Methods to Determine the End of an Infectious Disease Epidemic: A Short Review

Hiroshi Nishiura

Abstract Deciding the end of an epidemic is frequently associated with forthcoming changes in infectious disease control activities, including downgrading alert level in surveillance and restoring healthcare workers' working shift back to normal. Despite the practical importance, there have been little epidemiological and laboratory methods that were proposed to determine the end of an epidemic. This short review was aimed to systematically discuss methodological principles of a small number of existing techniques and understand their advantages and disadvantages. Existing epidemiological methods have been mostly limited to a single-and-brief exposure setting, while the application to human-to-human transmissible disease epidemic with stochastic dependence structure in the observed case data has remained to be a statistical challenge. In veterinary applications, a large-scale sampling for laboratory testing has been commonly adapted to substantiate a freedom from disease, but such study has only accounted for binomial sampling process in estimating the error probability of elimination. Surveillance and mathematical modeling are two complementary instruments in the toolbox of epidemiologists. Combining their strengths would be highly beneficial to better define the end of an epidemic.

Keywords Epidemic · Ebola · Epidemic elimination · Incubation period · Exposure · Polio · Heuristic method

1 Prologue

Rather than declaring the start of an epidemic, it has been harder to determine the end of the epidemic. Erroneous declaration of the start might be understood as part of errors in risk assessment practice, and such an occasional error might not impose serious irreversible damage to our society. However, deciding the end of an

H. Nishiura (✉)
Hokkaido University, Kita 15 Jo Nishi 7 Chome,
Hokkaido, Kita-ku Sapporo-shi 060-8638, Japan
e-mail: nishiurah@gmail.com

epidemic is frequently associated with forthcoming changes in infectious disease control activities, and its decision imposes a pressure to epidemiologists to a certain extent. Declaring the end of an epidemic, the alert level in surveillance system may be downgraded, and working shift of medical and public health experts in charge of control practice (e.g. contact tracing effort) may also be restored [1]. In the case of the end of a vaccine preventable disease, the declaration would always lead to a discussion over the cessation of routine immunization [2]. The impact of the end of an epidemic is not limited to healthcare settings. Reduced volume of travelers may be recovered to normal due to declaration of the end of an outbreak, and thus, the declaration of safety would involve a pressure from tourism industry and have substantial impact on associated economics. In the case of an epizootic event of a veterinary disease, especially among livestock animals, the freedom from the epizootic disease indicates a permission to restart international transportation or trade of specific animals [3]. Getting along with these social and political interests, the end of an epidemic must be determined without serious errors and the announcement should be made carefully and appropriately.

Despite the importance of the determination of the end of an epidemic, there have been little available methods to explicitly judge the end of an epidemic [4]. In particular, published studies have been mostly limited to a setting with single and brief exposure, e.g. a point source outbreak of food-borne disease. While methods are scarce, there have been multiple practical events on the ground that did require explicit methodological assistance in deciding the end of an epidemic. Nevertheless, it is also true that practical side has involved a number of complications that cannot be immediately addressed by epidemiological modeling only. For instance, many epidemics have involved a substantial number of asymptomatic infections, ascertainment biases and underreporting issues. In passive surveillance, diagnosed cases are notified to the public health authority. In addition to passive information, there might be datasets based on active surveillance (e.g. case finding effort through outbreak investigation) or laboratory testing of (a part of) possible exposed individuals, but their utilities have not been taken into account in the determination of the end of an epidemic. Moreover, one may ultimately wish to judge the end of an epidemic, not using notified case data but using other informative resources such as event-based or syndromic surveillance data.

Facing these complexities in empirical observation, what can epidemiological modelers offer to the society? The purpose of this short review is to understand methodological principles of available criteria of the end of an epidemic, identifying their advantages and disadvantages. This exercise will shed light on future path of the objective judgment of the end of infectious disease epidemics.

