Infectious Disease Modelling 10 (2025) 680–690

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Diphtheria transmission dynamics – Unveiling generation time and reproduction numbers from the 2022–2023 outbreak in Kano state, Nigeria

Raoul Kamadjeu [a, b, *], Oyeladun Okunromade [c], Bola Biliaminu Lawal [c], Muzammil Gadanya [c], Salma Ali Suwaid [d], Eduardo Celades Blanco [e], Ifedayo Adetifa [c], Elizabeth A. Kelvin [b, f]

[a] UNICEF, Program Group, Health Emergency Preparedness and Response, New York City, United States
[b] Department of Epidemiology and Biostatistics, CUNY Graduate School of Public Health & Health Policy, City University of New York, New York City, United States
[c] Nigeria Centre for Disease Control and Prevention, Abuja, Nigeria
[d] Murtal Muhammad Specialist Hospital, Kano State Ministry of Health, Kano State, Nigeria
[e] UNICEF Nigeria Country Office, Nigeria
[f] CUNY Institute for Implementation Science in Population Health, City University of New York, New York City, United States

ARTICLE INFO

Article history:
Received 10 September 2024
Received in revised form 26 January 2025
Accepted 11 February 2025
Available online 12 February 2025
Handling Editor: Dr Daihai He

Keywords:
Diphtheria
Outbreak
Nigeria
Reproduction number
Generation time
Transmission dynamics

ABSTRACT

Background: Diphtheria, caused by Corynebacterium diphtheriae, remains a serious public health threat in areas with low vaccination coverage, despite global declines due to widespread immunization and improved clinical management. A major outbreak in Nigeria from 2022 to 2023 underscored the persistent risk in regions with inadequate vaccination. This study aims to assess the transmission dynamics of diphtheria in Kano State, the epicenter of the outbreak, by estimating key epidemiological parameters, including the generation time (GT), approximated in our study by serial interval, and effective reproduction number (R_t).

Methods: We analyzed diphtheria case-based data from Kano State, Nigeria, collected between August 18, 2022, and November 29, 2023. Generation time was approximated using serial intervals in confirmed cases within the same geographical areas. The effective reproduction number (R_t) was calculated using four methods: Maximum Likelihood Estimation (MLE), Exponential Growth (EG), Sequential Bayesian (SB), and Time-Dependent (TD), focusing on the period of maximum exponential growth. A sensitivity analysis was conducted to quantify the impact of uncertainties in the GT derived from our data on the estimation of R_t.

Results: Over the 469-day outbreak period, 13,899 diphtheria cases were reported, with complete data available for 9406 cases. The estimated mean generation time was 2.8 days (SD = 3.48 days), with 97% of cases having a GT of less than 21 days. The R_t estimates varied across methods, with the TD method producing the highest reproduction number of 2.21 during the peak growth period. Sensitivity analysis showed that R_t estimates increased with longer generation times. The models, except for the SB method, demonstrated a generally strong fit with the outbreak exponential growth period.

* Corresponding author. UNICEF, Program Group, Health Emergency Preparedness and Response, New York City, United States
E-mail address: raoul.kamadjeu22@sphmail.cuny.edu (R. Kamadjeu).
Peer review under the responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2025.02.007

Conclusion: The ongoing diphtheria outbreak in Nigeria highlights the critical threat posed by declining vaccination coverage. This study provides valuable insights into the transmission dynamics of diphtheria during a prolonged and widespread outbreak, enhancing our understanding of disease spread in this context. While certain limitations may influence the interpretation of our estimates, the findings offer valuable information for future diphtheria outbreak preparedness and response in the African context.

1. Introduction

Diphtheria is a bacterial infection caused by *Corynebacterium diphtheriae*, characterized by the production of a potent and potentially lethal toxin. The primary mode of transmission of the disease is through respiratory droplets, though cutaneous diphtheria can occur through contact with infected lesions (Pinkbook: Diphtheria, 2022). Diphtheria can lead to severe complications, including breathing difficulties, heart damage, and nerve damage. Although most people recover with treatment, the disease has a case fatality rate of 5–10%, with an even higher risk of death in children under 5 years of age (Mayo Clinic [Internet]). Once widespread, diphtheria has seen a substantial decline in incidence due to the development and widespread adoption of effective vaccines, along with advances in clinical management through diphtheria antitoxin and antibiotics (Clarke et al., 2019). Nevertheless, outbreaks persist in regions with inadequate or decreasing vaccination coverage, including in Nigeria, which is among the top African countries with the highest number of unvaccinated children (New UNICEF report shows 12; Nigeria zero) and the country with the highest number of children who have not receive any routine vaccine (zero-dose children) (New UNICEF report shows 12; The Zero). The declining immunization coverage poses a continued threat of diphtheriae, particularly in African countries that accounted for nearly 16% (910 cases) of the 5856 cases reported globally to the World Health Organization (WHO) in 2022 (WHO [Internet]. World Health Organization). Inadequate surveillance and limited diagnostic capabilities in some affected countries likely lead to a significant underestimation of the true burden of diphtheria (Nicholson et al., 2022; Why diphtheria presents a growing).

Although Nigeria had previously reported sporadic diphtheria cases annually, the country experienced a gradual increase in cases beginning in December 2022. This was followed by an unusual surge between June and August 2023, during which nearly 6000 suspected cases were reported across 59 local government areas (LGAs) in 11 states. The majority of these cases (99.4%) were from the six states of Kano, Katsina, Yobe, Bauchi, Kaduna, and Borno (Diphtheria-Nigeria). By November 30, 2023, nearly 14,000 suspected cases had been reported in the country, including 9895 cases in Kano state — a state located in the Northern region of Nigeria, and one of the most populous of the country with a 2023 population estimates at 16,253,549 (List of Nigerian states by, 2024) – making this outbreak one of the largest diphtheria outbreaks ever documented in Africa (Sutherland & CNN, 2023).

