Identifying exceptional malaria occurrences in the absence of historical data in South Sudan: a method validation

G. Benedetti,1 R. A. White,2 H. Akello Pasquale,3 J. Stassijns,4 W. van den Boogaard,1 P. Owiti,5 R. Van den Bergh1

Background: Detecting unusual malaria events that may require an operational intervention is challenging, especially in endemic contexts with continuous transmission such as South Sudan. Médecins Sans Frontières (MSF) utilises the classic average plus standard deviation (AV+SD) method for malaria surveillance. This and other available approaches, however, rely on antecedent data, which are often missing.

Objective: To investigate whether a method using linear regression (LR) over only 8 weeks of retrospective data could be an alternative to AV+SD.

Design: In the absence of complete historical malaria data from South Sudan, data from weekly influenza reports from 19 Norwegian counties (2006–2015) were used as a testing data set to compare the performance of the LR and the AV+SD methods. The moving epidemic method was used as the gold standard. Subsequently, the LR method was applied in a case study on malaria occurrence in MSF facilities in South Sudan (2010–2016) to identify malaria events that required a MSF response.

Results: For the Norwegian influenza data, LR and AV+SD methods did not perform differently (P > 0.05). For the South Sudanese malaria data, the LR method identified historical periods when an operational response was mounted.

Conclusion: The LR method seems a plausible alternative to the AV+SD method in situations where retrospective data are missing.

Malaria remains a worldwide concern, with 90% of cases and 91% of related deaths in 2016 occurring in sub-Saharan Africa. From an epidemiological perspective, the presence of malaria is classified as follows: 1) non-endemic, where true epidemics can occur; 2) hypo- and meso-endemic, when there is low-to-moderate transmission that can be exacerbated by seasonality; and 3) hyper- and holo-endemic, with constant high transmission and possible exceptional periods.

Although not considered epidemics, seasonal variations and ‘exceptional’ years are observed in endemic areas. However, a standardised definition of ‘exceptionality’ does not exist, making the interpretation of malaria trends somewhat arbitrary. Nevertheless, such occurrences may require interventions to mitigate their impact, and early identification of these situations is crucial for operational actors in order to design and mount an intervention.

Classic measures of malaria occurrence
Context-adapted thresholds can define malaria outbreaks. This is straightforward in areas prone to epidemics and with well-functioning surveillance systems. However, no consensus exists on the applicability of different methods to define thresholds. Classically, historical malaria cases are used to identify thresholds. These can include: 1) the weekly average of cases plus Standard Deviations (AV+SD); 2) the third quartile over the weekly median of cases; or 3) the cumulative sum, i.e., the average of cumulative sums of cases over a period of 3 months. These methods rely on extensive retrospective information and are challenged when the data are missing, fragmented or non-comparable.

Other methods consider malaria incidence, the slope of malaria incidence or the slide positivity rate. The history of fever or parasite prevalence can also be investigated using surveys. Population figures or operational resources, however, often remain inadequate for such approaches.

Médecins Sans Frontières malaria surveillance
Médecins Sans Frontières (MSF) operates in more than 70 countries, where all epidemiological scenarios for malaria are encountered. National surveillance data are often unavailable or biased by fluctuating completeness, and the data collected at health facility level remain the source for disease monitoring and for launching an operational response. MSF recommends using both quantitative and qualitative information to identify ‘exceptional’ malaria occurrences. Quantitative measurements (e.g., number of cases, case severity, diagnostic test positivity rate) should be considered, together with qualitative analysis at service and community levels (e.g., drug efficacy, vector transmission, population movements, occurrence of other epidemics). Retrospective data are used to determine if an observed pattern is ‘normal’ or ‘exceptional’. Thresholds are calculated using the AV+SD method, ideally over the 5 previous years. However, missing, fragmented and non-comparable data sets frequently impede the application of this method.

Circumventing common shortcomings
In malaria-endemic South Sudan, historical malaria data are often missing, fragmented or non-comparable. An empirical method based on linear regression (LR) of 8 weeks of notification data was proposed to remedy the inapplicability of the AV+SD calculation. As the performance of the new LR method had to be
tested against the AV+SD method, and no complete and coherent historical data at MSF facility level were available, we sought an alternative data set for testing. A data set of notified influenza cases from primary health care services in Norway represented an opportunity of convenience. Influenza and malaria are both widespread diseases occurring with seasonal patterns. While many epidemiological similarities and differences exist between influenza and malaria, these were considered irrelevant to this study, as influenza cases from Norway exclusively represented the verification data set for the LR method.

The objectives of the present study were 1) to compare the performance of two mathematical (the LR and AV+SD) methods on a complete series of weekly notified influenza cases in Norway; and 2) to observe the performance of the LR method in a case study on malaria occurrence from MSF facilities in South Sudan.

