Supplementary Appendix 2 (Technical Supplement): Clinical and Epidemiological Aspects of Diphtheria

Shaun A Truelove[1,‡], Lindsay T Keegan[1,‡,*], William J. Moss[1,2], Lelia H. Chaisson[1], Andrew S. Azman[1,3], Emilie Macher[3], and Justin Lessler[1,*]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[2]International Vaccine Access Center, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

[3]Médecins Sans Frontières, Geneva, Switzerland

[*]justin@jhu.edu lindsay.keegan@utah.edu

[‡]these authors contributed equally to this work

Contents

Table 1. Table of parameters used in the technical supplement, their meaning, and for fixed parameters, their value.

Parameter	Definition	Estimate
	Vaccination	
VE	Vaccine effectiveness	3 doses are 87.2%
p	Proportion of the infected population experiencing symptomatic disease	varies by outbreak
p_k	Proportion symptomatic after receiving k doses of the vaccine	$p_0 = 0.69$, $p_{1,2} = 0.52$, $p_3 = 0.12$
V_k	Proportion receiving k doses of vaccine	varies by outbreak
r	Reported vaccination coverage among incompletely or unvaccinated individuals in the population	varies by outbreak
	Relative transmissibility	
X	Ratio of the number of secondary infections from a symptomatic case to the number of secondary infections from an asymptomatic carrier	4.3
β_S	measure of daily infectiousness for symptomatic cases	
β_A	measure of daily infectiousness for asymptomatic carriers	
τ	Relative infectiousness of asymptomatic individuals	0.24
δ	Relative infectiousness of treated symptomatic cases	0.49
	Duration of infection	
$g(t)$	Proportion of untreated infections still infectious t days later	
$g_T(t)$	Proportion of treated cases still infectious t days later	
$T(t)$	Proportion of treated cases still infectious t days after treatment	
	Reproductive numbers	
R_0	Basic reproductive number	2.73
R	Effective reproductive number	varies by outbreak
R_S	The reproductive number for untreated, symptomatic cases	3.19
R_A	The reproductive number for asymptomatic carriers	0.76
$\tilde{R}_S$	The reproductive number for symptomatic cases where some proportion are treated	varies by outbreak
R_T	The reproductive number for treated, symptomatic cases	1.56
R_{3d}	The reproductive number for individuals fully vaccinated (3 doses)	1.05
R_{3d}	The reproductive number for individuals never vaccinated (0 doses)	2.44
	Intervention	
α	Proportion of symptomatic cases that are treated	varies by outbreak
x	the proportion of symptomatic infections that are not treated	varies by outbreak
V_c	Critical vaccination threshold in an untreated population	varies by outbreak
$V_{c,T}$	Critical vaccination threshold in an treated population	varies by outbreak
x	The proportion of symptomatic cases that need to be treated to reduce R below 1	varies by outbreak

1 Clinical Course and Natural History

1.1 Proportion asymptomatic, p_k

To estimate the proportion of diphtheria infections who have been vaccinated k times that are asymptomatic, we pooled data from all closed outbreaks (i.e., within schools) with information on the immunization status of all infections (2 studies).[1,2] We estimated the proportion of individuals who experience asymptomatic infection who have never received diphtheria vaccine p_0.

To calculate the proportion of individuals who experience asymptomatic infection with imperfect vaccination (1-2 doses), $p_{1,2}$ and the proportion of individuals who experience asymptomatic infection who have received full vaccination (3 doses), p_3, we use to the vaccine effectiveness. We know that the vaccine effectiveness is how effective the vaccine is at preventing symptoms relative to unvaccinated individuals. So we know that:

$$1 - \mathrm{VE}_3 = \frac{p_3}{p_0} \tag{1}$$

Re-arranging, we solve for p_3 and find:

$$p_3 = (1 - \mathrm{VE}_3)p_0 \tag{2}$$

likewise, given the vaccine effectiveness for incomplete vaccination,

$$1 - \mathrm{VE}_{1,2} = \frac{p_{1,2}}{p_0} \tag{3}$$

Re-arranging, we solve for $p_{1,2}$ and find:

$$p_{1,2} = (1 - \mathrm{VE}_{1,2})p_0 \tag{4}$$

We estimate p_3 and $p_{1,2}$ within the hierarchical Bayesian framework used to estimate R_0 (see section 4 below).

