Communication

Not peer-reviewed version

Underrepresentation of African Surveillance Data in Global Arbovirus Modelling

Rosheen S. Mthawanji *

Posted Date: 16 June 2026

doi: 10.20944/preprints202606.1194.v1

Keywords: arboviruses; surveillance; anopheles; vector ecology; predictive modelling

Communication

Underrepresentation of African Surveillance Data in Global Arbovirus Modelling

Rosheen S. Mthawanji

Department of Evolution, Ecology and Behavior, University of Liverpool, Liverpool, United Kingdom; rsm1@liverpool.ac.uk

Abstract

Recent expansion of arboviruses across Europe has intensified efforts to understand the ecological and environmental drivers of pathogen emergence. However, global interpretations of arbovirus dynamics remain disproportionately informed by data from Europe and North America, where surveillance systems are well established. In contrast, large regions of Africa despite hosting high mosquito diversity and extensive human–vector contact remain substantially under-sampled. In this perspective, we argue that this imbalance constitutes a critical blind spot in global arbovirus surveillance, with implications for both risk assessment and predictive modelling. We suggest that apparent emergence in data-rich regions may partly reflect improved detection rather than true geographic expansion, occurring against a background of poorly characterised transmission in under-sampled regions Furthermore, the exclusion of underrepresented vector systems, including Anopheles mosquitoes, may constrain understanding of broader transmission networks We highlight how fragmented surveillance can influence ecological inference, disease forecasting, and global risk assessment, particularly in increasingly interconnected transmission systems. Finally, we argue that integrating arbovirus surveillance into existing malaria monitoring infrastructure across Africa could provide a scalable and cost-effective strategy for improving global surveillance coverage. Addressing structural surveillance biases will be essential for developing more representative and predictive frameworks of arbovirus emergence.

Keywords: arboviruses; surveillance; anopheles; vector ecology; predictive modelling

Introduction

Recent decades have witnessed an increasing frequency of arbovirus emergence and re-emergence across multiple regions of the world, with particularly notable expansion into Europe [1–5].Viruses such as Usutu virus and West Nile virus, once considered geographically restricted, have now established transmission cycles in temperate regions, raising concerns about the drivers of emergence and the adequacy of current surveillance systems [6,7]. These patterns are commonly attributed to climate change, globalisation, and shifts in vector distribution, all of which can facilitate the spread of pathogens into new ecological niches [8]. While these explanations are well supported, they largely focus on processes occurring within regions where detection has already taken place, potentially overlooking broader structural biases in how arbovirus data are generated and interpreted.

A fundamental yet underexamined issue in arbovirus ecology is the uneven global distribution of surveillance effort. Contemporary understanding of arbovirus transmission is heavily shaped by data originating from Europe and North America, where sustained investment in vector monitoring, pathogen detection, and public health infrastructure has enabled detailed characterisation of transmission dynamics [9]. In contrast, large areas of Africa remain comparatively under-sampled, despite harbouring high mosquito biodiversity, diverse vertebrate host communities, and environmental conditions conducive to vector-borne disease transmission [10]. In many settings, surveillance is spatially fragmented, temporally inconsistent, or absent altogether, resulting in substantial gaps in empirical data.

Current interpretations of arbovirus emergence are strongly shaped by unequal surveillance intensity across regions. Observed increases in arbovirus detection in Europe may reflect genuine ecological change; however, they may also be influenced by improvements in surveillance sensitivity and coverage, which enhance the probability of detection relative to historically under-sampled regions [11,12]. In this context, an increase in arbovirus detection is not solely a biological phenomenon but also a function of observation, such that regions with limited surveillance capacity may harbour ongoing transmission that remains undetected, leading to a skewed perception of global disease distribution.

Africa represents one of the most significant underrepresented regions within current global arbovirus surveillance frameworks The continent supports a wide range of mosquito species, including both established arbovirus vectors and numerous understudied taxa, alongside dense human populations and extensive wildlife reservoirs [2]. These characteristics create conditions that are highly favourable for arbovirus maintenance and transmission. Yet, the absence of systematic, large-scale surveillance limits the ability to detect viral circulation and to assess its contribution to regional and global transmission dynamics. As a result, Africa is often underrepresented in global datasets used to parameterise models of arbovirus risk, potentially leading to biased or incomplete predictions [13].

