The Need to Develop a Holistic Infectious Disease Surveillance System

Callum R.K. Arnold[1,2,*], Alex C. Kong[3], Amy K. Winter[4], William J. Moss[3,5], Bryan N. Patenaude[3], Matthew J. Ferrari[1,2]

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA 16802
[2] Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802
[3] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205
[4] Department of Epidemiology, College of Public Health, University of Georgia, Athens, GA, USA 30602
[5] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205

* Corresponding author. Callum R.K. Arnold. Address: Department of Biology, Pennsylvania State University, University Park, PA, USA 16802. Email: cfa5228@psu.edu.

Word count: 488

Abstract

Background

Infectious disease surveillance and outbreak detection systems often utilize diagnostic testing to validate case identification. The metrics of sensitivity, specificity, and positive predictive value are commonly discussed when evaluating the performance of diagnostic tests, and to a lesser degree, the performance of outbreak detection systems. However, the interaction of the two levels' (the test and the alert system) metrics, is typically overlooked. Here, we describe how equivalent regions of detection accuracy can exist over a range of diagnostic test characteristics, examining the sensitivity to background noise structure and magnitude.

Methods

We generated a stochastic SEIR model with importation to simulate true measles and non-measles sources of febrile rash (noise) daily incidence. We generated time series of febrile rash (i.e., measles clinical case definition) by summing the daily incidence of measles and either independent Poisson noise or non-measles dynamical noise (consistent with rubella virus). For each time series we assumed a fraction of all cases were seen at a healthcare clinic, and a subset of those were diagnostically confirmed using a test with sensitivity and specificity consistent with either a rapid diagnostic test (RDT) or an enzyme-linked immunosorbent assay (ELISA). From the resulting time series of test-positive cases, we define an outbreak alert as the exceedance of a threshold by the 7-day rolling average of observed (test positive) cases. For each threshold level, we calculated percentages of alerts that were aligned with an outbreak (analogous to the positive predictive value), the percentage of outbreaks detected (analogous to the sensitivity), and combined these two measures into an accuracy metric for outbreak detection. We selected the optimal threshold as the value that maximizes accuracy. We show how the optimal threshold and resulting accuracy depend on the diagnostic test, reporting rate, and the type and magnitude of the non-measles noise.

Results

The optimal threshold for each test increased monotonically as the percentage of clinic visits who were tested increased. With Poisson-only noise, similar outbreak detection accuracies could be achieved with imperfect RDT-like tests as with ELISA-like diagnostic tests (c. 93%), given moderately high testing rates. With larger delays (14 days) between the ELISA test administration and result date, RDTs could outperform the ELISA. Similar numbers of unavoidable cases and outbreak alert delays could be achieved between the test types. With dynamical noise, however, the accuracy of ELISA scenarios was far superior to those achieved with RDTs (c. 93% vs. 73%). For dynamical noise, RDT-based scenarios typically favored more sensitive alert threshold than ELISA-based scenarios (at a given reporting rate), observed with lower numbers of unavoidable cases and detection delays.

Conclusions

The performance of an outbreak detection system is highly sensitive to the structure and the magnitude of background noise. Under the assumption that the noise is relatively static over time, RDTs can perform as well as ELISA in a surveillance system. However, when the noise is temporally correlated, as from a separate SEIR process, imperfect tests cannot overcome their accuracy limitations through higher testing rates.

Key words: Rapid-Diagnostic Tests; ELISA; Infectious Disease Surveillance; Outbreak Detection.

Word count: 4310

Background

At the heart of an outbreak detection system is a surveillance program built upon case detection, often utilizing individual diagnostic tests as necessary components of epidemiological investigations before the declaration of an outbreak [1–5]. For diseases with non-specific symptoms, accurate measurement tools are often necessary to confidently and correctly ascribe changes in symptom prevalence within a population to a particular disease, and therefore detect outbreaks of specific pathogens. As a result, it has been commonplace for surveillance systems to be developed around high-accuracy tests, such as Polymerase Chain Reaction (PCR) tests and immunoglobulin (Ig) tests, when financially and logistically feasible [6–11]. Depending on the disease in question, either sensitivity (the ability to correctly detect a true positive individual) or specificity (the ability to correctly discount a true negative) will be prioritized, as they often are at odds with each other [12–14]. This balance is commonly defined within the Target Product Profile (TPP) of a test [15], which is a set of minimum characteristics that should be met for production and widespread use, helping to guide research and development. For example, in the wake of the 2013 Ebola outbreak in Guinea a TPP was developed that listed the minimum acceptable sensitivity of 95% and specificity of 99% [16]. Recognizing that Ebola is not the major cause of fever and other non-specific symptoms in the region, it is arguably more important to prioritize the specificity of the disease, although the authors note that the severity of the infection requires a high level of sensitivity as the consequences of a missed case are dire at an individual and population level [16].

Much like the accuracy of an individual test, outbreak detection systems face the same issue regarding the prioritization of sensitive or specific alerts [17–19]. For many disease systems, particularly in resource constrained environments where the burden of infectious diseases is highest [20,21], cases are counted and if a pre-determined threshold is breached, be that weekly, monthly, or some combination of the two, an alert is triggered that may launch a further investigation and/or a response [4,18]. In effect, this discretizes a distinctly continuous phenomenon (observed cases) into a binary measure, outbreak or no outbreak, for decision making purposes. Reactive interventions, such as vaccination campaigns and non-pharmaceutical based interventions, designed to reduce transmission or limit and suppress outbreaks, early action has the potential to avert the most cases [22–27]. While this framing would point towards a sensitive (i.e., early alert) surveillance system being optimal, each action comes with both direct and indirect financial and opportunity costs stemming from unnecessary activities that limit resources for future response capabilities. Just as the balance of sensitivity and specificity of a test for an individual must be carefully evaluated, so must the balance at the outbreak level.

The concept of using incidence-based alert triggers to define the discrete event of an "outbreak" with characteristics analogous to individual tests has been well documented in the case of meningitis, measles, and malaria [4,19,26,28–31]. However, an overlooked, yet critical, aspect of an outbreak detection system is the interplay between the individual test and outbreak alert characteristics. With their success within malaria surveillance systems, and particularly since the COVID-19 pandemic, rapid diagnostic tests (RDTs) have garnered wider acceptance, and their potential for use in other disease systems has been gaining interest [32]. Despite concerns of their lower diagnostic accuracy slowing their adoption until recently [33], RDTs generally have

reduced cold-chain requirements [34] and faster speed of result that has been show to outweigh the cost of false positive/negative results in some settings [32,35–37].

In this paper we examine how the use of imperfect diagnostic tests affects the performance of outbreak detection for measles outbreaks in the context of a febrile rash surveillance system that includes both measles and non-measles cases. Because measles symptoms are non-specific, it is important to account for non-measles sources of febrile rash e.g., rubella, parvovirus, varicella, etc., producing the potential for false positive results in the context of imperfect tests. Currently, measles outbreaks are declared on the basis of either suspected measles cases (i.e., an individual with fever and maculopapular rash [38]) alone, cases confirmed by enzyme-linked immunosorbent assay (ELISA) to detect the presence of measles-specific IgM antibodies, or a combination of the two, depending on the elimination status of the region, with countries nearer elimination increasing the use of ELISA diagnostic confirmation for suspected cases [38]. Each of these detection systems have their flaws. Although clinical case definition is very fast and requires limited resources, it is highly sensitive, and in the face of high "background noise" from non-measles sources of febrile rash, can lead to low positive predictive value (PPV), i.e., the probability that an alert accurately captures an outbreak [39]. And while ELISA confirmation is the standard diagnostic test for measles surveillance and has higher specificity [38], the training and facility requirements generally mean that samples must be transported from the point of care to a separate laboratory, which incurs both costs and delays *[ADD REFERENCE HERE]*. In resource-poor settings these delays may be days to weeks. A rapid diagnostic test offers the opportunity to provide a compromise between diagnostic accuracy and timeliness, and recent developments show encouraging signs in the field as well as in theory [32,34].