2 Epidemiology

2.1 Proportion of secondary cases caused by asymptomatic individuals, τ

To calculate the relative contribution of asymptomatic carriers over the course of infection, we first assume that given someone is colonized with diphtheria, the relative transmissibility is constant over time. Then, we write

$$X = \frac{\beta_S \int g(t)dt}{\beta_A \int g(t)dt} \tag{5}$$

where X is the relative number of secondary infections caused by symptomatic cases compared to asymptomatic carriers (Doull and Lara 1925[3]); we estimated this to be $\frac{59/758}{14/779}$. β_S and β_A are measures of daily infectiousness for symptomatic cases and asymptomatic carriers, respectively; and $g(t)$ is the proportion of untreated infections still infectious t days later. Here we assume that ratios of secondary transmissions to household contacts between asymptomatic and symptomatic individuals is representative of the ratio of secondary infections by asymptomatic and symptomatic individuals overall.

We define $\tau := \beta_A/\beta_S$, thus, solving for τ, we find

$$\tau = \frac{1}{X} \tag{6}$$

Using the number of infected and uninfected contacts in Doull and Lara 1925[3], we estimate τ in a Bayesian framework where the number of secondary infections from either symptomatic or asymptomatic infections were binomially distributed.

3 The Reproductive Numbers, R_0 and R

Because vaccination and infection do not produce immunity to infection, but rather immunity to symptomatic disease, and because of asymptomatic infections, which are typically unobserved, we need to derive new equations for the basic and effective reproductive numbers that account for this unique transmission process.

3.1 Derivation of the theoretical equation for the effective reproductive number, R

The effective reproductive number, R, is the average number of secondary infections from a single (average) infected individual in a population with some level of existing immunity, such as from vaccination or previous infection. Because with diphtheria immunity is in terms of immunity to symptomatic disease, rather than immunity to infection, we needed to derive a new equation that describes the relationship between vaccination/prior infection and symptomology. To start with an "average" individual, we need to account for both symptomatic and asymptomatic transmission and treatment. We can write R as a

weighted sum of the contributions by symptomatic and asymptomatic infected individuals as follows:

$$R = \sum_{k=0}^{3} (V_k)\left((p_k((1-\alpha)R_S + \alpha R_T) + (1-p_k)R_A\right) \tag{7}$$

where V_k is the proportion of the population receiving k doses (such that $\sum V_k = 1$), p_k is the proportion of the population which are symptomatic with k doses of the diphtheria vaccine, R_S is the average number of infections caused by a single infected symptomatic individual in an otherwise totally susceptible population (the symptomatic reproduction number), R_T is the average number of secondary infections caused by a single treated symptomatic case in an otherwise totally susceptible population, R_A is the analogous term for untreated asymptomatic infections, and α is the proportion treated.

3.2 Derivation of the basic reproductive number, R_0

From the equation for the effective reproductive number we can derive an equation for the basic reproductive number, R_0, where we assumes a completely immunologically naive population (i.e., no vaccination or prior infection) and no treatment with antibiotics. Setting $V_{1,2} = 0\%$ and $V_3 = 0\%$, and thus $V_0 = 100\%$, we are left with the following equation for R_0:

$$R_0 = p_0 R_S + (1-p_0)R_A \tag{8}$$

where p_0 is the proportion of the population that are symptomatic (despite being completely naive to diphtheria infection, a significant portion of infections will be asymptomatic).

4 Estimating Reproductive Numbers from Data

We estimated the basic and effective reproductive numbers for diphtheria using data from 23 outbreaks occurring from 1901-2018. From these data, we first estimated outbreak specific effective reproductive numbers, R, using methods described by White and Pagano which estimate R from the early exponential growth period of the epidemic.[4] Using these estimates of R, we estimate a basic reproductive number, R_0, across all outbreaks using the diphtheria-specific equations derived above, which account for the contribution of asymptomatic infections to transmission and the impact of vaccination on rates of symptomatic/asymptomatic infection. All of this was implemented in a hierarchical Bayesian framework, which we implemented using `R`, `Stan`, and the `R` package `rstan`.

4.1 Outbreak data

We used data from 23 separate outbreaks occurring from 1901-2018 for which daily or weekly case counts were available. For outbreaks in which case count data were aggregated by week, we estimated daily case counts by fitting a spline to the cumulative density function of the epidemic curve to produce probabilities for each day of each week, from which daily cases

were imputed to generate a dataset of daily case counts that was used in the model[5,6].