The consequences of this bias extend beyond regional knowledge gaps. Predictive models of disease emergence rely on representative data to infer relationships between vectors, hosts, pathogens, and environmental drivers [15]. When substantial portions of the global system are under-sampled, these models may fail to capture critical transmission pathways, including those involving less-studied vector taxa or complex ecological interactions. This limitation is particularly relevant in the context of increasing global connectivity, where movement of hosts, vectors, and pathogens can link distant geographic regions into a single epidemiological network.

Despite growing recognition of arbovirus emergence, quantitative evidence highlights substantial geographic bias in surveillance effort. Global datasets of vector–pathogen interactions and disease occurrence are disproportionately populated by records from Europe and North America, reflecting long-standing investment in public health infrastructure and research capacity [12,14]. In contrast, much of Africa remains underrepresented in these datasets, with surveillance often limited to outbreak-driven investigations or geographically restricted studies [15]. This imbalance is particularly striking when contrasted with malaria research, where extensive entomological surveillance networks exist across the continent [10]. The divergence between well-developed malaria surveillance systems and comparatively limited arbovirus monitoring highlights a missed opportunity to leverage existing infrastructure for broader vector-borne disease detection.

In this perspective, we argue that Africa constitutes a major blind spot in global arbovirus surveillance and that this bias is not only a limitation of data availability but a structural issue that systematically distorts ecological inference and predictive modelling of arbovirus emergence. By recognising and addressing the structural biases seen in current surveillance approaches, it may be possible to develop a more accurate and comprehensive understanding of arbovirus emergence at a global scale.

Surveillance Bias and the Construction of Global Arbovirus Knowledge

Current models of arbovirus emergence and transmission are heavily influenced by data availability, yet this dependence is rarely acknowledged explicitly in model interpretation. Regions with established surveillance infrastructure contribute disproportionately to global datasets, shaping both empirical understanding and predictive modelling efforts [12]. In practice, this means that a relatively small number of well-studied vector–virus systems primarily involving *Aedes aegypti* and *Culex pipiens* form the backbone of global arbovirus knowledge [9]. These systems have been intensively investigated through longitudinal surveillance, laboratory competence studies, and genomic sequencing, particularly in Europe and North America.

By contrast, in many parts of Africa, arbovirus data are derived from sporadic outbreak investigations or limited cross-sectional surveys rather than sustained surveillance programmes. For example, viral detection often occurs in response to clinical cases or sentinel events, rather than through routine vector screening or longitudinal monitoring. This creates a fragmented evidence base that is difficult to integrate into global analyses. Limited surveillance capacity in many African settings complicates interpretation of apparent pathogen absence and may contribute to under-detection of ongoing transmission.

This imbalance has several consequences. First, it creates a perception that arbovirus emergence is geographically concentrated in data-rich regions, reinforcing existing research priorities and funding patterns. For example, global arbovirus datasets and vector competence studies are heavily skewed toward Europe and North America, with comparatively limited representation from African countries despite their ecological suitability for transmission [14]. This imbalance is further reflected in the distribution of longitudinal surveillance programmes, which are concentrated in high-income regions.

Second, it limits the identification of transmission pathways that fall outside well-characterised systems, including those involving understudied vector species or alternative host networks. Third, it constrains the development of predictive models, which are often trained on incomplete datasets and may therefore fail to capture the full ecological complexity of arbovirus transmission [13].

Importantly, these biases are self-reinforcing. Regions that generate more data are more likely to be included in future studies, while under-sampled regions remain marginalised, further entrenching gaps in global understanding. Addressing this issue requires not only increased data collection but also explicit recognition of how surveillance bias shapes the construction of scientific knowledge.

Implications for Emergence and Risk Prediction

The underrepresentation of African data has direct implications for how arbovirus emergence is interpreted and predicted. Emergence is typically framed as a process driven by ecological change, including climate variability, land use transformation, and shifts in vector distribution [8]. However, these interpretations depend fundamentally on the availability of baseline data against which change can be measured. In regions where historical data are sparse or absent, it becomes difficult to distinguish between true emergence and increased detection.

