;155:315–23. <https://doi.org/10.1016/j.ijbiomac.2020.03.196> PMID: 32224183
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;130:536–44. <https://doi.org/10.1016/j.ijbiomac.2019.02.128> PMID: 30802520
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. <https://doi.org/10.1016/j.car-res.2012.12.018> PMID: 23360863
29. Stanton T, Jacobsen M, Gymoese P, Basu S, Chan G, Dangor Z. Technical report: accuracy of *Klebsiella pneumoniae* capsule serotype predictions from whole genome sequences. Zenodo. 2025. Available from: <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. <https://doi.org/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 Genom.* 2024;10(7):001271. <https://doi.org/10.1099/mgen.0.001271> PMID: 39037209
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(1):11. <https://doi.org/10.1186/s13073-019-0706-y> PMID: 31948471
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(3):227–39. <https://doi.org/10.15252/emmm.201404767> PMID: 25712531
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(8):e1170–8. [https://doi.org/10.1016/S2214-109X\(22\)00246-7](https://doi.org/10.1016/S2214-109X(22)00246-7) PMID: 35839815
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(10):921–6. <https://doi.org/10.1097/INF.0000000000004008> PMID: 37364138
36. Phillips LT, Bates M, Coffin SE, Foster-Nyarko E, Kapasa M, Machona S, et al. Transmission dynamics of *Klebsiella pneumoniae* in a neonatal intensive care unit in Zambia before and after an infection control bundle. *medRxiv.* 2025;2025(11):1. <https://doi.org/10.1101/2025.11.12.25340082>
37. Aiken AM, Rehman AM, de Kraker MEA, 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(11):1280–90. [https://doi.org/10.1016/S1473-3099\(23\)00233-5](https://doi.org/10.1016/S1473-3099(23)00233-5) PMID: 37454672
38. 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(5):e870–8. [https://doi.org/10.1016/S2214-109X\(24\)00564-3](https://doi.org/10.1016/S2214-109X(24)00564-3) PMID: 40023188
39. 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(1):8043. <https://doi.org/10.1038/s41598-025-92517-4> PMID: 40055469
40. Stryzko J, Hu W, Mochankana K, John-Thubuka J, Zankere T, Gopolang B, et al. Using Genomic and Traditional Epidemiologic Approaches to Define Complex Transmission Pathways of *Klebsiella pneumoniae* Infection in a Neonatal Unit in Botswana, 2022–2023. *medRxiv.* 2025;2025(11):1. <https://doi.org/10.1101/2025.11.06.25339637>
41. Amulele AV, Orindi B, Ndumba M, Tigoi C, Omuoyo D, Kahindi E, et al. Genomic epidemiology of *Klebsiella pneumoniae* neonatal and infant sepsis in Kenyan hospitals. *medRxiv.* 2025;2025(10):1. <https://doi.org/10.1101/2025.10.30.25339129>
42. 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;2025-06. <https://doi.org/10.1101/2025.06.11.25329357>
43. 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(1):4188. <https://doi.org/10.1038/s41467-021-24448-3> PMID: 34234121
44. United Nations Statistics Division. Standard country or area codes for statistical use (M49) [cited 15 Mar 2025]. Available from: <https://unstats.un.org/unsd/methodology/m49/>
45. 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(Suppl\_4):S325–35. <https://doi.org/10.1093/cid/ciab784> PMID: 34850838
46. 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(9):e00573-19. <https://doi.org/10.1128/JCM.00573-19> PMID: 31315956
47. 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(11):1919–29. <https://doi.org/10.1038/s41564-019-0492-8> PMID: 31358985
48. Gorrie CL, Mirceta 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(2):208–15. <https://doi.org/10.1093/cid/cix270> PMID: 28369261
49. 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 Genom.* 2023;9(2):mgen000936. <https://doi.org/10.1099/mgen.0.000936> PMID: 36752781
50. Antonov M, Csárdi G, Horvát S, Müller K, Nepusz T, Noom D. igraph enables fast and robust network analysis across programming languages. *arXiv.* 2023. <https://doi.org/10.48550/arXiv.2311.10260>
51. Bürkner PC. Advanced Bayesian multilevel modeling with the R package brms. *R J.* 2018;10:395–411.
52. 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. <https://doi.org/10.18637/jss.v076.i01> PMID: 36568334
53. Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.

54. 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(7):1203–11. <https://doi.org/10.1086/423820> PMID: [15346329](#)
55. 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;107:105302. <https://doi.org/10.1016/j.ebiom.2024.105302> PMID: [39178743](#)
56. 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(6):606–14. <https://doi.org/10.1111/j.1365-2796.2006.01641.x> PMID: [16704562](#)
57. 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(1):36. <https://doi.org/10.1186/s13073-025-01461-5> PMID: [40205597](#)
58. 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(5):2402–9. <https://doi.org/10.1128/IAI.68.5.2402-2409.2000> PMID: [10768923](#)
59. 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 (Basel)*. 2024;12(10):1177. <https://doi.org/10.3390/vaccines12101177> PMID: [39460343](#)
60. 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(5):e1011367. <https://doi.org/10.1371/journal.ppat.1011367> PMID: [37146068](#)
61. 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(3):401–12. [https://doi.org/10.1016/S1473-3099\(21\)00399-6](https://doi.org/10.1016/S1473-3099(21)00399-6) PMID: [34767753](#)
62. 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(12):2276–82. <https://doi.org/10.1093/cid/ciab006> PMID: [33411882](#)
63. 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):898. <https://doi.org/10.1038/s42003-025-08258-7> PMID: [40490475](#)
64. Wantuch PL, Knoot CJ, Marino EC, Harding CM, Rosen DA. *Klebsiella pneumoniae* bioconjugate vaccine functional durability in mice. *Vaccine*. 2025;43(Pt 2):126536. <https://doi.org/10.1016/j.vaccine.2024.126536> PMID: [39571358](#)
65. 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 U S A*. 2019;116(37):18655–63. <https://doi.org/10.1073/pnas.1907833116> PMID: [31455739](#)
66. Choi M, Tennant SM, Simon R, Cross AS. Progress towards the development of *Klebsiella* vaccines. *Expert Rev Vaccines*. 2019;18(7):681–91. <https://doi.org/10.1080/14760584.2019.1635460> PMID: [31250679](#)
67. Doua J, Ndembu N, Auerbach J, Kaseya J, Zumla A. Advancing local manufacturing capacities for vaccines within Africa—Opportunities, priorities and challenges. *Vaccine*. 2025;50:126829. <https://doi.org/10.1016/j.vaccine.2025.126829> PMID: [39914258](#)
68. 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(8):e347–8. [https://doi.org/10.1016/S2666-5247\(21\)00126-9](https://doi.org/10.1016/S2666-5247(21)00126-9) PMID: [34179857](#)