4.2 Outbreak vaccination data

For each outbreak, vaccination coverages for each stratum of doses (0, 1-2, 3+) were specified from one of three forms of available data: reported vaccination coverage for the population (reported from the paper), reported vaccination among cases, or estimated vaccination coverage from the WHO (for country-level outbreaks). We did not include outbreaks where these data were not available.

For studies which only reported the vaccination among cases, we calculated the population-level vaccination coverage using our estimates for proportion symptomatic for each vaccination dose group (p_k). For early studies prior to the existence of the 3-dose primary vaccination series, we estimated the proportion receiving 0 or $1-2$ doses (V_0 and $V_{1,2}$) using the following system of equations:

$$r = \frac{p_{1,2}V_{1,2}}{p_{1,2}V_{1,2} + p_0(1-V_{1,2})} \tag{9}$$

$$1 = V_0 + V_{1,2} \tag{10}$$

Where r is the reported vaccination coverage among incompletely or unvaccinated individuals in the population, V_0 is the proportion of the population with 0 doses, and $V_{1,2}$ is the proportion of the population with $1-2$ doses of the vaccine.

For studies from the three-dose vaccine series era, we can estimate these proportions when the population coverage with DTP3 is reported but incomplete or unvaccinated proportions are not reported. We use *Eq. 9* again, but where $V_0 + V_{1,2} + V_3 = 1$ and V_3 is known.

4.3 Outbreak R estimation

Our Bayesian framework first uses an adaptation of previously described methods from White and Pagano to estimate the R for each outbreak during its initial exponential growth phase.[4] This method also allows for simultaneous estimation of the serial interval (time between symptom onsets of consecutive infections in a transmission chain) using surveillance data (i.e., cases by time). However, for the the final results, we specified the serial interval as what we previously estimated from individual-based data (gamma distribution with mean=1.36 and sd=5.97)). This was done to maintain consistency with our previous results, in which we have more confidence due to coming from individual data (further, the resulting estimated serial interval was similar to our individual-based estimate).

The likelihood function for R was as follows:

$$L(R) = \prod_{t=1}^{T} \frac{e^{-\mu_t} \mu_t^{N_t}}{N_t!} \tag{11}$$

where $\mu_t = R \sum_{j=1}^{min(k,t)} N_{t-j} p_i$, and N is the number of new cases at time t, k is a constraint such that $k < T$, and p is the probability vector of the multinomial distribution describing the probability of a case occurring on a specific day, assuming the serial interval.

4.4 Bayesian hierarchical framework

We integrated the White and Pagano methods for estimating R into a Bayesian hierarchical framework to estimate R_S (and thus R_0), based on a similar decomposition of R as shown in 7, where $R_A = \tau R_S$:

$$R = \sum_{k=0}^{3} (V_k)\,((p_k((1-\alpha)R_S + \alpha R_T) + (1-p_k)\tau R_S) \tag{12}$$

In this model, the $\log(R_s)$ from each of the 23 outbreaks was assumed to come from a common normal distribution with mean $\log(R_s^\star)$ and standard deviation σ. Both $\log(R_s^\star)$ and σ were assigned weakly informative priors (normal/truncated normal). We assigned a normal prior distribution to, $\log(\tau)$, the (log) proportion of secondary cases caused by asymptomatic individuals, with a mean of -1.437 and standard deviation of 0.09, based on our estimates described above. The relative transmission contribution of treated symptomatic individuals, δ, was assumed come from a uniform distribution from 0.4-0.6, which implies an average delay to antibiotic treatment of 2-5 days. For outbreaks before discovery of antibiotics, we set δ to 0. $VE_{1,2}$, VE_3, and p_0 (used to estimate p_k) were assumed to follow truncated normal distributions with means and standard deviations drawn from our separate analyses ($VE_{1,2}$: mean=0.705, sd=0.110; VE_3: mean=0.872, sd=0.047; p_0: mean=0.69, sd=0.097). V_k parameter estimates were allowed to follow normal distributions using the vaccination coverage data above for the means, and standard deviation assumed to be 10% of these means. See *Supplementary Appendix 1, Table S3* for full parameter estimates and data sources.