According to the World Health Organization, the diphtheria outbreak in Nigeria was driven by low routine immunization coverage with the pentavalent vaccine, particularly among the pediatric population of whom an estimated 43% are unvaccinated, a figure well below the 80–85% coverage required to ensure community protection and preventing such large-scale outbreaks (Diphtheria vaccines).

Understanding disease transmission dynamics during outbreaks is crucial to inform public health interventions. Several metrics, including the generation time (GT) (Svensson, 2007a) and the basic reproduction number (R_0) (Delamater et al.) are used to describe the transmission dynamics of infections. This information can help public health officials understand the severity of the outbreak and disease transmission patterns to inform the development of potential control strategies.

1.1. Generation time (GT) and basic reproduction number (R_0)

1.1.1. Generation time (GT)

The generation time is defined as the time between the infection of a primary case and one of its secondary cases (Svensson, 2007a), or the time interval between the infection of a primary case (infector) and the subsequent infection of a secondary case (infectee) within a transmission pair (Lehtinen et al., 2021). GT is an important concept in infectious diseases modeling and has been used to estimating the basic reproduction number. GT is typically estimated by analyzing the time lag between identified infection pairs; however, such data are often unavailable during outbreaks, especially during large, geographically spread and prolonged outbreaks. Consequently, GT is often approximated using serial intervals (SI) defined as the average time between the infection of a primary case and the subsequent infection of others by that individual, or the time between two similar events such as the onset of disease (Lehtinen et al., 2021; Svensson, 2007b). Consequently, SI are usually less cumbersome to measure during outbreaks and are expressed as units of time (e.g. hours, days, weeks). A systematic review by Truelove et al.(Truelove et al., 2020) reported an average GT of 7.8 days (95% CI: 6.3–9.7 days) for diphtheria, while

Flavio et al. (Finger et al., 2019) estimated a GT of 4–5 days in a short and localized outbreak setting, illustrating how the GT can vary depending on the characteristics of the outbreak and the methods and assumptions underlying their measure. GT and SI play a crucial role in public health planning and interventions by indicating the transmission window for an infected individual. Throughout the paper, we use SI as a proxy to the GT.

1.1.2. Reproduction number

The reproduction number (R) defines the average number of secondary cases generated by a single primary case (Svensson, 2007a; Delamater et al.). This metric helps gauge outbreak severity and informs control strategies. Two variants of the reproduction number exist, including.

- R_0: The basic reproduction number represents the average number of secondary cases during the initial outbreak phase when the entire population is susceptible and without control measures.
- R_t: The effective reproduction number is a time-varying adjustment to R_0, reflecting changes in population immunity over time and the impact of control measures.

In practice, R_t is estimated during outbreaks and varies considerably due to factors like vaccination coverage, public health interventions, and population behavior. It then changes over time as more people become infected and subsequently immune to the disease and as control measures such as expanded vaccination, in the case of vaccine preventable diseases, are implemented. Truelove et al. (Truelove et al., 2020) reported an R_t range of 1.7–4.3 for diphtheria, while Matsuyama et al. (Matsuyama et al., 2018) found R_t to range from 4.7 to 14.8 in a specific refugee camp setting in Bangladesh, demonstrating the potential for significant variation in R_t over time, across contexts and analytical approaches.

Data on diphtheria transmission dynamics in Africa remains limited, largely due to the scarcity of data from well-documented outbreaks. This gap in knowledge constrains our understanding of diphtheria's spread to inform public health response. This study estimates diphtheria transmission parameters, notably the generation time and effective reproduction number (R_t), using data from the 2022–2023 diphtheria outbreak in Kano State. Our research aims to highlight the challenges associated with estimating these key infectious disease parameters in an Africain setting, while contributing valuable insights to inform outbreak preparedness and response strategies in similar contexts.

2. Methods

2.1. Data and data source

We utilized case-based data (line-list) from Kano State, covering the period from August 18, 2022, to November 29, 2023. These data were collected as part of the diphtheria outbreak response through the surveillance system established by the Nigeria Centre for Disease Control (NCDC) in collaboration with the Kano State health authorities.

The raw dataset included 7 variables and 13899 record. A detailed description o the dataset is provided in the supplemental material. The variables in the diphtheria outbreak dataset relevant to our analysis were the following.

- State of residence of the case
- Date of onset of symptoms
- Final classification of the case laboratory confirmed, clinically compatible, epidemiological link, discarded (non-diphtheria)
- Outcome: death or alive
- Othe variables included in the dataset provided by NCDC, but not used in our analysis included the age and sex of the cases, and their vaccination status against diphtheria.

Cases were identified through a dual approach, combining passive reporting from health facilities with active reporting involving community health workers conducting house-to-house active case search and contact tracing. Each case in the line-list could be assigned one of the following classifications of the disease status.

- Laboratory confirmed: for cases with a laboratory confirmation of diphtheria through laboratory analysis of a collected sample
- Clinically compatible: for cases with no laboratory diagnosis, but with signs and symptoms compatible with diphtheria, in the context of the outbreak,
- Epidemiological linked (epi-linked): for cases with no laboratory diagnosis, but with direct connection (close contact, shared exposure) with a laboratory confirmed or clinically compatible cases

For our analysis, confirmed diphtheria cases included laboratory-confirmed, clinically compatible and epi-linked cases.

2.2. Estimation of the serial time distribution

The generation time, defined as the time interval between the infection of a primary case (infector) and the subsequent infection of a secondary case (infectee) within a transmission pair, in an interrupted chain of transmission, has been used to estimating the basic reproduction number. The infector-infectee pair could not be derived from our data since the actual time of infection was unknown; therefore, we used the serial interval, defined as the onset of symptoms in a primary case (infector) and the onset of symptoms in a secondary cases (infectee) to approximate the generation time for confirmed cases (Svensson, 2007b) within the same smallest administrative unit of residence (LGA), The infector-infectee pair was determined by the order of the symptom onset date of the cases within a cluster, and used to generate a distribution of serial intervals within a cluster such as:

$$Serial\ Interval = T_{Symptom\ onset(x)} - T_{Symptom\ onset(y)}$$

with $T_{Symptom\ onset(y)}$ and $T_{Symptom\ onset(y)}$ representing the dates of symptom onset for two consecutive cases (infectee, infector), with data sorted in ascending dates of symptom onset.