The recent expansion of arboviruses in Europe provides a clear example of this challenge. Enhanced surveillance systems, including coordinated vector monitoring and molecular diagnostics, have increased the likelihood of detecting viral circulation. While environmental change has undoubtedly contributed to transmission dynamics, the role of improved detection should not be underestimated [6]. Without comparable surveillance intensity in Africa, it is not possible to determine whether similar transmission patterns are occurring undetected.

This limitation has important consequences for predictive modelling. Many models of arbovirus risk rely on correlations between environmental variables and observed transmission events. When these models are trained predominantly on data from Europe and North America, they may implicitly assume that these regions are representative of global dynamics. However, this assumption is unlikely to hold in ecologically diverse regions such as Africa, where vector communities, host availability, and environmental conditions differ substantially.

Furthermore, incomplete data may obscure key components of transmission systems, including potential reservoir hosts and alternative vector pathways. For example, migratory birds are known to play a role in the spread of viruses such as West Nile virus, linking African and European transmission systems. However, limited data on viral circulation within African bird populations restricts our ability to quantify this connection. Similarly, human mobility and trade networks may facilitate pathogen movement, yet these pathways are difficult to model without comprehensive surveillance data. As illustrated in Figure 1, this uneven distribution of surveillance effort leads to biased representations of global arbovirus risk.

Finally, an additional, largely unexplored dimension of this overlap is the potential for co-exposure and interaction between arboviruses and *Plasmodium* within shared vector and host systems. In regions where *Anopheles* mosquitoes coexist with established arbovirus vectors, the possibility of concurrent transmission or co-infection remains poorly understood. While *Anopheles* are primarily studied in the context of malaria, their ecological overlap with arbovirus transmission systems raises important questions regarding how viral and parasitic infections may interact within vectors or vertebrate hosts. These interactions could influence vector competence, pathogen dynamics, or transmission efficiency, yet they remain largely uncharacterised, particularly in African settings where surveillance is limited. Addressing this gap represents an important direction for future research in integrated vector–pathogen systems.

Taken together, these gaps suggest that current models may underestimate both the complexity and the scale of arbovirus transmission. Addressing this issue requires the incorporation of under-sampled regions into both empirical studies and modelling frameworks, as well as the development of approaches that explicitly account for data uncertainty.

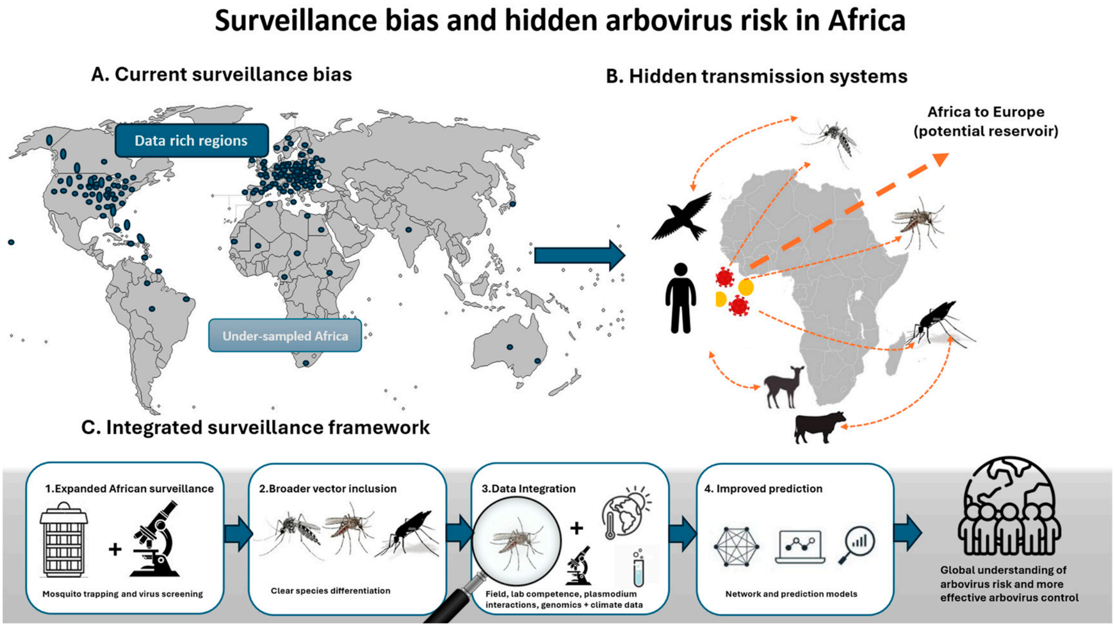

Figure 1. Surveillance inequality and global arbovirus risk prediction (A) Arbovirus surveillance is concentrated in data-rich regions such as Europe and North America, while Africa remains under-sampled. (B) This imbalance may obscure under-characterised transmission systems involving multiple vectors and hosts. (C) Expanding surveillance and integrating diverse data sources can improve global risk prediction.