By examining the combination of alert threshold and individual test characteristic in a modeling study that explicitly incorporates dynamical background noise, we aim to illustrate the need to develop a TPP for the whole detection system, not just one component. To evaluate the alert system performance, we develop a set of outbreak definition criteria and surveillance metrics, drawing inspiration from acceptance sampling, ecological surveillance systems, and epidemiological surveillance systems guidelines and reviews [40–45]. Using these metrics we overcome issues encountered by early warning systems that rely on dynamical values such as $R_{\text{effective}}$ in defining outbreaks [46–50], for example, characterizing the end of an epidemic period is important in a time series where multiple outbreaks will occur.

Methods

Model Structure

We constructed a stochastic compartmental non-age structured Susceptible-Exposed-Infected-Recovered (SEIR) model of measles and simulated using a modified Tau-leaping algorithm with a time step of 1 day, utilizing binomial draws to ensure compartment sizes remained positive valued [51,52]. We assumed that transmission rate (β_t) is sinusoidal with a period of one year and 20% seasonal amplitude. R_0 was set to 16, with a latent period of 10 days and infectious period of 8 days. The population was initialized with 500,000 individuals with Ghana-like birth and vaccination rates, and the final results were scaled up to the approximate 2022 population size of Ghana (33 million) [53]. We assumed commuter-style imports at each time step to avoid extinction; the number of imports each day were drawn from a Poisson distribution with mean proportional to the size of the population and R_0 [54]. The full table of parameters can be found in Table 1. All simulations and analysis was completed in Julia version 1.10.5 [55], with all code stored at https://github.com/arnold-c/OutbreakDetection.

Parameters	Measles	Dynamical noise
R_0	16	5
Latent period (σ)	10 days	7 days
Infectious period (γ)	8 days	14 days
Seasonal amplitude	0.2	0.2
Birth/death rate (μ)	27 per 1000 per annum	27 per 1000 per annum
Vaccination rate at birth (ρ)	80%	(0.5-85)%
Importation rate	$\frac{1.06*\mu*R_0}{\sqrt{N}}$	
Population size (N)	500,000, scaled to 33M	
Initial proportion susceptible	0.05	
Initial proportion exposed	0.0	
Initial proportion infected	0.0	
Initial proportion recovered	0.95	

Table 1: Compartmental model parameters

To examine the sensitivity of the detection system to background noise, we layered the measles incidence time series with one of four noise time series structures: Poisson-only noise; or dynamical noise with rubella-like parameters that could be in- or out-of-phase, or with independent seasonality to the measles dynamics. For Poisson-only noise, the number of non-measles febrile rash cases each day were independent draws from a Poisson distribution with mean λ. For dynamical noise, we generated time series of cases from an SEIR model that matched the measles model in structure, but had $R_0 = 5$, mean latent period of 7 days, and mean infectious period of 14 days, and added some additional noise independently drawn from a Poisson distribution with mean equal to 15% of the average daily rubella incidence from the SEIR time series to account for non-rubella sources of febrile rash (Table 1). The seasonality for the dynamical noise was assumed to be in-phase with measles, anti-phase with measles (peak timing 6 months later), or non-seasonal. Only dynamical in-phase noise and Poisson-only noise are presented in the main text; the anti-phase and non-seasonal dynamical noise scenarios are presented in the supplement.

For each noise structure, we simulated five different magnitudes of noise (Λ). Λ was calculated as a multiple (c) of the average daily measles incidence ($\langle \Delta I_M \rangle$): $\Lambda = c \cdot \langle \Delta I_M \rangle$ where $c \in \{1, 2, 4, 6, 8\}$. For the Poisson-noise scenarios, independent draws from a Poisson distribution with mean $\lambda = \Lambda$ were simulated to produce the noise time series. For the dynamical noise scenarios, the vaccination rate at birth was set to 85.38%, 73.83%, 50.88%, 27.89%, and 4.92% to produce equivalent mean daily noise values (to within 2 decimal places): $\Lambda = \langle \Delta I_R \rangle = c \cdot \langle \Delta I_M \rangle$. 100 time series of 100 years were simulated for each scenario, before summarizing the distributions of outbreak detection methods.

Defining Outbreaks

It is common to use expert review to define outbreaks when examining empirical data, but this is not feasible in a modeling study where tens of thousands of years are being simulated. To account for this, many studies only simulate a single outbreak within a time series (repeating this short stochastic simulation multiple times to ensemble results), define an outbreak as a period where $R_{\text{effective}} > 1$, or use a threshold of > 2 standard deviations (s.d.) over the mean seasonal incidence observed in empirical data (or from a 'burn-in' period of the simulation) [43,46,48,50,56,57]. Each method has its uses, but to evaluate the performance of an outbreak detection system in an

endemic region where multiple sequential epidemics are expected it is important to clearly define the bounds of the outbreak, which can only be achieved by 2 s.d. > mean ($R_{effective}$ will be less than 1 after an outbreak's peak, but still within what can be reasonably defined as the outbreak's bounds). This, however, assumes strong seasonal forcing and regular periodicity of incidence to produce a smooth enough baseline, which is not present as countries near measles elimination status [58]. Here we define a true measles outbreak as a region of the time series that meets the following three criteria:

- The daily measles incidence must be greater than, or equal to, 5 cases
- The daily measles incidence must remain above 5 cases for greater than, or equal to, 30 consecutive days
- The total measles incidence must be great than, or equal to, 500 cases within the bounds of the outbreak

Each of these is a necessary, but not sufficient, condition to the definition of an outbreak; all must be met. For example, a region may produce 1000 total cases within the lower and upper bounds defined by the daily incidence being $\geq$ 5 cases, but if this period only lasts for 20 days it would not be considered an outbreak as it would not be possible to mount a response to. The incidence of non-measles febrile rash (i.e., noise) does not affect the outbreak status of a region but may affect the alert status triggered by the testing protocol.

Triggering Alerts

Need to mention that an alert is any time the threshold is exceeded, regardless of the number of consecutive days of exceedance. Clearly state that 1 day of exceedance is and 30 days of consecutive exceedance are both treated as a single alert i.e., the same.

The use of alert thresholds is common in locations burdened by measles, where access to high performance computing is limited making nowcasting-style approaches to outbreak detection impractical. For this reason, we define an "alert" as any single day where the 7-day moving average of the daily incidence is greater than, or equal to, a pre-specified alert threshold. To examine the interaction of test and surveillance systems we vary the alert threshold, between 1 and 15 cases per day, before computing the evaluation metrics and identifying the optimal threshold for each combination of individual test, testing rate, noise structure, and noise magnitude. We also examined an alert method that required either the daily incidence or the 7-day moving average of daily incidence to exceed the alert threshold (see supplemental results).