The serial intervals calculated for each of the LGAs reporting at least 2 confirmed cases of diphtheria were accumulated to construct a distribution of serial intervals. We modeled the serial interval distribution using several candidate distributions selected based on the observed form of the serial interval distribution. These candidate distributions were evaluated for goodness of fit using criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The parameters of the best-fitting distribution (mean, standard deviation) were then used in our estimation of the reproduction numbers.

2.3. Modelling the diphtheria outbreak

We described the dynamic diphtheria incidence for the entire period of data availability through a combination of visual inspection and modeling approaches. We identified periods of increasing and decreasing transmission using the regression of log-incidence method (Cori et al., 2013) implemented with the *incidence* package in R-Studio (Jombart et al.). The log-incidence method involves three steps: first, log-transformation of the incidence data; second, fitting a linear regression model to the log-transformed data as a function of time and examining the slope of the regression line to determine whether the incidence of cases is increasing (phase of growth), decreasing (phase of decay), or remains stable over time (Cori et al., 2013; Nishiura et al., 2009; Wallinga & Teunis, 2004) and third, identifying periods of increasing or decreasing disease transmission/incidence based on the value of the slope of the regressions and using the slope value to derive the outbreak doubling and halving times (days, weeks).

2.4. Identification of the period of maximum growth

To estimate the R_t, we focused on the period of maximum growth of the outbreak. The period of maximum growth for an outbreak refers to the time interval during which the incidence of cases increases most rapidly, typically reflecting the phase of exponential growth in the epidemic curve (Ma, 2020). Considering the length of the outbreak, the period of increasing transmission (i.e. the period from the date of the first case to the date of the maximum incidence) was further subdivided into intervals equal to the doubling time identified in growth phase of the outbreak, from the day of highest incidence towards the start of the outbreak. The trend of the outbreak was then examined for each of the intervals using the log-incidence method. The interval exhibiting the highest deviance r-squared statistic was identified as the period of maximum exponential growth and used for estimating the effective reproduction numbers.

2.5. Estimation of the effective reproduction numbers

We estimated the effective reproduction number (R_t), a time-varying reflection of R, considering changes in immunity and control measures in the population.

Four approaches were employed to estimate the effective reproduction number from the identified period of maximum exponential growth: the Maximum Likelihood Estimation (MLE) method, the Exponential Growth (EG) method, the Sequential Bayesian (SB) method, and the time-dependent (TD) methods. These four methods are commonly used to estimate key epidemiological parameters in infectious diseases dynamics (Anderson & May 1992, p. 766).

2.5.1. Maximum likelihood estimation (ML) method

We utilized the maximum likelihood estimation approach proposed by White and Pagano (White et al., 2021) to estimate *R*. The approach assumes the number of secondary cases caused by an index case is Poisson distributed with expected value R. Given observation of $(N_0, N_1, \ldots, N_T)$ incident cases over consecutive time units (days, weeks, etc …) and a generation time distribution w, R is estimated by maximizing the log-likelihood as:

$$LL(R) = \sum_{t=1}^{T} log\left(\frac{e^{-\mu_t}\mu_t^{N_t}}{N_t!}\right) \quad \text{Eq[1]}$$

Where $\mu_t = R\sum_{i=1}^{t} N_t - {}_i w_i$ with the likelihood calculated on a period of exponential growth and the deviance r-squared used to select the most appropriate method of exponential growth. The method makes no assumption on mixing in the population.

2.5.2. Exponential growth (EG) method

We used the approach summarized by Wallinga and Lipsitch (Wallinga & Lipsitch, 2007), demonstrating that the exponential growth rate during the early phase of an outbreak can be linked to the initial reproduction ratio. We denoted the exponential growth rate as r as the per-capita change in number of new cases per day and was estimated using a Poisson regression, with the reproduction number estimated as

$$Eq[2] : R = \frac{1}{M(-r)}$$

with M as the moment generating function of the generation time distribution, and r the exponential growth rate. The approach requires choosing a period in the epidemic curve over which growth is exponential, guided by the deviance-based r-squared statistics, with no assumption made on mixing in the population.

2.5.3. Sequential Bayesian method (SB)

The SB method allows a sequential and real time estimation of the initial reproduction number(Bettencourt & Ribeiro, 2008). The SB approximates an SIR model with incidence at time $t+1, N(t+1)$ can be approximated to a Poisson distribution with mean $N(t)e^{(\gamma(R-1))}$ where $\frac{1}{\gamma}$ is the average duration of the infectious period. The SB updates the R distribution as new data is obtained using the algorithm:

$$P(R|N_0, \ldots, N_{t+1}) = \frac{P(N_{t+1}|R, N_0, \ldots, N_t)P(R|N_0, \ldots N_t)}{P(N_0, \ldots N_{t+1})} \quad \text{Eq[3]}$$

The SB approach uses an exponential distribution of the generation time and assumes random mixing in the population.

2.5.4. Time dependent reproduction numbers (TD)

Time-dependent reproduction number approach (Fraser, 2007; Wallinga & Teunis, 2004) computes reproduction numbers by averaging over all transmission networks compatible with observations. The probability P_{ij} that case i with onset at time t_i was infected by case j with onset at time t_j is calculated as:

$$P_{ij} = \frac{N_i w(t - i - t_j)}{\sum_{i \neq k} N_i w(t_i - t_k)}. \quad \text{Eq[4]}$$

The effective reproduction number for case j is estimated as $R_j = \sum iP_{ij}$ and is averaged as $R_j = \frac{1}{N_t}\sum_{t_j=t} R_j$ over all cases with the same date of onset. Estimates of the number of cases at time t as a function of the number of prior cases multiplied by their infectiousness, as described by equation (5) below.

$$Eq[5] : R_0 = \frac{N(t)}{\sum_{t=1}^{k} w(\tau)N(t-\tau)}$$

We applied the R_t estimation methods to the diphtheria epidemic curve, using the distribution best fitted to the generation time. To assess the reproduction rates, we defined a targeted time window selected from the day with the highest case incidence, backward, by periods of 60 days, purposely chosen to be around the tenth of the entire outbreak period. Using the deviance r-squared statistic across each of the temporal segments of interest, we identify the segment exhibiting the highest deviance r-squared statistic as the period of maximum exponential growth. This period of maximum exponential growth of the outbreak was used in our estimations of the reproduction numbers.