MATERIAL AND METHODS

Study design
This was an observational study utilising routinely collected, retrospective influenza and malaria surveillance data from Norway and South Sudan.

Settings
Norway is a malaria-free, northern European country with a well-functioning system for disease surveillance. Febrile states of influenza are notified from all public primary health care sources through the Kontroll og Utbetaling av Helserefusejoner (KUHR) register (control and payment of reimbursements) using the International Classification of Primary Care, second edition (ICPC–2) health codes. The KUHR is a register of all reimbursement records sent by public primary care physicians in Norway.10,11

South Sudan is an African country that has been torn apart by decades of civil war before and after independence from Sudan in 2011. Conflicts and a chronic lack of resources hamper the possibilities for development, including the national health system. Malaria is endemic, with seasonal variations in transmission and different levels of endemicity, from hypo- to holo-endemic.2,12 The National Malaria Control Programme implements the National Malaria Strategic Plan, with the guidance of the World Health Organization.13

Study populations
Norway
The study population consisted of the weekly number of primary care consultations for influenza (code R80 on ICPC–2) from Week 1 2006 to Week 40 2015 for 19 counties in Norway. The data were provided by the Norwegian Health Directorate, which has ownership of the KUHR register.14

South Sudan
The study population consisted of the weekly number of malaria cases (confirmed by rapid testing) at six MSF facility-level sites in South Sudan: Gogrial Town, Gogrial (data available 2010–2016); Doro Refugee Camp, Maban (2012–2016); Bunj Town, Maban (2014–2016); Pibor Town, Pibor (2010–2016); Gumuruk Town, Pibor (2010–2016); and Leukongole Town, Pibor (2010–2016). Data were retrieved from the Medical Information Network for Operational Support (MINOS) system (v3.0.13, a web-based application used at field level for routine programme data capturing, by MSF Operational Centre Brussels. The data were extracted and imported into an electronic spreadsheet (Excel 2010; Microsoft Corp, Redmond, WA, USA).

Comparing methods to determine ‘extreme periods’ in Norway
Alongside the comparison of the AV+SD and the LR methods on weekly notified influenza cases in Norway, the moving epidemic method (MeM) was used in this analysis as gold standard for comparison purposes. MeM is widely applied to monitor influenza occurrence and uses retrospective data to define thresholds for an epidemic. For each county in Norway, MeM was applied using all available data to define extreme periods of influenza activity. This was taken as the gold standard and used as the baseline for comparison.

As per the AV+SD method, a number of historically comparable weeks are chosen as baseline, their means and SDs are calculated, and prediction intervals are generated. Any observation that is higher than a predetermined cut-off is considered to be extreme.3 For each county, the AV+SD method was run twice: first in a retrospective manner for each year using all previous data (up to 5 years maximum); second, using 2 years of prior data. Cut-offs corresponding to a two-sided α of 0.1, 0.05 and 0.01 were selected.

The novel method proposed here consisted of fitting a moving LR model to previous weekly data (8 weeks) that generated prediction intervals for each week of observation as a way to expect a future value based on true previous occurrences (and not on a population sample), regardless of their distribution. If the new observation exceeded a cut-off corresponding to the higher limit of the prediction interval, the new observation was determined to be extreme. Given 8 weeks of observation \(x_1, \ldots, x_8\) and \(y = y_1, \ldots, y_{\text{nth}}\), the higher limit of the prediction interval was calculated as:

\[
\text{AV+SD} = \left\{ \left( \text{slope}_{1, \ldots, 8} + \text{intercept}_{1, \ldots, 8} \right) + \frac{1}{8} \left( \sum (x_i - \bar{x})^2 + \text{SEE} \right) \right\}
\]

in which SEE is the standard error in estimate, \(\bar{x}\) the mean of \(x_1, \ldots, x_8\) and \(\sum (x_i - \bar{x})^2\) the sum of squared deviations from the mean. For each county, the LR method was run once, using 8 weeks of retrospective information. Cut-offs corresponding to a two-sided \(\alpha\) of 0.1, 0.05 and 0.01 were selected.

Finally, a method in which weeks were randomly classified as extreme/not-extreme (using a MeM frequency distribution based on all available data) was applied to examine by how much each method surpassed a random classification.