Each day, 60% of the measles and non-measles febrile rash cases visit the clinic for treatment, and a percentage of these clinic visits are tested, as all clinic visits are deemed to be suspected measles cases because they meet the clinical case definition. This percentage of clinic visits that are tested is varied between 10% and 60%, in 10% increments, for all combinations of diagnostic test (except clinical case definition) and alert threshold, defining the "testing scenario". Each testing scenario uses one of the following tests:

- An RDT equivalent with 85% sensitivity and specificity, and 0-day lag in result return. That is, 85% of true measles cases will be correctly labelled as positive, and 15% of non-measles febrile rash individuals that are tested will be incorrectly labelled as positive for measles. This acts as a lower bound of acceptability for a new measles RDT.
- An RDT equivalent with 90% sensitivity and specificity, and 0-day lag in result return.
- An ELISA equivalent test with 100% sensitivity and specificity with a 0-day test result delay.
- An ELISA equivalent test with 100% sensitivity and specificity with a 14-day test result delay.

Optimal Thresholds

For each of the previously defined tests, we calculate the number of test positive cases (that will include false positive and negative cases resulting from imperfect diagnostic tests), which is used to categorize the time series by alert status in conjunction with the alert threshold. The overlap with true outbreak status, or lack thereof, is used to classify the following metrics, and has been illustrated in Figure 1:

- The percentage of alerts that are 'correct'. This is analogous to the PPV of the alert system. Note that it is possible to trigger multiple alerts within a single outbreak, and each would be considered correct.
- The percentage of outbreaks that are detected. This is analogous to the sensitivity of the alert system.
- The detection delay. If an alert precedes the start of the outbreak, as long as it successfully "captures" the outbreak within its bounds, it is considered to be a 'correct' alert, resulting in a negative detection delay i.e., an early warning triggered by false positives.
- The number of unavoidable and avoidable outbreak cases and deaths. Avoidable cases are defined as those that occur within an outbreak after a correct alert is first triggered i.e., cases that could theoretically be prevented with a perfectly effective and timely response. Unavoidable cases are the inverse: those that occur before a correct alert, or those that occur in an undetected outbreak. Using a fitted age distribution of measles cases in Ghana [59] and the age-specific case fatality ratio for Ghana in 2022 [60], we scaled the values for each one-year age cohort, per annum. In practice, not all cases deemed avoidable are (due to imperfect and delays in responses), but to minimize the sensitivity of the results to the response implementation and operational constraints we are counting them as such.
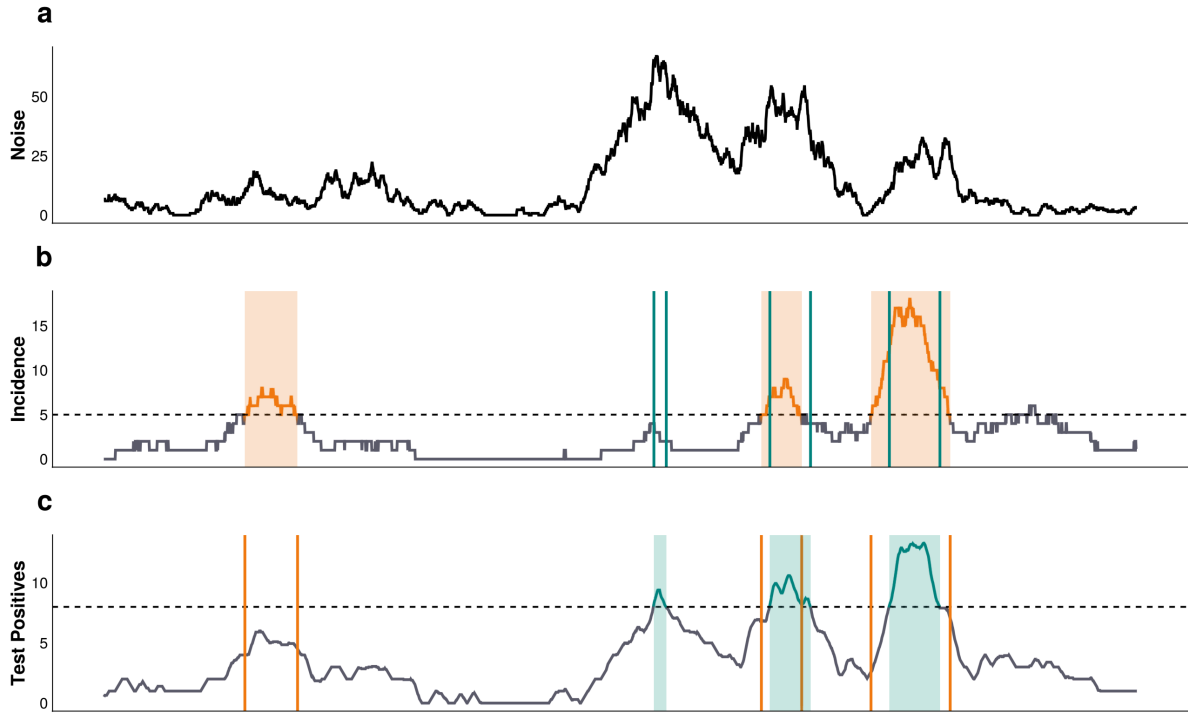
Figure 1: A schematic of the outbreak definition and alert detection system. A) Noise time series. B) Measles incidence time series. C) Observed time series resulting from testing noise & measles cases that visit the healthcare facility. The orange bands/vertical lines represent regions of the measles time series that meet the outbreak definition criteria. The green bands/vertical lines represent regions of the observed (measles - noise) time series that breach the alert threshold (the horizontal dashed line), and constitute an alert.

For each combination of diagnostic test and testing rate, the optimal alert threshold is calculated by selecting the threshold that produces the highest accuracy. Each of the 100 simulations produces an accuracy (the mean of the percentages of alerts that are correct and the percentage of the outbreaks that are detected), and the median accuracy of this distribution is used to determine the optimal alert threshold. Each testing scenario (the combination of the diagnostic test, testing rate, and alert threshold), is then compared (using the metrics above) at the optimal threshold that is specific for the scenario. This allows for conclusions to be made about the surveillance system as a whole, instead of just single components.

Results

The threshold that maximized outbreak detection accuracy given a testing scenario depends heavily on the noise structure and magnitude. For example, when the ratio of average noise incidence was 8 times higher than the average measles incidence, the optimal threshold for an RDT with 90% sensitivity/specificity and 40% testing of clinic visits was 4 cases per day with the dynamical in-phase noise, whereas it is 5 cases per day when Poisson-only noise was simulated (Table 2). This corresponds to an accuracy of 72% and 93%, respectively (Figure 2). Much higher accuracy can be achieved with imperfect tests under Poisson noise because large spikes of non-measles febrile rash that would lead to many false positives and trigger erroneous alerts do not occur. With dynamical noise, this possibility can and does occur and, when the ratio of non-measles to measles cases is sufficiently high, drastically impacts the accuracy of an alert system. The results from ELISA tests are identical between the different noise structures: with a perfect

test there are no false positive results. For dynamical noise, RDTs never perform as well as an ELISA, even under scenarios that include a 14-day lag between test procurement and result receipt. However, with Poisson noise, once testing reaches 20% of clinic visits, equivalence can be observed at approximately 91% accuracy (Figure 2).