Using the joint posterior for the universal R_S, p_0, p_3, and τ, we calculated R_0 following *Eq. 8* above, and V_c following *Eq. 17* below. This model and estimation methods can been found in the source code repository (https://github.com/shauntruelove/Diphtheria) in the scripts `Run_R0stan.R` and `R0stan_knownGT_fixedCoV.stan`.

4.5 Sensitivity analysis: Variability of vaccination coverage estimates

For our initial estimation of R_0 from outbreak data, we wanted to allow the model to have flexibility in the vaccination coverage parameter. As such, for each outbreak we input priors for the mean vaccination coverage of 0, 1-2, and 3 or more doses, and we allow individual sampling distributions for coverages with each dose to follow a normal distribution with a standard deviation equivalent to the standard across all outbreaks for that dose, normalized by the mean for the outbreak ($sd_m = CoV \cdot \mu_m$). As a result, the posterior distributions for the V parameters were broad. In order to not completely discard vaccination coverage data, we conducted a we conducted a sensitivity analysis restricting the standard deviations of the sampling distributions to 10% of the means. As a result, V_c was reduced by 1.8% and R_0 by 12.3%. To maintain a balance between allowing the model to freely estimate the parameters and maximizing the data that we have, we use the estimates from the restricted variance model for our final calculations.

5 Diphtheria Control and Outbreak Response

5.1 Critical vaccination threshold, V_c

To determine the critical vaccination threshold, we first determine what proportion of the population needs to have received three doses of diphtheria vaccine (V_3) to achieve herd immunity. For the critical vaccination threshold, we assume that individuals either have full vaccination or no vaccination, so we can re-write *Eq. 7* as:

$$R = V_0(p_0 R_S + (1-p_0)R_A) + V_3(p_3 R_S + (1-p_3)R_A) \tag{13}$$

$$\tag{14}$$

To achieve herd immunity, we want to estimate V_3 when $R < 1$, so we solve for V_3:

$$1 > V_0(p_0 R_S + \tau(1-p_0)R_S) + V_3(p_3 R_S + \tau(1-p_3)R_S) \tag{15}$$

$$\tag{16}$$

Thus, the critical vaccination threshold is:

$$V_c = \frac{1 - p_0 R_S - \tau R_S + p_0 \tau R_S}{(p_0 - p_3)R_S(\tau - 1)} \tag{17}$$

5.2 Proportion reduction in transmission from vaccination

To determine by what proportion three doses of diphtheria vaccine reduces transmission, we calculate the reproductive number for fully vaccinated individuals, R_{3d}, and the reproductive number of never vaccinated individuals, R_{0d}, and take the ratio.

We define R_{3d} as:

$$R_{3d} = p_3 R_S + (1 - p_3)\tau R_S \tag{18}$$

and R_{0d} as:

$$R_{0d} = p_0 R_S + (1 - p_0)\tau R_S \tag{19}$$

Thus the ratio is:

$$\frac{p_3 R_S + (1 - p_3)\tau R_S}{p_0 R_S + (1 - p_0)\tau R_S} = 0.41 \tag{20}$$

Thus, we show that fully vaccinated individuals cause 69% fewer secondary infections on average, over the course of their infection.

5.3 Critical vaccination threshold with treatment, $V_{c,T}$

The critical vaccination threshold determines what proportion of the population needs to be vaccinated to stop the spread of diphtheria, however, this assumes no treatment. To adjust for antibiotic treatment, we assume that only symptomatic individuals are treated and thus we can adjust the reproductive number of symptomatic individuals, where

$$\tilde{R_S} = (1 - \alpha)R_S + \alpha R_T \tag{21}$$

$\tilde{R_S}$ is the reproductive number for symptomatic individuals assuming some proportion, α, are treated and R_T is the reproductive number for treated, symptomatic individuals. *Note:* we assume that asymptomatic individuals are not detected as cases and therefore are not treated.

We can then relate R_T and R_S as follows:

$$R_S = \delta R_T \tag{22}$$

where δ is the relative infectiousness of treated individuals. To calculate δ we follow the process in equation 5, where

$$delta = \frac{\beta_S \int_0^{30} (\alpha g_T(t) + (1-\alpha) g(t)) dt}{\beta_S \int_0^{30} g(t) dt} \tag{23}$$

and $\delta = \beta_S / \beta_T$. We assume that $g_T(t)$ follows the distribution of carriage of untreated cases, $g(t)$, described in section 2.1, prior to treatment and $g(t) \cdot T(t)$ after the initiation of treatment. From this equation, we estimate δ in our hierarchical Bayesian framework. From this, we find $\delta = 0.49$.