2   Classical WHO Approach

The most stimulating practice has been seen in the adoption of classical criteria by the World Health Organization (WHO) on its definition of zero Ebola cases from 2013-15 [6]. In that criteria, the outbreak of EVD is considered ended in any one of affected

Fig. 1 *Probability density function of the incubation period of Ebola virus disease (EVD)*. The daily frequency of the incubation period, the time from infection to illness onset, for EVD is shown [5]. The mean and variance of the incubation period were assumed at 9.7 and 30.3 days^2, respectively. A lognormal distribution was employed, and parameters μ and σ were thus 2.13 and 0.53, respectively
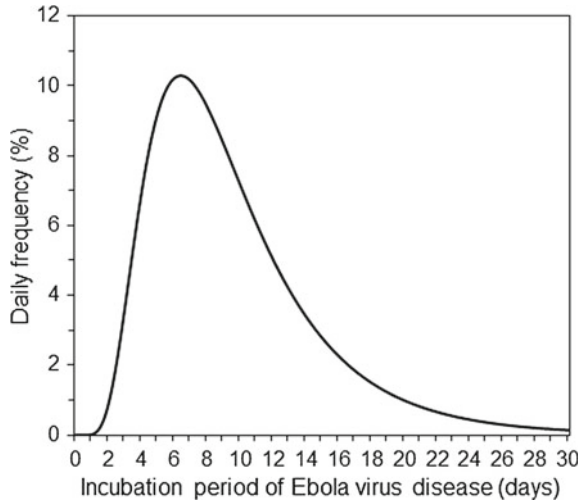
countries (e.g. Guinea, Liberia and Sierra Leone) after 42 days have passed since the last confirmed case has tested negative twice for the virus on blood samples. Along with this criterion, it has been also suggested that, after the 42-day period has elapsed, each country should maintain a system of heightened surveillance for a further 90 days, and ensure that ongoing EVD surveillance and notification thereafter will be conducted. Moreover, it is determined that the end of the EVD outbreak in the West African sub-region will be declared when the 42-day period has elapsed in the last affected country.

The choice of 42 days stems from the right tail of the incubation period. Figure 1 shows the probability density function of the incubation period of EVD [5]. Empirically observed certain maximum of the incubation period has been 21 days. Taking a double of this empirical maximum value, 42-day waiting period has been determined. Unfortunately, there has been no additional justification of using the incubation period and taking twice the empirical maximum, but the choice of a fixed length has been very transparent to public health societies and the criteria were made easy to follow for those working on the ground in West African countries.

Incubation period is the time from infection to illness onset [7]. As long as the time of potential exposure among traced contacts is known, the incubation period could indicate the length of time to be waited to ensure that no more symptomatic case exists. Even provided that the time of potential exposure is not directly observed, the use of the latest time at which an exposure could have occurred (e.g. the last date of PCR positive outcome in the last confirmed case) as "clock zero" point would offer a conservative suggestion to ascertain the absence of additional symptomatic infections [8].

Nevertheless, despite the simple and transparent fixed length, the classical approach suffers from a number of technical problems. First, the use of empirically observed maximum would be vulnerable to sample size of the incubation period.

Namely, the greater the sample size, the greater the observed value of maximum would be [9]. Second, it is evident that the use of the incubation period is not justified for diseases that involve substantial number of asymptomatic infections [10]. There could be unrecognized chains of transmission among asymptomatic cases. Third, due to the shortage of objectiveness, the waiting period does not directly measure the probability of the end of an epidemic. For instance, it appeared that viable Ebola virus could be maintained in semen of infected males even after recovery from convalescent phase. A number of sexual transmission events have been reported to have fueled local reemergence of EVD, but such event has never been captured by the right tail of the incubation period. As a consequence, several erroneous declarations of the end of Ebola epidemic were unfortunately observed in West Africa.

3   Single Exposure Approach

Food-borne outbreak is frequently referred to as the common source outbreak, because the causative food is mostly shared among exposed individuals. The point source outbreak is a special case of common source outbreak in that the exposure is very brief in time (e.g. sharing an identical party lunch menu on the same day). The point source outbreak has been well studied by statisticians, because the resulting epidemic curve can be assumed as identical to the density function of the incubation period (Fig. 2), permitting us to estimate the time of exposure and analyze a variety of statistical features of that distribution.