	Test Characteristic		Testing Rate					
Noise Type	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise: in-phase	RDT Equivalent (85.0%)	0	1	2	3	4	4	5
Dynamical noise: in-phase	RDT Equivalent (90.0%)	0	1	2	3	4	4	5
Poisson noise	RDT Equivalent (85.0%)	0	1	3	4	5	6	7
Poisson noise	RDT Equivalent (90.0%)	0	1	2	4	5	5	6
Common	ELISA Equivalent	0	1	2	3	4	4	5
Common	ELISA Equivalent	14	1	2	3	4	4	5

Table 2: Optimal threshold for each testing scenario, when the average noise incidence is 8 times higher than the average measles incidence. A) the noise structure is dynamical, and the seasonality is in-phase with the measles incidence. B) the noise structure is Poisson only

Moving across a single row of Table 2, as the testing rate increases, so does the optimal threshold, because more individuals are testing positive. However, the changes in accuracy are not monotonically increasing, as one might anticipate. Even with a perfect test, there are sequential cells that increase the testing rate, yet accuracy decreases (Figure 2). For example, the ELISA equivalent with a 0-day test result lag outbreak detection accuracy decreases from 91% to 87% when clinic testing rates increase from 30% to 40% of clinic visits. The reason behind this unintuitive result stems from the use of integer valued thresholds. While computing the moving average test positive incidence helps to smooth out the alert system, an integer valued threshold can result in step-changes of accuracy between two threshold values, and the expected increase in the alert system's PPV from a higher threshold value is outweighed by the loss in the alert system's sensitivity to detecting outbreaks. Even with a perfect test, the alert system needs to discriminate between endemic/imported cases and epidemic cases; increasing the testing rate will result in higher numbers of test positive individuals, and a lower threshold can result in an overly sensitive alert system, triggering for measles infections, but not those within an outbreak. A higher threshold will face the opposite issue; not triggering for smaller outbreaks, which is more likely to be an issue at lower testing rates. Examining the number of unavoidable cases (Figure 3) and the detection delays (Figure 4), we can see that the decrease in accuracy results from a decrease in the sensitivity of the alert system: the detection delay increases by 6 days, and the unavoidable cases by c. 2300. Although the testing rate increases, so does the optimal threshold, from 3 test positives to 4 test positives per day. The subsequent change in testing rate (40 - 50%) is associated with no change in the optimal threshold for the ELISA, and as expected, the number of unavoidable cases and detection delay decrease, resulting in a more sensitive alert system, with an higher system accuracy indicating that the increase more than offset the decrease in the system's PPV.
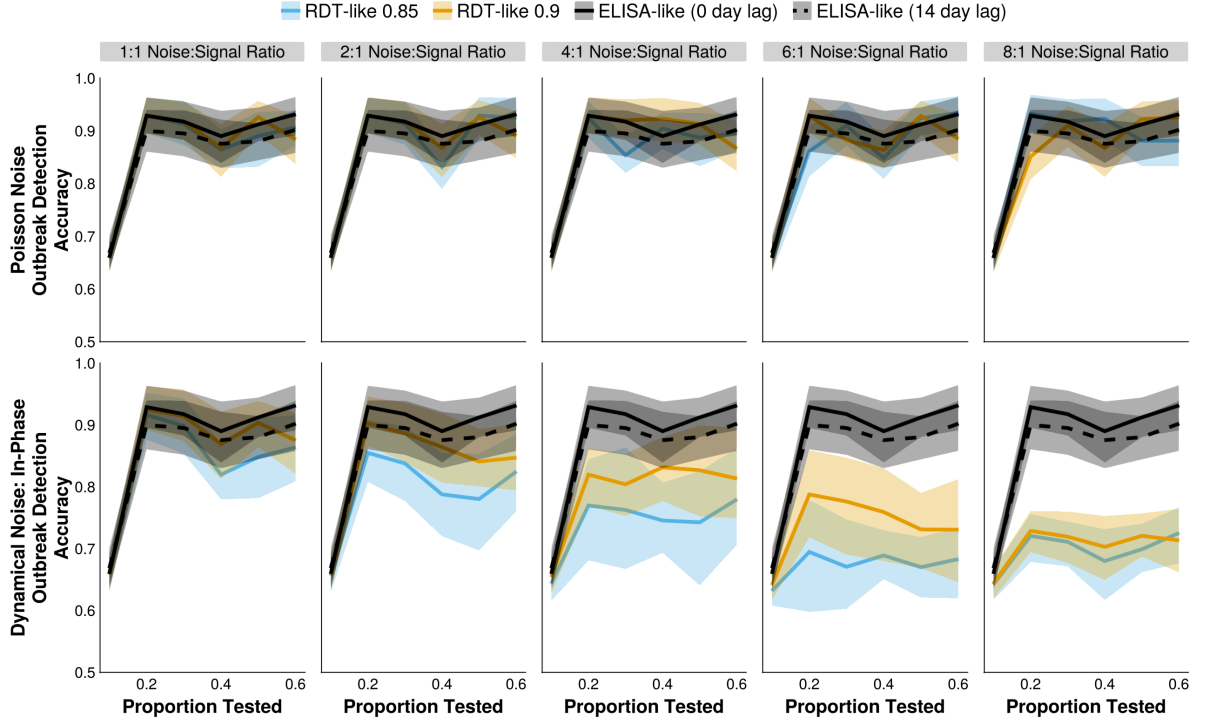
Figure 2: The accuracy of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate.
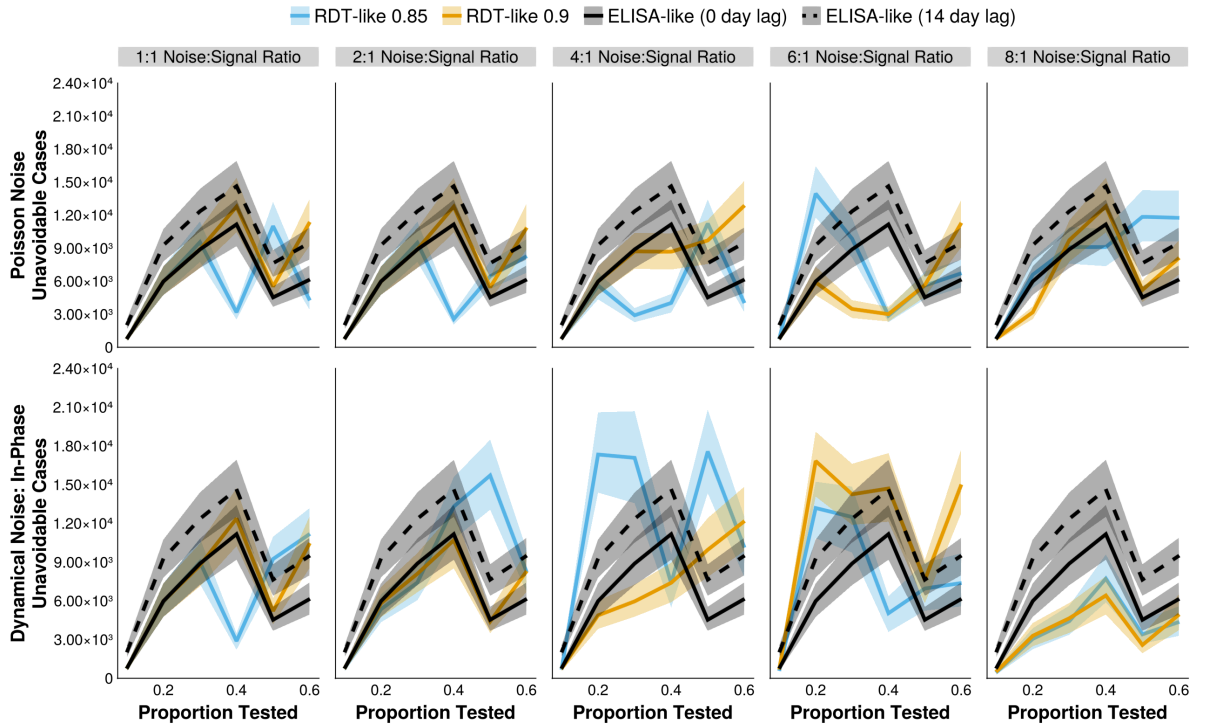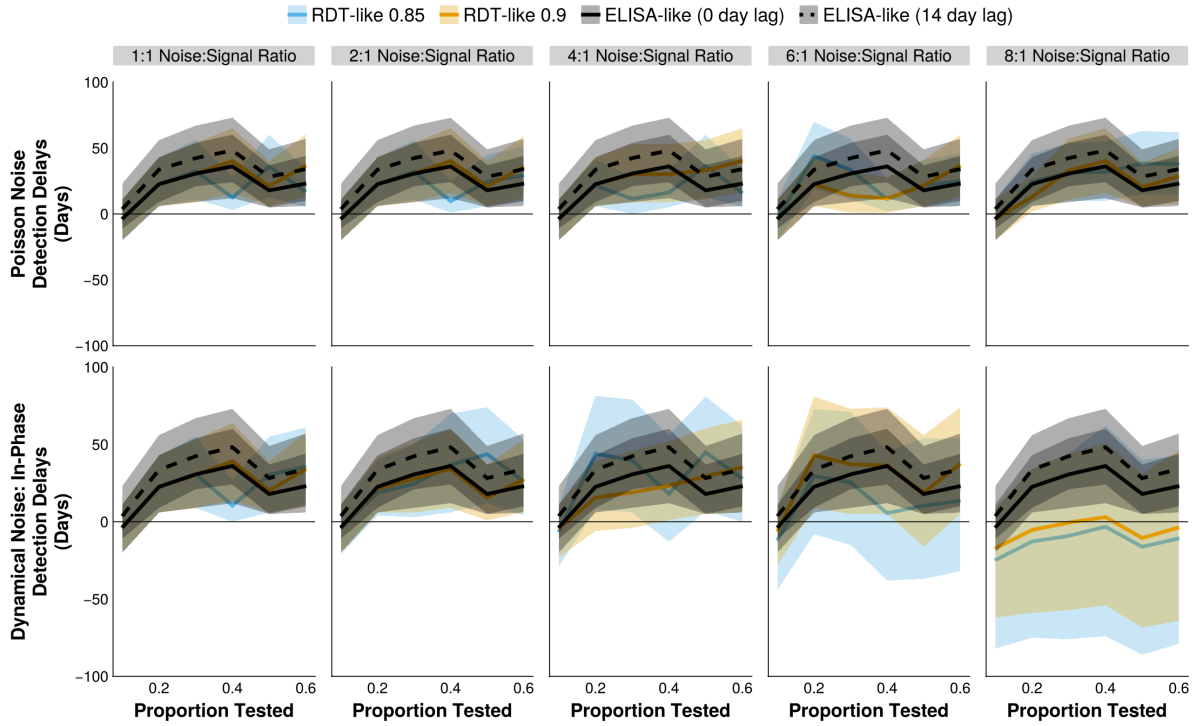