Integrating Overlooked Vector Systems

A further limitation of current surveillance frameworks is the narrow focus on a small subset of vector species. While *Aedes aegypti* and *Culex pipiens* have been extensively studied due to their established roles in arbovirus transmission, other mosquito taxa remain comparatively understudied. This focus reflects historical research priorities rather than the full diversity of ecological interactions that may influence transmission.

In Africa, this limitation is particularly evident in the case of *Anopheles* mosquitoes. These species are central to malaria transmission and are therefore among the most intensively studied vectors in the world. Extensive infrastructure exists for their monitoring, including entomological surveillance, insecticide resistance tracking, and intervention programmes [10]. However, despite this detailed knowledge base, *Anopheles* mosquitoes are rarely included in arbovirus surveillance or predictive models.

This exclusion is difficult to justify from an ecological perspective. *Anopheles* species frequently coexist with known arbovirus vectors, share vertebrate hosts, and occupy overlapping habitats. In many settings, they are abundant and exhibit opportunistic feeding behaviour, increasing the likelihood of exposure to arboviruses. Experimental studies have demonstrated that some Anopheles species can support arbovirus infection under laboratory conditions, although their efficiency as vectors may vary [16,17]. The systematic exclusion of *Anopheles* from arbovirus frameworks represents not only a taxonomic gap but a conceptual limitation in how vector–pathogen systems are defined.

The separation of malaria and arbovirus research systems further reinforces this gap. Surveillance programmes are typically organised around specific diseases, with limited integration across pathogen types. As a result, opportunities to detect co-exposure or shared transmission pathways are often missed. This is particularly relevant in high-transmission settings, where multiple pathogens may circulate within the same ecological system.

Recognising and addressing these overlaps does not imply that Anopheles mosquitoes are major arbovirus vectors in all contexts. Rather, it highlights the need to move beyond rigid classifications of “primary” and “secondary” vectors and to consider a broader range of ecological interactions. Even limited contributions from neglected vectors could influence transmission dynamics, particularly in regions with high vector density and environmental variability.

Toward an Integrated Surveillance Framework

Addressing the blind spot in African arbovirus surveillance requires a shift from fragmented, pathogen-specific approaches toward integrated and scalable surveillance frameworks. We propose a conceptual framework based on three interconnected components.

First, surveillance must expand from reactive, outbreak-driven sampling to systematic and longitudinal monitoring across diverse ecological settings. This includes routine mosquito sampling across environmental gradients, coupled with pathogen screening using both targeted diagnostics and untargeted genomic approaches. Such strategies would enable detection of both known and previously uncharacterised viruses.

Second, surveillance systems must move beyond a narrow focus on established vector species to incorporate broader vector communities. This requires standardised approaches to assessing vector competence and the inclusion of under-studied taxa, including *Anopheles*, *Mansonia*, and other regionally abundant genera. Expanding vector coverage would improve detection of alternative transmission pathways and reduce taxonomic bias in surveillance data.

Third, integration of heterogeneous data sources is essential for improving predictive capacity. Combining field surveillance data with laboratory-derived vector competence data, environmental variables, plasmodium interactions and host distribution information would enable the development of more robust predictive models. Machine learning and network-based approaches offer promising tools for identifying potential vector–virus associations and prioritising surveillance efforts under conditions of data scarcity [14].

Importantly, existing malaria surveillance infrastructure provides a practical foundation for implementing such integrated systems. Leveraging established entomological monitoring programmes to include arbovirus screening could provide a cost-effective and scalable approach to expanding surveillance coverage across Africa. By aligning surveillance efforts across vector-borne diseases, it may be possible to generate more comprehensive datasets while maximising the utility of limited resources.