Determination of the end of point source outbreak is perhaps the most well studied statistical subject in the context of the end of outbreak. Figure 2 shows the typical
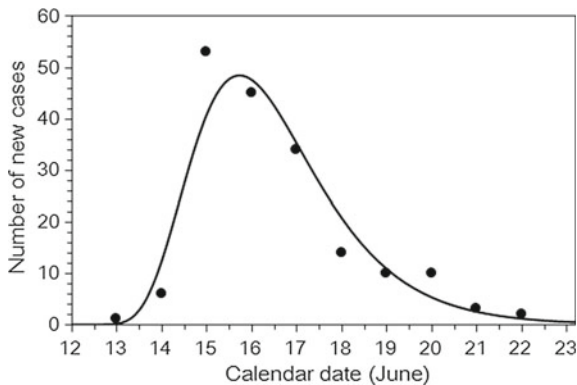
Fig. 2 *Fitting a three-parameter log-normal distribution to the epidemic curve of Salmonellosis in Gifu prefecture, Japan, 2003.* An outbreak of food-borne Salmonellosis was observed in Gifu prefecture involving a total of 178 cases [7]. A three-parameter log-normal distribution includes not only μ and σ but a threshold parameter that determines the time at which an exposure occurred

epidemic curve of food-borne outbreak, caused by Salmonellosis in Gifu, Japan, 2003. To capture the epidemic pattern, one can fit the following three-parameter log-normal distribution:

$$f(t; \gamma, \mu, \sigma^2) = \frac{1}{(t - \gamma)\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln(t - \gamma) - \mu)^2}{2\sigma^2}\right), \tag{1}$$

for $t - \gamma > 0$, where t is the calendar time and γ is the so-called threshold parameter indicating the time at which an exposure occurred. In the Gifu outbreak example, the maximum likelihood estimate of γ was 11.7 on the calendar time scale in June 2003, indicating that the most likely brief exposure may have happened at lunch or dinner on 11 June. In many food-borne outbreaks, food traceback effort during the outbreak investigation involves a serious problem of recall bias. However, employing the model (1), one could dramatically narrow down the scope of food menus to be recalled [7].

In addition to estimating the time of exposure, one can subsequently assess the right tail in detail, because the percentile of the incubation period distribution directly indicates the proportion of cases that we have already observed by a given calendar time. Brookmeyer and You [4] have exploited this knowledge to develop a hypothesis testing method. Suppose that the total outbreak size is N among which we have already observed n cases. We have the ordered calendar time of disease onset of cases, $y_1, \ldots, y_n$ and suppose that T days have passed since the last case (y_n) occurred. The hypotheses are $H_0 : N > n$ versus $H_1 : N = n$.

For the hypothesis testing, we consider the jth spacing $s_j = y_{j+1} - y_j$. Assuming that the incubation period follows a two-parameter exponential model with a guarantee time G, i.e., $f(u) = \lambda \exp(-\lambda(u - G))$ for $u > G$ and 0 for $u < G$, jth spacing arising from a sample size of N from the two parameter exponential model also has an exponential distribution with parameter $\lambda(N - j)$, and thus, the density function of the spacing is

$$f(s_j) = \lambda(N - j) \exp(-\lambda(N - j)s_j). \tag{2}$$

The probability that the nth spacing is greater than t days is

$$\Pr(s_n > t) = \exp(-\lambda(N - n)t) \tag{3}$$

Let α be the level of significance test. The length of waiting time t is set such that the Eq. (3) is equal to α at the particular null hypothesis when $N = n + 1$. Then, we obtain

$$T \geq -\frac{1}{\lambda}\ln(\alpha) \tag{4}$$

In general for any $N > n$, the probability of rejecting H_0 is

$$\Pr\left(s_n \geq -\frac{1}{\lambda}\ln(\alpha)\right) = \alpha^{(N-n)} \tag{5}$$

While the method is statistically very solid, the range of direct application is unfortunately limited to the point source outbreak. Moreover, the application is justified only when all of infected individuals develop symptoms and all cases are reported. Despite these problems, the proposed method is maintained very simple and can be implemented in some other settings with a little extensions, especially when the spacing of a single distribution can be applied.

4   Laboratory Testing to Ensure the Absence of Cases

In veterinary epidemiological practice, a mass laboratory testing may be more easily implemented than in human population. Due to economic interest to urge the government to be acknowledged as being free from a specific disease and resume trade, the cost that is required for laboratory testing may be justified well. Obtaining laboratory samples even from a part of the population, the following assessment would be feasible.