Figure 3: The number of unavoidable cases of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate.

Figure 4: The detection delay of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate.

For Poisson noise, similar detection delays are observed for all test and noise magnitudes, with variation by testing rate (mean of -3.7 to 36.1 days). Under dynamical noise, there are clearer differences in the performance of ELISA and RDTs, with the separation of outcomes occurring later than observed for detection accuracy (8 time noise magnitude vs. 2 times, respectively) (Figure 2). With large amounts of dynamical noise (8 times the measles incidence), the mean detection delay of the 90% and 85% RDTs range from -17.5 days to 3.2 days, and from -25.2 days to -3.4 days, respectively. Negative delays indicate that alerts are being triggered before the start of the outbreak and is correlated with the proportion of the time series that is under alert, with larger negative delays associated with more and/or longer alert periods (Figure 5, Supplemental Figure 2).
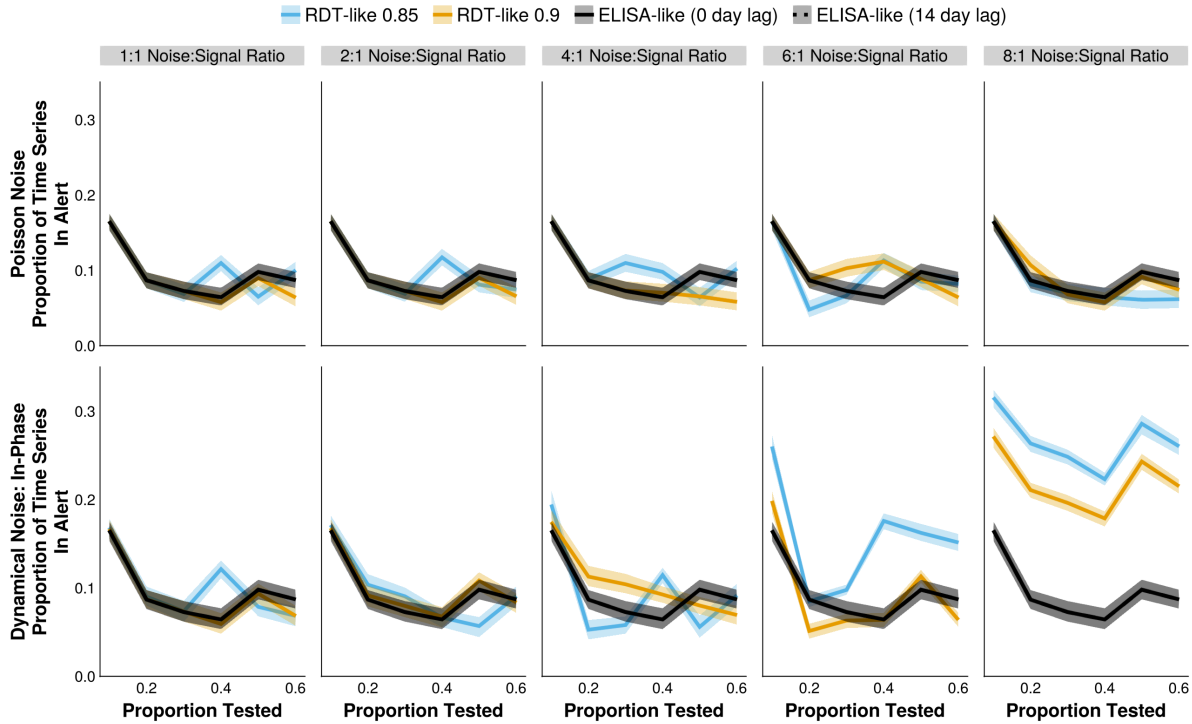
Figure 5: The difference between proportion of the time series in alert for outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate.

Discussion

The performance of an outbreak detection system is highly sensitive to the structure and level of background noise in the simulation. Despite the mean daily noise incidence set to equivalent values between the dynamical and Poisson-only simulations, drastically different results are observed.