Conclusion

Global understanding of arbovirus emergence is shaped as much by where we look as by what we find. The apparent expansion of arboviruses into Europe has prompted important advances in surveillance and modelling, but these efforts remain constrained by a fundamental imbalance in data availability. Africa, despite its ecological importance, remains a major blind spot in global arbovirus surveillance.

Addressing this gap is essential for improving both empirical understanding and predictive capacity. By expanding surveillance in underrepresented regions, incorporating diverse vector systems, and adopting integrated, data-driven approaches, it will be possible to develop a more accurate and comprehensive picture of global arbovirus dynamics. Such efforts are not only necessary for understanding current transmission patterns but are critical for anticipating and mitigating future emergence in an increasingly interconnected world.

Ultimately, improving arbovirus prediction is not solely a matter of collecting more data, but of collecting the right data in the right places. Without addressing the structural biases in surveillance that underrepresent regions such as Africa, global models will continue to provide an incomplete picture of transmission risk. Until surveillance inequalities are addressed, global arbovirus models will continue to provide only a partial representation of transmission risk in an increasingly interconnected world.

Author Contributions: RSM conceptualized and designed the manuscript outline; wrote the initial draft; edited, revised, and finalized the manuscript.

Funding: This research was funded by MRC Discovery Medicine North (DiMeN) Doctoral Training Partnership, MR/W006944/1.

Ethical approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Reference

1. Rezza G. (2018). Chikungunya is back in Italy: 2007-2017. *Journal of travel medicine*, *25*(1), 10.1093/jtm/tay004. https://doi.org/10.1093/jtm/tay004
2. Murray, N. E., Quam, M. B., & Wilder-Smith, A. (2013). Epidemiology of dengue: past, present and future prospects. *Clinical epidemiology*, *5*, 299–309. https://doi.org/10.2147/CLEP.S34440
3. Calzolari, F., & Berninger, B. (2021). cAAVe phaenomena: Beware of appearances! Cell, 184(21), 5303–5305. https://doi.org/10.1016/j.cell.2021.09.027
4. Koenraadt, C.J.M., Münger, E., Schrama, M.J.J. *et al.* Overwintering of Usutu virus in mosquitoes, The Netherlands. *Parasites Vectors* **17**, 537 (2024). https://doi.org/10.1186/s13071-024-06620-y
5. Brustolin, M. et al. (2018) 'Experimental infection of mosquitoes with arboviruses: a systematic review', *Parasites & Vectors*, 11, p. 109.
6. Carlson, C.J. et al. (2019) 'Climate change increases cross-species viral transmission risk', *Nature*, 607, pp. 555–562.
7. Carlson, C.J. et al. (2021) 'The future of zoonotic risk prediction', *Nature Communications*, 12, pp. 1–10.
8. Ferguson, H.M. et al. (2010) 'Ecology: a prerequisite for malaria elimination', *PLoS Medicine*, 7(8), e1000303.
9. Kramer, L.D. and Ciota, A.T. (2015) 'Dissecting vectorial capacity for mosquito-borne viruses', *Current Opinion in Virology*, 15, pp. 112–118.
10. Morse, S.S. et al. (2012) 'Prediction and prevention of the next pandemic zoonosis', *The Lancet*, 380(9857), pp. 1956–1965.
11. Paz, S. (2015) 'Climate change impacts on West Nile virus transmission in Europe', *Current Opinion in Virology*, 13, pp. 1–7.
12. Sinka, M.E. et al. (2012) 'A global map of dominant malaria vectors', *Parasites & Vectors*, 5, p. 69.
13. Vilibic-Cavlek, T. et al. (2020) 'Emerging and re-emerging arboviruses in Europe', *Pathogens*, 9(3), 187.

doi:10.20944/preprints202606.1194.v1

14. Wardeh, M. et al. (2021) ‘Database of host–pathogen and related species interactions’, *Scientific Data*, 8, p. 297.

15. Woolhouse, M.E.J. et al. (2012) 'Human viruses: discovery and emergence', *Philosophical Transactions of the Royal Society B*, 367, pp. 2864–2871.
16. Mthawanji, R.S., Baylis, M., Wardeh, M., & Blagrove, M.S. (2026). Evaluating the Role of Anopheles Mosquitoes in the Global Spread of Arboviruses: A Review of Laboratory-Confirmed Viral Competence. *bioRxiv*.