2.6. Sensitivity analysis

Recognizing that the assumption of fully linked chains of transmission within LGAs is unlikely to hold for the diphtheria outbreak in Kano—potentially introducing bias in the estimation of the generation time (GT)—we conducted a sensitivity analysis. This involved calculating values of the effective reproduction number (R_t) across a range of GT values. The analysis aimed to quantify the impact of uncertainties in the GT derived from our data on the estimation of R_t.

2.7. Data analysis

For all our analysis of R_t, we followed the approach described by Boelle et al. (Boelle & Obadia, 2023). All analyses were done using R-Studio (RStudio) and the *R0* (Boelle & Obadia, 2023)and *Incidence* (Jombart et al.) packages. Results are presented as graphs and tables.

3. Results

3.1. Description of the diphtheria epidemic

The available data for the outbreak covered the period from August 18, 2022, to November 29, 2023, spanning 469 days. During that period, a total of 13,899 diphtheria cases were reported through the outbreak surveillance system in Kano State of which 9406 had data with a final diagnosis. Most cases (9178, 97.6%) were classified as clinical compatible with diphtheria, 143 (1.5%) laboratory confirmed diphtheria, and 85 (0.9%) epidemiologically linked to a clinically compatible or laboratory confirmed case. The cases ranged in age from 1 to 70 years and originated from 46 LGAs of Kano State, with three states accounting for more than half of all those cases (Fig. 1)

Overall, visual inspection of the epidemic curve reveals a clear pattern of increasing transmission leading to a peak incidence, followed by a subsequent decline. Closer examination of the phase of increased transmission shows a sharp and dramatic rise in case numbers from late May 2023, culminating in the peak incidence around mid-August 2023, before transitioning into a phase of declining transmission. However, the factors driving this rapid acceleration could not be determined from the available data.

3.2. Modeling the epidemic curve

The regression of log-incidence (r) over time (in days) applied to the 469-days period of the outbreak revealed two notable transmission patterns centered around the August 27, 2023 (375th day of the outbreak). The first phase of the outbreak (**Phase 1**) was characterized by a long period (from day 1–375) of low but gradual increase in incidence (deviance r-squared statistic (r) = 0.00981, 95%CI: 0.0078–0.0117, doubling time of 70.63 days (95%CI: 59.00–88.03 days)), peaking at 123 cases on day 358 (Aug 10, 2023); and a second phase (**Phase 2**) from day 376 (August 27, 2023) till the end of the series of steady

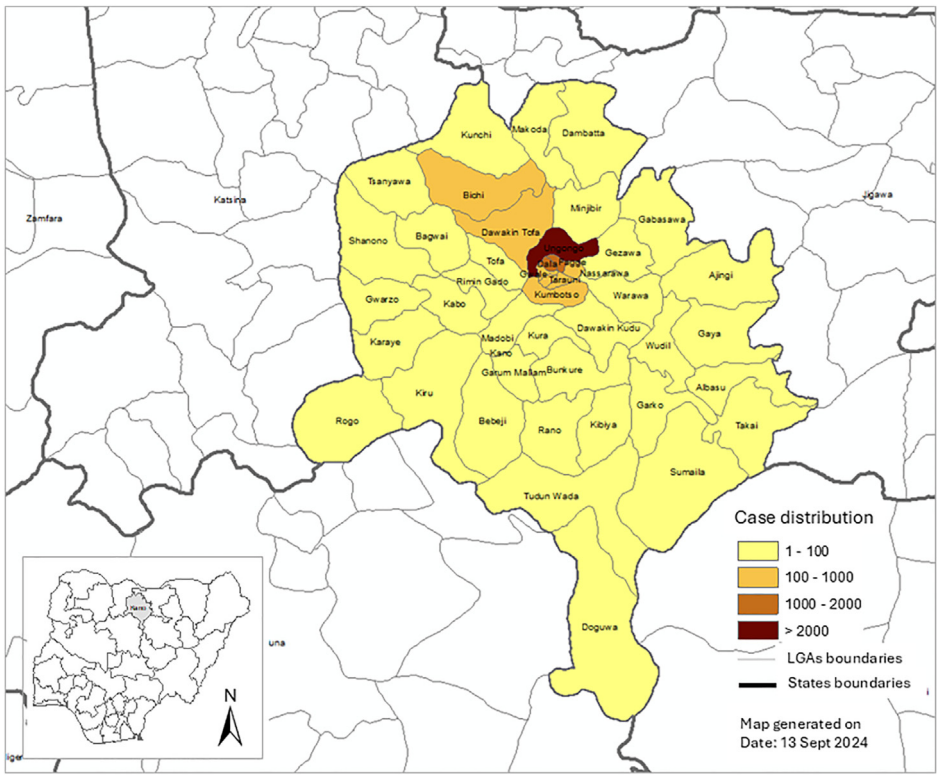

Fig. 1. Distribution of confirmed mpox diphtheria cases within local government authorities of Kano state, Nigeria, from Aug 18, 2022 to Nov 29, 2023.

decline in daily incidence (r = −0.0101; 95%CI: −0.0123 -0.0079, with a halving time of 68.414 days (95%CI: 56.3280–87.1057)) (Fig. 2).

3.3. Estimation of the generation time distribution

The generation time (GT) distribution, approximated in our study by the serial interval distribution, was accumulated from serial intervals from each of the 42 LGAs reporting at least two cases is presented in Fig. 3.