Under the assumption that non-measles febrile rash is relatively static in time (Poisson), RDTs can perform as well, if not better than ELISA tests at moderate to high testing rates, and at a fraction of the cost. However, if it is expected that the noise is non-stationary, imperfect tests cannot overcome their accuracy limitations through higher testing rates, saturating at c. 74% accuracy, relative to ELISA's 93%. This occurs because, despite the same average incidence of noise in each (comparable) scenario, the relative proportion of measles to noise varies differently throughout the time series, exacerbating the effects of imperfect diagnostic tests that will produce higher rates of false positives and negatives than ELISA-like diagnostics. Imperfect tests are also more susceptible to incorrect results due to the varying prevalence throughout the time series, so the PPV of the system will not be static throughout the time series.

- *Make point about balance of false positive/negative outbreak detection*
- *Analysis isn't true optimization:*
  - *Requires explicit decisions about preference for speed / false alerts vs higher PPV*
  - *Surveillance is counting for action (WHO quote)*

The purpose of routine surveillance is to characterize the infection landscape. With strong public health infrastructure and infectious disease surveillance programs, it is possible to develop a strong understanding of the shape of febrile rash cases, regardless of source. With this

information, countries can tailor their future activities to rely more or less heavily upon RDTs, depending on the dynamics of the target disease and its relationship to background noise, favoring RDTs when there are low levels of noise and ELISAs during large rubella outbreaks, for example.

Limitations and Strengths

To our knowledge, this is one of the first simulation studies to examine the relationship between individual test characteristics and the wider surveillance program. By explicitly modeling the interaction between the two, we make a case that surveillance systems should take a holistic approach; ignoring one component can lead to drastically different, and suboptimal, results. Additionally, by defining outbreak bounds concretely we have been able to calculate metrics of outbreak detection performance that draw parallels to those used when evaluating individual diagnostic tests, allowing for intuitive and simple implementation of this method in resource-constrained environments, something that is not possible with most outbreak detection and early warning system simulations in the literature. An evaluation of all outbreak detection algorithms is beyond the scope of this work, but a more computationally expensive approach based on nowcasting incidence may help overcome the shortcomings of RDTs in high-noise scenarios.

For computational simplicity, this paper did not include demography in the model structure. And while a simulation-based approach allows for complete determination of true infection status i.e., measles vs non-measles febrile rash cases, and therefore an accurate accounting of the outbreak and alert bounds, these simulations do not specifically represent any real-world setting. The evaluation of empirical data does provide this opportunity, but at the cost of knowing the true infection status of individuals and confounding of multiple variables, limiting analysis to only those who are observed (i.e., not those in the community who do not visit a healthcare center), and removing the possibility to explore the sensitivity of the results to parameters of interest to a surveillance program e.g., testing rate, and the test itself.

Additionally, is has been well documented that the performance of an individual test is highly sensitive to its timing within a person's infection cycle [6,36,37,61,62], so it is possible that different conclusions would be drawn if temporal information about the test acquisition was included in the simulation.

Finally, the optimal threshold for a testing scenario depends heavily on the costs ascribed to incorrect actions, be that failing to detect an outbreak or incorrectly mounting a response for an outbreak that doesn't exist. In the simulations we have weighted them equally, but it is likely that they should not be deemed equivalent: missing an outbreak leads to many thousands of cases and associated DALYs, whereas an unnecessary alert would generally launch an initial low-cost investigation for full determination of the outbreak status. This is particularly important in countries with vast heterogeneity in transmission: different weightings should be applied to higher vs. lower priority/risk regions to account for discrepancies in consequences of incorrect decisions. For example, a high priority zone could still benefit from a false alert as many high-risk healthcare regions within a country are targeted for Supplemental Immunization Activities, so a false alert would just hasten the (proactive) vaccination response.

Given these limitations, the explicit values (i.e., optimal thresholds, accuracies etc.) should be interpreted with caution, and the exact results observed in the real-world will likely be highly dependent on unseen factors, such as the proportion of measles and non-measles sources of febrile rash that seek healthcare. However, the general patterns should hold, and more importantly, the analysis framework provides a consistent and holistic approach to evaluating the trade-off between individual level tests and the alert system enacted to detect outbreaks.

Funding

- *Something about GAVI/Gates*

Acknowledgements

References

1. Murray J, Cohen A L. Infectious Disease Surveillance. International Encyclopedia of Public Health [Internet]. 2017 [cited 2024Feb7];:222–9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7149515/

2. Zhou X-N, Bergquist R, Tanner M. Elimination of Tropical Disease through Surveillance and Response. Infectious Diseases of Poverty [Internet]. 2013 [cited 2024Feb7];2:1. Available from: https://doi.org/10.1186/2049-9957-2-1

3. PAHO. An Integrated Approach to Communicable Disease Surveillance [Internet]. 2000 Mar. Available from: https://www3.paho.org/english/sha/eb_v21n1-vigil.htm

4. World Health Organization. Measles Outbreak Guide [Internet]. Geneva, Switzerland: World Health Organization; 2022 [cited 2024Feb6]. Available from: https://iris.who.int/bitstream/handle/10665/360891/9789240052079-eng.pdf?sequence=1

5. Cragg L. Outbreak Response. Cragg L, Nutland W, Rudge J, editors. Applied Communicable Disease Control. London: Open University Press; 2018. pp. 134–51.

6. Gastanaduy P A, Redd S B, Clemmons N S, Lee A D, Hickman C J, Rota P A, et al. Measles [Internet]. Manual for the Surveillance of Vaccine-Preventable Diseases. Center for Disease Control and Prevention; 2019 [cited 2024Feb7]. Available from: https://www.cdc.gov/vaccines/pubs/surv-manual/chpt07-measles.html

7. Commissioner O of the. Coronavirus (COVID-19) Update: FDA Informs Public About Possible Accuracy Concerns with Abbott ID NOW Point-of-Care Test [Internet]. FDA; 2020 [cited 2024Feb7]. Available from: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-informs-public-about-possible-accuracy-concerns-abbott-id-now-point

8. Grassly N C, Pons-Salort M, Parker E P K, White P J, Ferguson N M, Imperial College COVID-19 Response Team. Comparison of Molecular Testing Strategies for COVID-19 Control: A Mathematical Modelling Study. Lancet Infect Dis. 2020;20:1381–9.

9. Ezhilan M, Suresh I, Nesakumar N. SARS-CoV, MERS-CoV and SARS-CoV-2: A Diagnostic Challenge. Measurement (Lond) [Internet]. 2021 [cited 2024Feb7];168:108335. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7833337/

10. World Health Organization. Cholera [Internet]. 2023 [cited 2024Feb7]. Available from: https://www.who.int/news-room/fact-sheets/detail/cholera

11. Essential Programme on Immunization (EPI), Immunization, Vaccines and Biologicals (IVB). Clinical Specimens for the Laboratory Confirmation and Molecular Epidemiology of Measles, Rubella, and CRS [Internet]. 3rd ed. Mulders M, editor. Manual for the Laboratory-based Surveillance of Measles, Rubella, and Congenital Rubella Syndrome. 2018 [cited 2024Feb7]. Available from: https://www.who.int/docs/default-source/immunization/vpd_surveillance/lab_networks/measles_rubella/manual/chapter-3.pdf

12. Westreich D. Diagnostic Testing, Screening, and Surveillance [Internet]. Westreich D, editor. Epidemiology by Design: A Causal Approach to the Health Sciences. Oxford University Press; 2019 [cited 2024Feb7]. p. 0. Available from: https://doi.org/10.1093/oso/9780190665760.003.0005