Suppose that we have a perfect laboratory test and we handle infinitely large population of animals. The probability of selecting a given number of positives when randomly selecting n animals from a population with disease prevalence p is given by the binomial distribution [11]:

$$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{(n-x)} \tag{6}$$

Using the perfect test (i.e. with 100 % sensitivity and 100 % specificity), an epidemiological survey to substantiate freedom from disease requires that no positives are found. When $x = 0$, the Eq. (6) simplifies to:

$$\Pr(X = 0) = (1 - p)^n \tag{7}$$

Even in the case that we should consider imperfect laboratory testing, the abovementioned scheme can be easily extended [11]. The probability of observing x positive animals when testing n animals from an infinite population is given by the binomial distribution:

$$\Pr(X = x) = \binom{n}{x} (p\text{Se} + (1 - p)(1 - \text{Sp}))^x \, (p(1 - \text{Se}) + (1 - p)\text{Sp})^{(n-x)}, \tag{8}$$

Of course, the Eq. (8) is followed by the same argument in (7) to calculate the probability that substantiates freedom from disease. That equation or the Eq. (7)

would help veterinary epidemiologist to determine the minimum sample size of laboratory testing.

The abovementioned model is kept very simple. However, the method heavily relies on laboratory testing performance and sampling effort. As an important remark about the sampling, considering that clustering is common for directly transmitted infectious diseases, it is hard to truly achieve a random sampling. Another technical issue is that the prevalence is assumed to be a constant, and thus, the stationarity is inherently assumed. For the similar reason, it is quite unfortunate that the error probability of elimination is only based on binomial sampling error (without accounting for stochastic dependence structure in empirical data of cases). Despite these problems, it is worth noting that the use of laboratory testing can overcome the problem of involving asymptomatic infections.

5 An Explicit Method for Multiple Exposure Setting

Epidemiological methods to determine the end of an epidemic in the presence of multiple exposures (and thus, involving stochastic dependence structure) are very scarce. This might be attributable to a difficulty in capturing the complex epidemic dynamics using simple equations in the presence of human-to-human transmissions.

An exceptionally careful pioneering study in this context was conducted by Eichner and Dietz [12] on poliomyelitis. Polio virus infection involves a substantial number of asymptomatic infections, and it is believed that only one paralytic case would occur among a total of 200 infections in naive host. Besides, because polio eradication program is underway due to effective vaccines and routine immunization programs, the so-called endgame of polio has called for a solid method to determine the local elimination of polio.

In principle, a stochastic compartmental model was employed for simulations, and Eichner and Dietz examined the probability that silent infections are underway as a function of time since the observation of last paralytic case [12]. Using the Markov jump process and simulating from the endemic equilibrium, the probability of silent infections as a function of the time since the last paralytic case, as shown in Fig. 3, was obtained. Examining realistic range of the frequency of paralytic case, ranging from one among 300 infections to 100 infections, Fig. 3 indicated that the probability of silent infections would be less than 1 % if 5 years is secured as the waiting time since the last paralytic case.

Fitting the stochastic model to empirically observed epidemiological data would be perhaps the most straightforward method to estimate the probability of extinction (and thus, the probability that the epidemic is still going on). Such model could also have a potential to be fitted to the dataset both with and without case finding efforts on the ground. Nevertheless, in practice, it is extremely difficult to fit such a stochastic model to a portion of epidemic data. That is, fitting to the latest data only would force us to focus on a chopped epidemic curve (with unknown infection-age structure) and the determination of the end of epidemic without fully realizing the
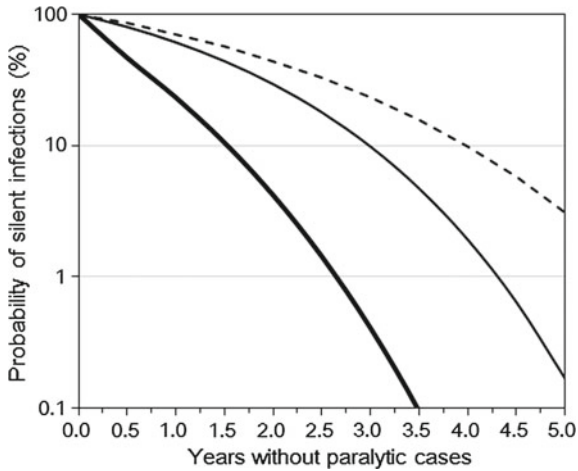
Fig. 3 *Probability of silent infection as a function of time since the last paralytic polio case.* Probability that silent infections still occur when no paralytic polio cases have been observed for a given period of time is shown [12]. The figure was reproduced by the author with reference to methods of Eichner and Dietz [12] for the scenario in which IPV (inactivated polio vaccine) was employed with the 80 % vaccination coverage. One case per 100 infections (*bold line*), one case per 200 infections (*solid line*), and one case per 300 infections (*dashed line*) were assumed

epidemiological dynamics might be too challenging. In fact, Fig. 3 is the result from simulations starting with a boundary condition and is not the time from the actual latest observation.