ACKNOWLEDGEMENTS
The authors give special thanks to L. Leyen and R. Embudo for the support they provided to the principal investigator in developing the reasoning behind the study. This research was supported through an operational research course, which was jointly developed and run by the Centre for Operational Research, the International Union Against Tuberculosis and Lung Disease (Paris, France) and the Operational Research Unit, Médecins Sans Frontières (Brussels, Belgium). Conflicts of interest: none declared.
To allow the AV+SD and LR methods to be evaluated over the same time periods, the first 2 years of results were removed for each site and method. For each site, method and year, the number of strikes (i.e., weeks above threshold) were counted in a cumulative manner. As the data did not meet the normality requirements of Pearson’s correlation coefficient, a Spearman’s correlation coefficient was calculated for each site and method in comparison to the gold standard. Unpaired t-tests were used to compare the mean of the correlation coefficients of each method against the randomly generated strikes (each site was considered as its own observation). Because MSF recommends the AV+SD method, a comparison was made using the unpaired t-test between this and the LR method. $P < 0.05$ was considered statistically significant. The analysis was performed with R Software v3.2.0 (R Foundation for Statistical Computing, Vienna, Austria).

**South Sudan case study**

Neither MeM nor AV+SD could be applied to the South Sudanese data because of missing, fragmented or non-comparable data across years. The LR method was run on the data set of each site up to Week 26 of 2016 (up to Week 20 of 2016 for the Gogrial data, as after that date the facility ceased to receive MSF support). For each site, the LR method was run once, at $\alpha = 0.01$ using eight consecutive non-zero weeks of data. The time series of weekly registered cases was plotted, with vertical lines indicating weeks above threshold (strikes). The strikes were visually observed and described. From an operational perspective, MSF considered the Gogrial 2015, Doro 2016 and Bunj 2016 to be ‘exceptional’ years, as they required an extraordinary response, including additional mobile clinics, mosquito net distributions and environmental spraying. The first strikes occurring in Gogrial 2015, Doro 2016 and Bunj 2016 were discussed from the perspective of the MSF response.

**Ethics**

As this study used routine, aggregate surveillance and programme data, consent was not required. This research fulfilled the exemption criteria set by the MSF Ethics Review Board (ERB; Geneva, Switzerland) for posteriori analyses of routinely collected data.

### TABLE

Mean yearly number of strikes (weeks above threshold) using the AV+SD method, the LR method and the random scenario and the test of difference between methods for the weekly notified influenza cases in 19 Norwegian counties (from Week 1 2006 to Week 40 2015)*

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Random scenario</th>
<th>LR (8 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.1$</td>
<td>Yearly strikes mean ± SD</td>
<td>2.8 ± 1.5</td>
</tr>
<tr>
<td></td>
<td>SCC mean ± SD</td>
<td>0.40 ± 0.12</td>
</tr>
<tr>
<td>AV+SD (5 years)</td>
<td>8.0 ± 6.5</td>
<td>0.73 ± 0.09</td>
</tr>
<tr>
<td>AV+SD (2 years)</td>
<td>9.7 ± 6.0</td>
<td>0.71 ± 0.08</td>
</tr>
<tr>
<td>LR (8 weeks)</td>
<td>6.4 ± 2.4</td>
<td>0.70 ± 0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Random scenario</th>
<th>LR (8 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.05$</td>
<td>Yearly strikes mean ± SD</td>
<td>2.8 ± 1.5</td>
</tr>
<tr>
<td></td>
<td>SCC mean ± SD</td>
<td>0.40 ± 0.12</td>
</tr>
<tr>
<td>AV+SD (5 years)</td>
<td>7.0 ± 6.4</td>
<td>0.71 ± 0.10</td>
</tr>
<tr>
<td>AV+SD (2 years)</td>
<td>8.8 ± 6.0</td>
<td>0.69 ± 0.08</td>
</tr>
<tr>
<td>LR (8 weeks)</td>
<td>3.9 ± 1.8</td>
<td>0.69 ± 0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cut-off point</th>
<th>Random scenario</th>
<th>LR (8 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.01$</td>
<td>Yearly strikes mean ± SD</td>
<td>2.8 ± 1.5</td>
</tr>
<tr>
<td></td>
<td>SCC mean ± SD</td>
<td>0.40 ± 0.12</td>
</tr>
<tr>
<td>AV+SD (5 years)</td>
<td>5.7 ± 6.4</td>
<td>0.66 ± 0.12</td>
</tr>
<tr>
<td>AV+SD (2 years)</td>
<td>7.4 ± 5.9</td>
<td>0.67 ± 0.10</td>
</tr>
<tr>
<td>LR (8 weeks)</td>
<td>1.6 ± 1.1</td>
<td>0.72 ± 0.13</td>
</tr>
</tbody>
</table>

*Differences are based on the mean SCC of the sum of strikes per method against the moving epidemic method (gold standard). The significance of differences is reported as $P$ values using the unpaired t-test.