13. Shreffler J, Huecker M R. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios [Internet]. StatPearls. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024Feb8]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK557491/

14. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and Using Sensitivity, Specificity and Predictive Values. Indian J Ophthalmol [Internet]. 2008 [cited 2024Feb8];56:45–50. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636062/

15. World Health Organization. Target Product Profiles [Internet]. [cited 2024Feb6]. Available from: https://www.who.int/observatories/global-observatory-on-health-research-and-development/analyses-and-syntheses/target-product-profile/who-target-product-profiles

16. Chua A C, Cunningham J, Moussy F, Perkins M D, Formenty P. The Case for Improved Diagnostic Tools to Control Ebola Virus Disease in West Africa and How to Get There. PLOS Neglected Tropical Diseases [Internet]. 2015 [cited 2024Feb6];9:e3734. Available from: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003734

17. German R R. Sensitivity and Predictive Value Positive Measurements for Public Health Surveillance Systems. Epidemiology [Internet]. 2000 [cited 2024Feb8];11:720–7. Available from: https://www.jstor.org/stable/3703831

18. World Health Organization. Operational Thresholds [Internet]. Meningitis Outbreak Response in Sub-Saharan Africa: WHO Guideline. World Health Organization; 2014 [cited 2024Feb6]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK274178/

19. Lewis R, Nathan N, Diarra L, Belanger F, Paquet C. Timely Detection of Meningococcal Meningitis Epidemics in Africa. The Lancet [Internet]. 2001 [cited 2024Feb6];358:287–93. Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140673601054848/fulltext

20. GBD 2019 Child and Adolescent Communicable Disease Collaborators. The Unfinished Agenda of Communicable Diseases among Children and Adolescents before the COVID-19 Pandemic, 1990–2019: A Systematic Analysis of the Global Burden of Disease Study 2019. The Lancet [Internet]. 2023 [cited 2024Feb6];402:313–35. Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(23)00860-7/fulltext

21. Roser M, Ritchie H, Spooner F. Burden of Disease. Our World in Data [Internet]. 2023 [cited 2024Feb6];. Available from: https://ourworldindata.org/burden-of-disease

22. Atkins B D, Jewell C P, Runge M C, Ferrari M J, Shea K, Probert W J M, et al. Anticipating Future Learning Affects Current Control Decisions: A Comparison between Passive and Active Adaptive Management in an Epidemiological Setting. Journal of Theoretical Biology [Internet]. 2020 [cited 2021Oct17];506:110380. Available from: https://www.sciencedirect.com/science/article/pii/S0022519320302356

23. Tao Y, Shea K, Ferrari M. Logistical Constraints Lead to an Intermediate Optimum in Outbreak Response Vaccination. :20.

24. Grais R, Conlan A, Ferrari M, Djibo A, Le Menach A, Bjørnstad O, et al. Time Is of the Essence: Exploring a Measles Outbreak Response Vaccination in Niamey, Niger. J R Soc Interface [Internet]. 2008 [cited 2023Jun27];5:67–74. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2605500/

25. Ferrari M J, Fermon F, Nackers F, Llosa A, Magone C, Grais R F. Time Is (Still) of the Essence: Quantifying the Impact of Emergency Meningitis Vaccination Response in Katsina State, Nigeria.

International Health [Internet]. 2014 [cited 2024Feb6];6:282–90. Available from: https://doi.org/10.1093/inthealth/ihu062

26. World Health Organization. Confirming, Investigating and Managing an Outbreak [Internet]. Response to Measles Outbreaks in Measles Mortality Reduction Settings: Immunization, Vaccines and Biologicals. World Health Organization; 2009 [cited 2024Feb6]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK143959/

27. Minetti A, Kagoli M, Katsulukuta A, Huerga H, Featherstone A, Chiotcha H, et al. Lessons and Challenges for Measles Control from Unexpected Large Outbreak, Malawi. Emerg Infect Dis [Internet]. 2013 [cited 2024Feb6];19:202–9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3559033/

28. Trotter C L, Cibrelus L, Fernandez K, Lingani C, Ronveaux O, Stuart J M. Response Thresholds for Epidemic Meningitis in Sub-Saharan Africa Following the Introduction of MenAfriVac®. Vaccine [Internet]. 2015 [cited 2024Feb6];33:6212–7. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X15014097

29. Cooper L V, Stuart J M, Okot C, Asiedu-Bekoe F, Afreh O K, Fernandez K, et al. Reactive Vaccination as a Control Strategy for Pneumococcal Meningitis Outbreaks in the African Meningitis Belt: Analysis of Outbreak Data from Ghana. Vaccine. 2019;37:5657–63.

30. Zalwango M G, Zalwango J F, Kadobera D, Bulage L, Nanziri C, Migisha R, et al. Evaluation of Malaria Outbreak Detection Methods, Uganda, 2022. Malar J [Internet]. 2024 [cited 2024Feb6];23:18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10787982/

31. Kaninda A-V, Belanger F, Lewis R, Batchassi E, Aplogan A, Yakoua Y, et al. Effectiveness of Incidence Thresholds for Detection and Control of Meningococcal Meningitis Epidemics in Northern Togo. International Journal of Epidemiology [Internet]. 2000 [cited 2024Feb6];29:933–40. Available from: https://doi.org/10.1093/ije/29.5.933

32. Warrener L, Andrews N, Koroma H, Alessandrini I, Haque M, Garcia C C, et al. Evaluation of a Rapid Diagnostic Test for Measles IgM Detection; Accuracy and the Reliability of Visual Reading Using Sera from the Measles Surveillance Programme in Brazil, 2015. Epidemiology & Infection [Internet]. 2023 [cited 2024Jan14];151:e151. Available from: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/evaluation-of-a-rapid-diagnostic-test-for-measles-igm-detection-accuracy-and-the-reliability-of-visual-reading-using-sera-from-the-measles-surveillance-programme-in-brazil-2015/70D966F864717174092310EBA37929DD

33. Miller E, Sikes H D. Addressing Barriers to the Development and Adoption of Rapid Diagnostic Tests in Global Health. Nanobiomedicine (Rij). 2015;2:6.

34. Brown D W, Warrener L, Scobie H M, Donadel M, Waku-Kouomou D, Mulders M N, et al. Rapid Diagnostic Tests to Address Challenges for Global Measles Surveillance. Current Opinion in Virology [Internet]. 2020 [cited 2024Feb6];41:77–84. Available from: https://www.sciencedirect.com/science/article/pii/S1879625720300286

35. McMorrow M L, Aidoo M, Kachur S P. Malaria Rapid Diagnostic Tests in Elimination Settings—Can They Find the Last Parasite?. Clin Microbiol Infect [Internet]. 2011 [cited 2024Feb6];17:1624–31. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4821879/

36. Larremore D B, Wilder B, Lester E, Shehata S, Burke J M, Hay J A, et al. Test Sensitivity Is Secondary to Frequency and Turnaround Time for COVID-19 Screening. Science Advances

[Internet]. 2021 [cited 2022Mar29];7:eabd5393. Available from: https://www.science.org/doi/10.1126/sciadv.abd5393