6   A Heuristic Method for Multiple Exposure Setting

The last approach to be reviewed is a heuristic approach in the presence of stochastic dependence structure with an application to the Middle East respiratory syndrome (MERS) in the Republic of Korea [1]. Not involving any additional cases of MERS for several weeks in the South Korea, the government and the WHO discussed an appropriate timing to declare the end of the outbreak. As discussed in the second section, a widely acknowledged criteria of the WHO to decide the end of an epidemic has been to ensure no further report of cases, setting twice the long incubation period (i.e. 14 days for MERS) as the standard waiting period since the latest date of diagnosis or recovery. Adopting 28 days as the waiting time and count days from 4 July, the date on which the latest case was diagnosed, the earliest date that Korean government could have declared the end of outbreak was 2 August adhering to the WHO criteria. If we count the days from the last PCR positive date, the date of declaration would even have been in late December 2015. Nevertheless, to emphasize the safety to the nation as well as forthcoming international travelers at

an earlier time, the Korean government made an original decision to announce the end of MERS outbreak on 27 July due to the fact that the last quarantined case was freed from movement restriction. To judge the appropriateness of these decisions, the probability of observing additional cases as a function of calendar time was explicitly calculated and such objective judgment was compared against that based on the WHO criteria.

The probability of observing additional cases was derived, using the serial interval, i.e. the time from illness onset in a primary case to illness onset in the secondary case, and the transmissibility of MERS. Let $F(t)$ be the cumulative distribution function of the serial interval. If time t is elapsed since the last case and provided that the last case were able to produce only one secondary case, the probability that at least one additional case is observed at time t since the illness onset of last case would be $1 - F(t)$. To address the potential of observing multiple secondary transmissions produced by a single primary case, we use the offspring distribution $p_y = \Pr(Y = y)$. Then, the risk of observing at least one additional case at time t since the illness onset of primary case is

$$\Pr(X \geq 1) = 1 - \sum_{y=0}^{\infty} p_y F(t)^y \tag{9}$$

Using the dataset of t_i, the calendar date of illness onset of diagnosed cases i ($i = 0, 1, \ldots, 185$), the probability of observing additional cases in future at calendar date t is calculated as

$$\Pr(X \geq 1) = 1 - \prod_{i=1}^{185} \sum_{y=0}^{\infty} p_y F(t - t_i)^y \tag{10}$$

It should be noted that the Eq. (10) does not manually subtract all existing secondary transmissions from the model, despite the fact that the observed cases have already generated a part of secondary cases that they have been supposed to cause. For that reason, the probability that is derived from the Eq. (10) may be slightly an overestimate.

As practiced in the determination of the length of quarantine [8, 10], one can declare the end of outbreak if that probability is smaller than 5 %, a threshold value. Our analysis showed that the first date on which the posterior median probability decreased to less than 5 % was 21 July (Fig. 4). The first date on which the posterior median lowered 1 % was 23 July. Namely, compared with 2 August as calculated from the WHO criteria, the declaration date of the end of outbreak could have been 11 and 9 days earlier, respectively.

The calculated probability is interpreted as the risk of observing at least one more case on or after a specified date and has a good potential to assist objective determination of the end of outbreak. The model efficiently addressed three practical problems in objectively calculating the probability that an outbreak leads to the end: (i) multiple cases on the same date, (ii) several recent cases with different illness
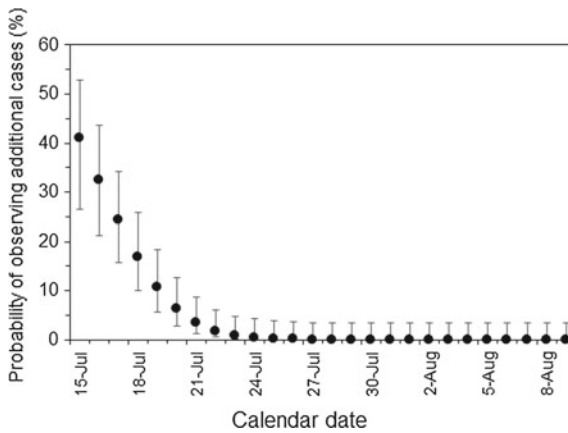
Fig. 4 *Estimated probability of observing additional cases of the Middle East respiratory syndrome coronavirus infection in the Republic of Korea, 2015*. Probability of observing additional cases on each calendar date, given no illness onset has been observed, is calculated [1]. *Circles* represent posterior median values that were calculated from resampled parameters governing the offspring distribution and serial interval. Whiskers extend to upper and lower 95 % credible intervals