Figure S9 shows the distribution $g_T(t)$ compared to $g(t)$ for a treatment delay of 5.9 days.[7–17]

To calculate the critical vaccination threshold with treatment, $V_{c,T}$, we return to equation 17 where:

$$R = V_0(p_0 \tilde{R_S} + (1-p_0) R_A) + V_3(p_3 \tilde{R_S} + (1-p_3) R_A) \tag{24}$$

We set $R < 1$ to find the critical vaccination threshold, and solve for V_3, knowing $V_0 = 1 - V_3$:

$$1 > V_0(p_0 \tilde{R_S} + (1-p_0) R_A) + V_3(p_3 \tilde{R_S} + (1-p_3) R_A) \tag{25}$$

Thus, the critical vaccination threshold with treatment is is:

$$V_{c,T} = \frac{-1 + p_0 R_S - p_0 \alpha R_S + p_0 \alpha \delta R_S - \tau R_S - p_0 \tau R_S}{(p_0 - p_3) R_S (1 - \alpha + \alpha \delta - \tau)} \tag{26}$$

5.4 Isolation of Symptomatic Cases Only

In many cases, it is more practical to treat only symptomatic infections. To determine what proportion of symptomatic cases need to be isolated to halt transmission, we write

$$x p_0 R_S + (1-p_0) R_A < 1 \tag{27}$$

Solving for x, we find

$$x = \frac{1 - (1-p_0) R_A}{p_0 R_S} \tag{28}$$

where x is the proportion of symptomatic infections that are not treated.