37. Middleton C, Larremore D B. Modeling the Transmission Mitigation Impact of Testing for Infectious Diseases. 2023.

38. Essential Programme on Immunization (EPI), Immunization, Vaccines and Biologicals (IVB). Measles [Internet]. Vaccine Preventable Diseases Surveillance Standards. Geneva, Switzerland: World Health Organization; 2018 [cited 2024Feb6]. p. 30. Available from: https://cdn.who.int/media/docs/default-source/immunization/vpd_surveillance/vpd-surveillance-standards-publication/who-surveillancevaccinepreventable-11-measles-r2.pdf?sfvrsn=6d8879f9_10&download=true

39. Hutchins S S, Amler R, Maes E F, Grabowsky M, Bromberg K, Glasglow V, et al. Evaluation of the Measles Clinical Case Definition. The Journal of Infectious Diseases [Internet]. 2004 [cited 2024Feb6];189:S153–9. Available from: https://doi.org/10.1086/379652

40. Yemshanov D, Haight R G, Liu N, Chen C, MacQuarrie C J K, Ryall K, et al. Acceptance Sampling for Cost-Effective Surveillance of Emerald Ash Borer in Urban Environments. Forestry: An International Journal of Forest Research [Internet]. 2020 [cited 2021Dec1];93:280–96. Available from: https://doi.org/10.1093/forestry/cpz028

41. Christensen J, Gardner I A. Herd-Level Interpretation of Test Results for Epidemiologic Studies of Animal Diseases. Preventive Veterinary Medicine [Internet]. 2000 [cited 2021Dec2];45:83–106. Available from: https://www.sciencedirect.com/science/article/pii/S0167587700001185

42. Murato Y, Hayama Y, Shimizu Y, Sawai K, Yamamoto T. Evaluation of Sampling Methods for Effective Detection of Infected Pig Farms during a Disease Outbreak. PLOS ONE [Internet]. 2020 [cited 2021Dec1];15:e241177. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241177

43. Stern L, Lightfoot D. Automated Outbreak Detection: A Quantitative Retrospective Analysis. Epidemiol Infect [Internet]. 1999 [cited 2024Feb6];122:103–10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809594/

44. Calba C, Goutard F L, Hoinville L, Hendrikx P, Lindberg A, Saegerman C, et al. Surveillance Systems Evaluation: A Systematic Review of the Existing Approaches. BMC Public Health [Internet]. 2015 [cited 2024Feb6];15:448. Available from: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1791-5

45. Guidelines Working Group. Updated Guidelines for Evaluating Public Health Surveillance Systems [Internet]. Atlanta, Georgia; 2001 Jul pp. 1–35. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm

46. Jombart T, Ghozzi S, Schumacher D, Taylor T J, Leclerc Q J, Jit M, et al. Real-Time Monitoring of COVID-19 Dynamics Using Automated Trend Fitting and Anomaly Detection. Philosophical Transactions of the Royal Society B: Biological Sciences [Internet]. 2021 [cited 2021Oct17];376:20200266. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0266

47. Proverbio D, Kemp F, Magni S, Gonçalves J. Performance of Early Warning Signals for Disease Re-Emergence: A Case Study on COVID-19 Data. PLOS Computational Biology [Internet]. 2022 [cited 2023Oct29];18:e1009958. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009958

48. Stolerman L M, Clemente L, Poirier C, Parag K V, Majumder A, Masyn S, et al. Using Digital Traces to Build Prospective and Real-Time County-Level Early Warning Systems to Anticipate COVID-19 Outbreaks in the United States. Science Advances [Internet]. 2023 [cited 2023Oct29];9:eabq199. Available from: https://www.science.org/doi/10.1126/sciadv.abq0199

49. Brett T S, O'Dea E B, Marty É, Miller P B, Park A W, Drake J M, et al. Anticipating Epidemic Transitions with Imperfect Data. PLoS Comput Biol [Internet]. 2018 [cited 2023Sep29];14:e1006204. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 6010299/

50. Salmon M, Schumacher D, Höhle M. Monitoring Count Time Series in \emph{R} : Aberration Detection in Public Health Surveillance. J Stat Soft [Internet]. 2016 [cited 2023Sep24];70. Available from: http://www.jstatsoft.org/v70/i10/

51. Chatterjee A, Vlachos D G, Katsoulakis M A. Binomial Distribution Based τ-Leap Accelerated Stochastic Simulation. The Journal of Chemical Physics [Internet]. 2005 [cited 2023Sep29];122:24112. Available from: https://pubs.aip.org/jcp/article/122/2/024112/307261/ Binomial-distribution-based-leap-accelerated

52. Gillespie D T. Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. The Journal of Chemical Physics [Internet]. 2001 [cited 2023Sep29];115:1716–33. Available from: https://pubs.aip.org/jcp/article/115/4/1716/451187/Approximate-accelerated-stochastic-simulation-of

53. World Bank. Ghana [Internet]. [cited 2024Feb6]. Available from: https://data.worldbank.org/

54. Keeling M J, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton: Princeton University Press; 2008.

55. Bezanson J, Edelman A, Karpinski S, Shah V B. Julia: A Fresh Approach to Numerical Computing. SIAM Rev [Internet]. 2017 [cited 2023Aug24];59:65–98. Available from: https:// epubs.siam.org/doi/10.1137/141000671

56. Teklehaimanot H D, Schwartz J, Teklehaimanot A, Lipsitch M. Alert Threshold Algorithms and Malaria Epidemic Detection. Emerg Infect Dis [Internet]. 2004 [cited 2024Feb6];10:1220–6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3323320/

57. Leclère B, Buckeridge D L, Boëlle P-Y, Astagneau P, Lepelletier D. Automated Detection of Hospital Outbreaks: A Systematic Review of Methods. PLOS ONE [Internet]. 2017 [cited 2024Feb6];12:e176438. Available from: https://journals.plos.org/plosone/article?id=10.1371/ journal.pone.0176438

58. Graham M, Winter A K, Ferrari M, Grenfell B, Moss W J, Azman A S, et al. Measles and the Canonical Path to Elimination. Science [Internet]. 2019 [cited 2023Nov20];364:584–7. Available from: https://www.science.org/doi/10.1126/science.aau6299

59. Minta A A. Progress Toward Measles Elimination — Worldwide, 2000–2022. MMWR Morb Mortal Wkly Rep [Internet]. 2023 [cited 2024Jan9];72. Available from: https://www.cdc.gov/ mmwr/volumes/72/wr/mm7246a3.htm

60. Sbarra A N, Mosser J F, Jit M, Ferrari M, Ramshaw R E, O'Connor P, et al. Estimating National-Level Measles Case–Fatality Ratios in Low-Income and Middle-Income Countries: An Updated Systematic Review and Modelling Study. The Lancet Global Health [Internet]. 2023 [cited 2024Feb6];11:e516–24. Available from: https://www.thelancet.com/journals/langlo/article/ PIIS2214-109X(23)00043-8/fulltext

61. Kissler S M, Fauver J R, Mack C, Olesen S W, Tai C, Shiue K Y, et al. Viral Dynamics of Acute SARS-CoV-2 Infection and Applications to Diagnostic and Public Health Strategies. PLOS Biology [Internet]. 2021 [cited 2024Feb7];19:e3001333. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001333

62. Ratnam S, Tipples G, Head C, Fauvel M, Fearon M, Ward B J. Performance of Indirect Immunoglobulin M (IgM) Serology Tests and IgM Capture Assays for Laboratory Diagnosis of Measles. J Clin Microbiol [Internet]. 2000 [cited 2024Feb7];38:99–104. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC86030/