onset dates, and (iii) variations in the number of secondary cases generated by a single primary case.

Of course, missing undiagnosed or mild cases is not taken into account in this method, and under-diagnosis would considerably extend the time to declare the end of outbreak (and thus, the proposed method is not directly applicable to EVD in West Africa to which we are presently developing an alternative method), all possible contact of diagnosed cases in the late phase of MERS outbreak in Korea were all traced, and thus, it was appropriate to ignore ascertainment bias in this specific setting. Important limitations include (i) the absence of dependence between serial interval and offspring distribution (as long as the two were estimated separately from independent datasets) and (ii) need to infer the offspring distribution precisely, perhaps requiring us to analyze contact tracing data or outbreak size distribution.

7   Conclusions

Epidemiological and laboratory methods of ascertaining the end of an epidemic were reviewed. To declare the end of an epidemic, it has been shown that multitude of methods might be used in combination with or without case finding efforts and biological samples for laboratory testing. To achieve this task, it is evident that surveillance and mathematical modeling are two complementary instruments in the toolbox of epidemiologists. Combining their strengths would be highly beneficial to

better define the end of an epidemic so that necessary public health actions can be taken properly.

Lastly, it is inevitable that the decision for declaring the end of an epidemic is highly politicized, and thus, the final decision must not solely be based on mathematical modeling results alone. Nevertheless, offering scientific evidence would make a big difference in epidemiological capacity and definitely ease the decision by policymakers. Ideally, there should be regular opportunity for modeling experts and policymakers to sit together to work on and discuss this matter.

References

1. Nishiura, H., Miyamatsu, Y., Mizumoto, K.: Objective determination of end of MERS outbreak, South Korea, 2015. Emerg. Infect. Dis. **22**, 146–148 (2016)
2. Arita, I., Wickett, J., Nakane, M.: Eradication of infectious diseases: its concept, then and now. Jpn. J. Infect. Dis. **57**, 1–6 (2004)
3. Greiner, M., Dekker, A.: On the surveillance for animal diseases in small herds. Prev. Vet. Med. **70**, 223–234 (2005)
4. Brookmeyer, R., You, X.: A hypothesis test for the end of a common source outbreak. Biometrics **62**, 61–65 (2006)
5. Ebola virus disease in West Africa: The first 9 months of the epidemic and forward projections. N. Engl. J. Med. **371**, 1481–1495 (2014)
6. World Health Organization: Definition of zero Ebola cases, World Health Organization: Geneva (2014). http://www.who.int/csr/disease/ebola/declaration-ebola-end/en/
7. Nishiura, H.: Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. Emerg. Themes. Epidemiol. **4**, 2 (2007)
8. Nishiura, H.: Determination of the appropriate quarantine period following smallpox exposure: an objective approach using the incubation period distribution. Int. J. Hyg. Environ. Health. **212**, 97–104 (2009)
9. Farewell, V.T., Herzberg, A.M., James, K.W., Ho, L.M., Leung, M.: SARS incubation and quarantine times: when is an exposed individual known to be disease free? Stat. Med. **24**, 3431–3445 (2005)
10. Nishiura, H., Wilson, N., Baker, M.G.: Quarantine for pandemic influenza control at the borders of small island nations. BMC Infect. Dis. **9**, 27 (2009)
11. Cameron, A.R., Baldock, F.C.: A new probability formula for surveys to substantiate freedom from disease. Prev. Vet. Med. **34**, 1–17 (1998)
12. Eichner, M., Dietz, K.: Eradication of poliomyelitis: when can one be sure that polio virus transmission has been terminated? Am. J. Epidemiol. **143**, 816–822 (1996)