The serial interval distribution was fitted to three theoretical distributions: gamma, lognormal, and Weibull. Among the tested distributions, the lognormal distribution showed the best overall fit, with the lowest Akaike's Information Criterion (AIC) (6294.71) and Bayesian Information Criterion (BIC) (6306.17) values compared to the gamma (AIC: 7570.43, BIC: 7581.89) and Weibull (AIC: 7763.12, BIC: 7774.58) distributions. Additionally, the lognormal distribution performed slightly better in the C-vM (Cramer-von Mises) statistic (88.44) compared to gamma (94.44), although Weibull had the lowest C-vM statistic (87.00), indicating a strong overall fit. Weibull distribution had the smallest Kolmogorov-Smirnov statistic (0.379), suggesting it aligned most closely with the empirical cumulative distribution function. However, the Anderson-Darling statistic, which is sensitive to differences in the tails, favored the lognormal distribution (438.47) over gamma (461.59) and Weibull (422.06). Based on the combination of goodness-of-fit criteria and statistical diagnostics, we concluded that the serial interval distribution was best fitted with a lognormal distribution, with a mean of 2.80 days and a standard deviation of 3.48 days, a minimum value of 1 day, and 97% of the GT distribution below 21 days (Fig. 3-A).

3.4. Identification of period of maximum growth

Because of the length of the outbreak, that first period was secondarily, arbitrarily, subdivided into approximately to 5-time intervals of 70 days each, corresponding to the outbreak doubling time, starting from the day of highest incidence, moving toward the start of the outbreak. The 70-day interval, ranging from day 288 to day 358 (from June 1, 2023 to August 23, 2023) of the outbreak yielded the highest deviance r-squared statistic (r = 0.078; 95% CI: 0.072–0.086) with a doubling time of 8.77 days (95% CI: 8.07–9.60). This timeframe was identified as the period of maximum exponential growth and was used for estimating the reproduction numbers. Fig. 2 overlays on the epidemic curve, the modeled trends for the two phases of the outbreak and the period of maximum exponential growth.

3.5. Calculation of reproduction numbers

We confined our analysis of reproductive numbers to the phase marked with the highest exponential growth (288 to day 358 of the outbreak). Primary inputs in our calculations included the generation time estimated for the entire outbreak and the case incidence series during the period of maximum exponential growth.

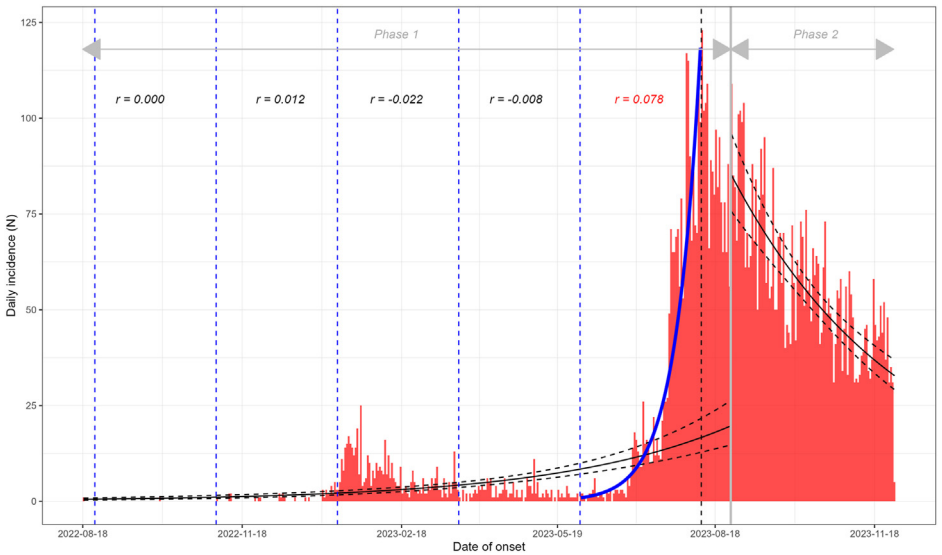

Fig. 2. Overall epidemic curve of the diphtheria outbreak in Kano State, by day of onset, from Oct 18, 2022 to Nov 29, 2023. The grey vertical line set on Aug 27, 2023 splits the outbreak into Phase 1 (growth) and Phase 2 (decay). The solid blue line marks the day (Aug 10, 2023) with highest incidence. The dashed blue lines mark the limits of the interval used to identify the period of maximum growth, with their corresponding r-squared values (*r*). The curves model the incidence of diphtheria (with 95% confidence interval) for the entire outbreak (black curves), and the period of maximum growth (solid blue curve, r ≈ 0.078).

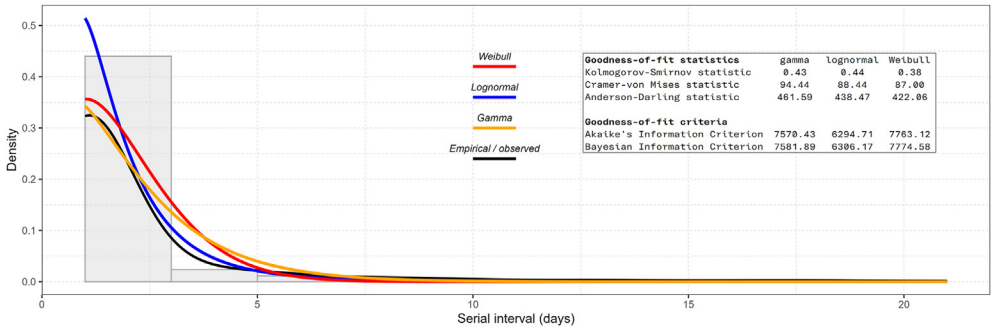

Fig. 3. Modeling the diphtheria outbreak serial interval distribution of the 2022–2023 outbreak in Kano (Nigeria). The curves show the empirical/observed, Weibull, lognormal, and gamma distributions, with associated goodness-of-fit estimates. The serial interval distribution is best approximated with a lognormal distribution with mean 2.80 and a standard deviation of 3.48 days.