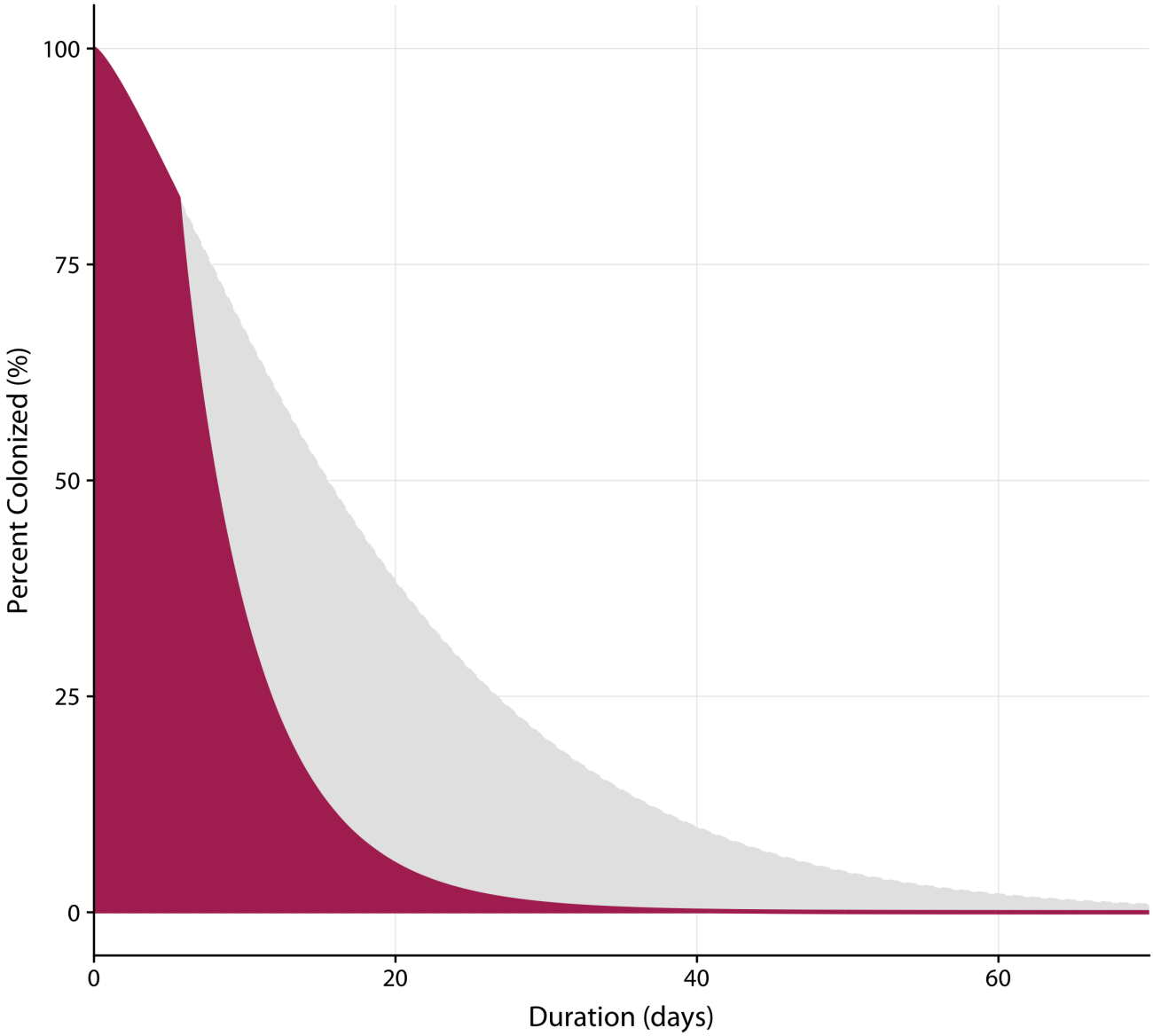

Figure S9. Clearance of C. diphtheriae in cases treated with antibiotics 2 days after developing characteristic symptoms (maroon), and untreated symptomatic infections (grey)

References

1. Fanning, J. Diphtheria in an immunized community. *Br. Med. J.* **1**, 371 (1947).

2. Miller, L. W., Older, J. J., Drake, J. & Zimmerman, S. Diphtheria immunization: effect upon carriers and the control of outbreaks. *Am. J. Dis. Child.* **123**, 197–199 (1972).

3. Doull, J. A. & Lara, H. The epidemiological importance of diphtheria carriers. *Am. J. Epidemiol.* **5**, 508–529 (1925).

4. Forsberg White, L. & Pagano, M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat. medicine* **27**, 2999–3016 (2008).

5. Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, 569–593 (Springer, 1992).

6. Harel, O. *et al.* Multiple imputation for incomplete data in epidemiologic studies. *Am. journal epidemiology* **187**, 576–584 (2017).

7. CG Gutherie, BC Marshall, and WL Moss. Experimental inoculation of human throats with virulent diphtheria bacilli. *Bull. Johns Hopkins Hosp.* **XXXII** (1921).

8. Webb, J. W. & MacFADYEN, N. Field investigation of an outbreak of diphtheria. *Br. Med. J.* **2**, 248 (1947).

9. Heath Jr, C. W. & Zusman, J. An outbreak of diphtheria among skid-row men. *New Engl. J. Medicine* **267**, 809–812 (1962).

10. Lys, H. G. AN EPIDEMIC OF DIPHTHERIA. *Lancet* **136**, 968–969 (1890).

11. Chalmers Parry, S. An investigation of a small epidemic of diphtheria in the hop-fields. *Public Heal.* **53**, 223–224 (1939).

12. Kehr, M. J., Tannahill, R. W., Elias-Jones, T. F. & Whittaker, L. Observations on a present-day outbreak of diphtheria. *Public Heal.* **74**, 294–303 (1960).

13. An outbreak of milk-borne diphtheria due to carrier cases. *Public Heal.* **22**, 149–150 (1908).

14. Miller, L. W., Bickham, S., Jones, W. L., Heather, C. D. & Morris, R. H. Diphtheria carriers and the effect of erythromycin therapy. *Antimicrob. Agents Chemother.* **6**, 166–169 (1974).

15. Beach, M. W., Gamble, W. B., Jr, Zemp, C. H., Jr & Jenkins, M. Q. Erythromycin in the treatment of diphtheria and diphtheria carrier state. *Pediatrics* **16**, 335–344 (1955).

16. Bixby, E. W. The effectiveness of penicillin in the treatment of nasopharyngeal diphtheria. *Am. J. Med. Sci.* **215**, 509–513 (1948).

17. Kocher, R. A. & Siemsen, W. J. Diphtheria carriers treated with penicillin. *Ann. Intern. Med.* **24**, 883–886 (1946).