The distribution of reproduction numbers generated using the TD and SB approaches is depicted in Fig. 3-B, with a comprehensive summary of estimates provided in Table 1. Notably, point estimates and median values from the SD and TD methods did not exhibit significant differences. Among the methods employed, the Time-Dependent approach yielded the highest reproduction number with a maximum value of 2.21.

When fitting the four models to the exponential growth phase of the epidemic, a generally strong fit was observed, except for the SB method (Fig. 4-A).

3.6. Assessing the influence of the generation time on R_t estimation - sensitivity analysis

To evaluate the impact of the chosen generation time on our reproduction number estimates, we performed a sensitivity analysis by computing values of R_t across a range of generation times, varying from 1 to 6 days in increments of 1 day for the ML and EG methods only. The outcome of this analysis is depicted in Fig. 4-B.

As expected, there is an increase in the value of R_t with increasing values of generation times (Fig. 3B). Our results indicate that the estimated reproduction number of the epidemic ranges from 1.1 to 1.40 for the ML method and from 1.14 to 1.55 for the EG method, with incremental values of the generation time from 1 to 6 days.

4. Discussion

This study provides a pragmatic assessment of the reproduction number and generation time of diphtheria during a large and lengthy outbreak in Kano State, Nigeria.

The effective reproduction number for the Kano 2022–2023 diphtheria outbreak was estimated at 1.25 and 1.18 using the EG and ML methods, respectively, with higher maximum values of 1.7 and 2.21 observed using the SB and TD methods. While most methods showed a strong fit with the data during the period of maximum exponential growth, the SB performed less effectively. We stipulate that this discrepancy can be attributed to the method's assumption of a smooth and gradual change in R_t over time, which does not align with the observed fluctuations in transmission dynamics during the outbreak. These fluctuations may reflect changes in surveillance sensitivity, the implementation of localized interventions, and variations in public behavior, all of which can introduce variability in transmission that the SB method may fail to capture adequately.

Our calculations of R_t were consistent with Trulove et al. study (Truelove et al., 2020) in their pool analysis of 23 outbreaks in which R_t ranged from 1.7 to 4.3, but they significantly diverged from the median estimate of 7.2 obtained by Matsuyama et al. (Matsuyama et al., 2018) based on a diphtheria outbreak in a refugee camp in Bangladesh.

Our estimates of the serial interval, used as proxy to the generation time (mean GT 2.8 days) in our analysis differ from those reported in the literature; Finger et al. estimated a diphtheria generation time of 4–5 days, while Truelove et al. reported a serial interval—a proxy for generation time—of 7.8 days, suggesting faster cycles of transmission of diphtheria in

Table 1
Estimates of reproductive numbers derived from the period of highest exponential growth (day 298–358) of the 2022–2023 diphtheria outbreak in Kano (Nigeria).

	Methods (Mean R_t and 95% CI)			
Estimates	EG	ML	SB	TD
R_t	1.25 (1.23–1.26)	1.18 (1.10–1.25)	1.14 (1.08–1.21)	1.26 (1.19–1.35)

Bootstrap was used to estimate the mean and 95% CI (normal) of the R_t for SB and TD. Median and extremes values of R_t distributions generated using SD and TD methods were 1.16 (Max: 1.95) and 1.17 (Max: 2.20) respectively.
R_t: effective reproduction number, EG: Exponential growth; ML: Maximum likelihood; SB: Sequential Bayesian, TD: Time-dependent.

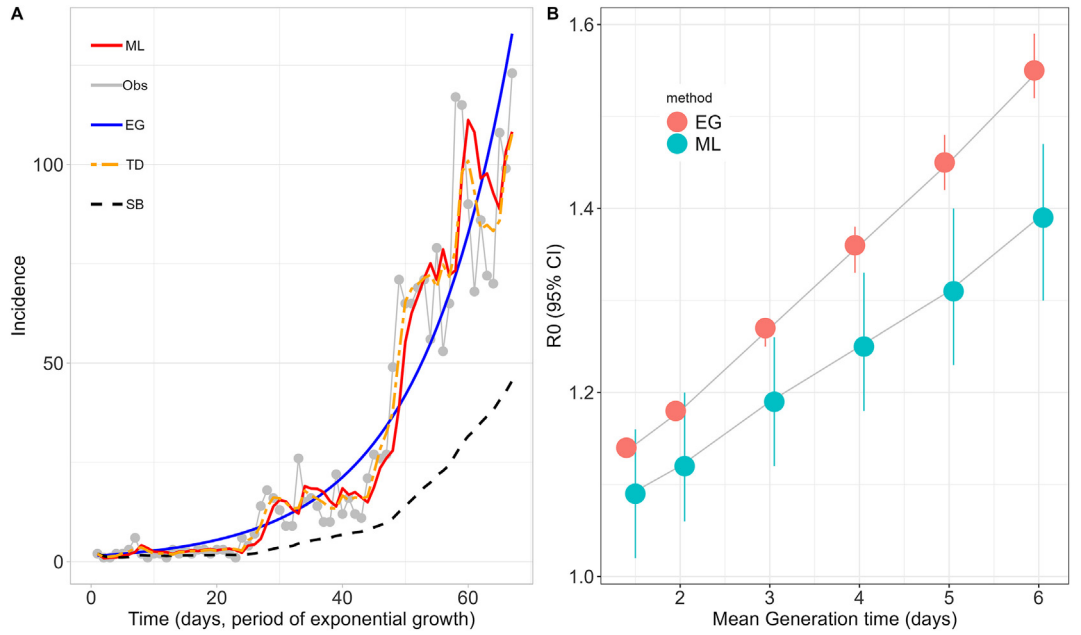

Fig. 4. Modeling diphtheria exponential growth, and sensitivity analysis of changes in R_t with generation time. A) Modeling the period of exponential growth of the epidemic with the four approaches (EG, ML, SB, and TD) to estimate the reproduction numbers. The red dashed vertical line represents the mean of the distribution; B) Sensitivity analysis assessing the change in estimation of Rt (EG and ML methods) with generation time.
EG: Exponential growth, ML: maximum likelihood. TD: time-dependent, SB: Sequential Bayesian, Obs: Observed.

Kano Truelove's analysis found that 95% of cases had a serial interval shorter than 21 days, which falls within the range of our estimates for 95% of cases. The disparities in the estimates of the GT are unsurprising given the differences in the estimation methods, variabilities in transmission contexts, scale and duration of the outbreaks and differences in case ascertainment. Truelove et al. estimates are based on a literature review, and Flavio et al. estimates are derived from a localized outbreak, whereas our estimates are based on data from a prolonged outbreak spanning a large geographic area (Kano State). Furthermore, the estimation of generation time assumes a continuous chain of transmission within the same geographical areas, limited to confirmed cases, and without case importation. Given the geographic extent of some local government areas (LGAs), these assumptions may not hold true for most cases. Furthermore, most of the cases in our database were clinically compatible rather than laboratory-confirmed, suggesting that some of the cases may not have been diphtheria, which could further contribute to the differences in our estimates.

The relationship between the GT and R_t underscores the importance of accurately estimating the generation time, as over- or underestimating GT can lead to significant bias in the calculation of Rt. In outbreak settings, such biases may affect the assessment of transmission potential and the design of public health interventions. To assess the potential of such bias and the impact of uncertainties in the GT on the estimation of Rt, we conducted a sensitivity analysis across a range of GT values.

The sensitivity analysis revealed variability in R_t estimates based on the selected GT values, underscoring the dependency of R_t on the choice of GT. This finding suggests that biases in GT estimation will propagate into R_t calculations, potentially affecting the accuracy and interpretation of transmission dynamics.

Even when using GT values closer to those reported in the literature (Finger et al., 2019; Truelove et al., 2020), our R_t estimates remained on the lower end of the spectrum compared to previously published findings (Islam et al., 2022). These discrepancies highlight potential data reliability issues that have likely influenced our estimation of the GT. First, the assumption of fully linked chains of transmission within Kano is unlikely to hold, leading to missed transmission chains. Additionally, uncertainties surrounding case diagnostics and the heavy reliance on clinically compatible cases—which constituted much of our dataset—may have resulted in the inclusion of non-diphtheria cases. These factors, coupled with general data quality challenges, likely contributed to compressed GT estimates and subsequently lower R_t values. This contrasts with the longer GT and higher R_t values observed in other outbreaks with eventually more robust data.

5. Limitations

Our study is one of the few to describe the diphtheria transmission dynamics during a prolonged and geographically widespread outbreak in the African context. It is important to acknowledge its limitations, which warrant caution in interpreting our findings.

The approach we used in our analysis assumed that all diphtheria cases were detected by the surveillance system, linked in a continuous chain of transmission, and that no cases were imported. However, these assumptions are unlikely to fully apply to the diphtheria outbreak in Kano, given the outbreak's duration and wide geographical scope with increased likelihood of external case importation, unrelated, broken and missed transmission chains. To maximize the likelihood of capturing linked chains of transmission, we estimated the generation time at the LGA Level, the lowest administrative level for cases available in the dataset. However, LGAs are not the smallest administrative subdivisions in Kano State and cover substantial geographic areas. As a result, our GT and R_t estimates likely reflect not a single, continuous localized outbreak, but rather multiple outbreaks occurring in various locations. This suggests that our parameters will not fully capture the transmission dynamics at the more localized levels.

Moreover, the prolonged periods of low incidence may suggest ongoing transmission that was not detected by the surveillance system. Variations in population heterogeneity, disparities in the ascertainment of diphtheria diagnoses — particularly with less than 2% of cases confirmed through laboratory testing — and spatial differences in transmission dynamics across Kano State, possibly due to non-uniform implementation of control measures, all contribute to the observed differences in our findings. Finally, the methods used to estimate reproduction numbers may account for differences in the estimates obtained; for instance, Matsuyama et al. (Matsuyama et al., 2018) employed a Latin Hypercube Sampling Method and accounted for the loss of susceptibility over time.

6. Conclusion

The diphtheria outbreak in Nigeria, one of the largest and longest in recent history, is a reminder of the persistent threat posed by vaccine-preventable diseases amid declining vaccination coverage. While certain limitations may influence the interpretation of our estimates, the findings offer valuable information for future diphtheria outbreak preparedness and response strategies in the African context. This study sheds light on the transmission dynamics of this unprecedented outbreak, contributing to the limited body of knowledge on diphtheria transmission during protracted and geographically widespread outbreaks in Africa. It also provides critical insights to support outbreak control efforts, particularly in settings often marked by suboptimal surveillance systems and constraints in data quality and availability.

CRediT authorship contribution statement

Raoul Kamadjeu: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Oyeladun Okunromade:** Writing – review & editing, Conceptualization. **Bola Biliaminu Lawal:** Writing – review & editing, Data curation. **Muzammil Gadanya:** Writing – review & editing, Investigation, Data curation. **Salma Ali Suwaid:** Writing – review & editing, Investigation. **Eduardo Celades Blanco:** Writing – review & editing, Validation, Conceptualization. **Ifedayo Adetifa:** Writing – review & editing, Validation, Project administration, Investigation, Conceptualization. **Elizabeth A. Kelvin:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Data availability statement

Permission to access the dataset was requested to and approved by the Surveillance Directorate of Nigeria Centres for Disease Control (Nigeria CDC). The data used for this study is not publicly available. Requests for data on the 2022–2023 diphtheria outbreak in Nigeria should be directed to the Surveillance Directorate of Nigeria CDC.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used AI (ChatGPT 4o) in order to improve language and readability. After using this ChatGPT 4o, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding sources

None.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2025.02.007.

References

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford, New York: Oxford University Press.

Bettencourt, L. M. A., & Ribeiro, R. M. (2008). Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One, 3*(5), Article e2185.

Boelle, P. Y., & Obadia, T. (2023). R0: Estimation of R0 and real-time reproduction number from epidemics [internet] [cited 2023 Dec 8]. Available from: https://cran.r-project.org/web/packages/R0/index.html.

Clarke, K. E. N., MacNeil, A., Hadler, S., Scott, C., Tiwari, T. S. P., & Cherian, T. (2019). Global epidemiology of diphtheria, 2000–20171. *Emerging Infectious Diseases, 25*(10), 1834–1842.

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505–1512.

Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R0) - Volume 25, Number 1—January 2019 - Emerging Infectious Diseases journal - CDC. [cited 2024 Mar 7]; Available from: https://wwwnc.cdc.gov/eid/article/25/1/17-1901_article.

Diphtheria vaccines: WHO position paper – August 2017 [Internet]. [cited 2023 Dec 8]. Available from: https://www.who.int/publications-detail-redirect/who-wer9231.

Diphtheria-Nigeria [Internet]. [cited 2024 Dec 4]. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON485.

Finger, F., Funk, S., White, K., Siddiqui, M. R., Edmunds, W. J., & Kucharski, A. J. (2019). Real-time analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in Bangladesh. *BMC Medicine, 17*(1), 58.

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One, 2*(8), e758.

Islam, Z., Ahmed, S., Rahman, M. M., Karim, M. F., & Amin, M. R. (2022). Global stability analysis and parameter estimation for a diphtheria model: A case study of an epidemic in Rohingya refugee camp in Bangladesh. *Computational and Mathematical Methods in Medicine, 2022*, Article 6545179.

Jombart T, N. Kamvar. Overview of the incidence package [Internet]. [cited 2024 Dec 5]. Available from: https://cran.r-project.org/web/packages/incidence/vignettes/overview.html.

Lehtinen, S., Ashcroft, P., & Bonhoeffer, S. (2021). On the relationship between serial interval, infectiousness profile and generation time. *J R Soc Interface, 18*(174), Article 20200756.

List of Nigerian states by population. In . *Wikipedia [internet]*, (2024) [cited 2024 Dec 5]. Available from: https://en.wikipedia.org/w/index.php?title=List_of_Nigerian_states_by_population&oldid=1261029014.

Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infect Dis Model, 5*, 129–141.

Matsuyama, R., Akhmetzhanov, A. R., Endo, A., Lee, H., Yamaguchi, T., Tsuzuki, S., et al. (2018). Uncertainty and sensitivity analysis of the basic reproduction number of diphtheria: A case study of a Rohingya refugee camp in Bangladesh, november-december 2017. *PeerJ, 6*, Article e4583.

Mayo Clinic [Internet]. . Diphtheria-Diphtheria - Symptoms & causes. [cited 2024 Aug 27], Available from:: https://www.mayoclinic.org/diseases-conditions/diphtheria/symptoms-causes/syc-20351897.

New UNICEF report shows 12.7 million children in Africa missed out on one or more vaccinations over three years [Internet]. [cited 2024 Sep 1]. Available from: https://www.unicef.org/esa/press-releases/new-unicef-report-shows-127-million-children-africa-missed-out-one-or-more.

Nicholson, L., Adkins, E., Karyanti, M. R., Ong-Lim, A., Shenoy, B., Huoi, C., et al. (2022). What is the true burden of diphtheria, tetanus, pertussis and poliovirus in children aged 3–18 years in asia? A systematic literature review. *International Journal of Infectious Diseases, 117*, 116–129.

Nigeria zero-dose landscape [Internet]. [cited 2024 Sep 1]. Available from: https://zdlh.gavi.org/sites/default/files/2024-01/ZDLH_Nigeria%20_Landscape_2023.pdf.

Nishiura, H., & Chowell, G. (2009). The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In G. Chowell, J. M. Hyman, L. M. A. Bettencourt, & C. Castillo-Chavez (Eds.), *Mathematical and statistical estimation approaches in epidemiology [internet]* (pp. 103–121). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2313-1_5.

Pinkbook: Diphtheria. (2022). CDC [Internet], [cited 2024 Mar 7]. Available from: https://www.cdc.gov/vaccines/pubs/pinkbook/dip.html.

RStudio: Integrated Development for R. RStudio, PBC, Boston, MA [Internet]. Posit. [cited 2023 Aug 15]. Available from: https://www.posit.co/.

Sutherland C. CNN. 2023 Available from:[cited 2024 Sep 3]. Fears as West Africa battles worst diphtheria outbreak in recent times. https://www.cnn.com/2023/12/08/africa/fears-as-west-africa-battles-worst-diphtheria-outbreak-in-recent-times/index.html.

Svensson, Å. (2007a). A note on generation times in epidemic models. *Mathematical Biosciences, 208*(1), 300–311.

Svensson, Å. (2007b). A note on generation times in epidemic models. *Mathematical Biosciences, 208*(1), 300–311.

Truelove, S. A., Keegan, L. T., Moss, W. J., Chaisson, L. H., Macher, E., Azman, A. S., et al. (2020). Clinical and epidemiological aspects of diphtheria: A systematic review and pooled analysis. *Clin Infect Dis Off Publ Infect Dis Soc Am., 71*(1), 89–97.

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B Biol Sci, 274*(1609), 599–604.

Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology, 160*(6), 509–516.

White, L. F., Moser, C. B., Thompson, R. N., & Pagano, M. (2021). Statistical estimation of the reproductive number from case notification data. *American Journal of Epidemiology, 190*(4), 611–620.

WHO [Internet]. World Health Organization; . GHO | By category | Diphtheria - Reported cases by WHO region. Available from:[cited 2024 Jan 30] https://apps.who.int/gho/data/view.main.1520_41.

Why diphtheria presents a growing threat to our health | Gavi, the Vaccine Alliance [Internet]. [cited 2024 Mar 7]. Available from: https://www.gavi.org/vaccineswork/why-diphtheria-presents-growing-threat-our-health.

o The Zero-Dose Child: Explained [Internet]. [cited 2024 Sep 3]. Available from: https://www.gavi.org/vaccineswork/zero-dose-child-explained.