AV+SD = average plus standard deviation; LR = linear regression; SCC = Spearman’s correlation coefficients.
FIGURE  Weekly number of confirmed malaria cases and strikes according to the linear regression method per Médecins Sans Frontières (MSF) facility in South Sudan (2010–2016). In each graph, the case count (vertical axis) is reported over time (horizontal axis, labelled as years). Vertical lines indicate weeks with strikes. *Years that required an extraordinary operational response by MSF.
RESULTS

Influenza in Norway
The Table shows the differences among the methods in terms of mean Spearman’s correlation coefficients of the sums of strikes against the gold standard in the 19 Norwegian counties. All methods performed significantly better than random assignment ($P < 0.01$). AV+SD did not perform better than the LR method ($P > 0.05$).

Malaria in South Sudan
The Figure shows the number of confirmed malaria cases and strikes according to the LR method per MSF facility in South Sudan from 2010 to 2016. By observation, strikes occurred immediately after the beginning of every seasonal wave (every year in Doro, Bunj and Gogrial) or during the seasonal wave, in correspondence of a new upsurge (e.g., Doro 2015). Strikes also occurred in Pibor, Leukongole and Gumuruk, where data were fragmented. Finally, strikes were registered when a descending trend stabilised (e.g., Gogrial 2014).

In Gogrial 2015, the first (and only) strike related to the 2015 malaria season occurred in Week 26. At that time, MSF had observed the “exceptionality” of the ongoing season (in comparison to previous years) and launched an emergency response (starting in week 34). Based on the Doro 2016 and Bunj 2016 data, the LR method indicated a strike in Week 21 and Week 22, respectively. At that time, MSF had remarked a sharp increase of cases and initiated an emergency response in these two sites.

DISCUSSION

This was an empirical study of different mathematical methods for monitoring disease occurrence. A set of complete influenza data from Norway allowed us to compare methods under the best conditions. Missing, fragmented or non-comparable malaria data from South Sudan could be used for the case study of only one method.

With the testing data set of influenza in Norway, the LR method only used 8 weeks of data, and yet it did not perform significantly worse than AV+SD methods, while handling the most common limitation of retrospective surveillance data i.e., missing data. These results justified the use of the LR method in real-world scenarios (as other methods would not work), i.e., a South Sudanese case study. Applied to South Sudan data, LR seemed to be able to indicate ‘exceptional’ occurrences of malaria at single facility level, and, on one occasion, to anticipate MSF operational response.

As a moving indicator, the LR method was sensitive to a variation of cases over a short time-period (unrelated to the magnitude of data). In South Sudan, the LR method gave strikes in all years regardless of operationally detected ‘exceptionality’, thus suggesting poor specificity. While the absence of operational responses could also be the result of various factors (e.g., environmental and logistic constraints faced in remote areas) and cannot therefore be considered an absolute measure of exceptionality, we suggest that the LR method could be used as a warning indicator, requiring a subsequent holistic quantitative-qualitative assessment. The method is not expected to be the sole trigger for an operational response, as no single observational method for case count can stand alone in malaria surveillance.

The literature reports different strategies for identifying “exceptional” malaria transmission. Classic methods are hampered by missing, fragmented or non-comparable data. In case of the MSF South Sudanese data set, the use of classic methods such as the AV+SD would have led to the exclusion of some subjectively identified outliers. As <5 years of retrospective data were available, the normality of case distribution and the applicability of a mean would have been affected. Even percentiles, which are applicable in case of non-normal distributions, could not be used due to the lack of consecutive observations. Where retrospective data are lacking, Teklehaimanot et al. proposed a moving indicator— the slope of the natural logarithm of the number of normalised cases—but reported its poor performance in raising alerts for outbreaks. We hypothesised that a LR of case counts was sufficiently reliable to observe potentially ‘exceptional’ short-term variations at a single facility level. Eight consecutive weeks of data were considered operationally relevant and sufficient without the need to use extensive antecedent data.

The main weakness of this analysis stems from its empirical methods. Also, influenza data from Norway may not be appropriate for testing a method intended to monitor the occurrence of malaria in endemic settings. However, it should be noted that this study was not designed to test the performance of a method on different diseases, but to compare different methods. Given the unreliability of malaria surveillance data in many low-resource settings, we used instead a reliable and complete data set of a disease occurring with a clear seasonal pattern. Because classic methods could not be applied to the data sets relevant to this analysis (i.e., South Sudanese data), a full comparison of LR against AV+SD could not be done. Surveillance data from other endemic contexts would not necessarily have been appropriate. MSF has been running long-term interventions in various refugee camps in sub-Saharan Africa. Despite being historically complete, these data are typically biased in their representativeness, as population movements and other contingent variations are common in these settings. Nonetheless, further research is required to test the performance and actual usefulness of the LR method using reliable and complete series of malaria cases.

MSF recommends using an AV+SD method but a plausible alternative seems to be available when retrospective data are missing. The performance of the LR method appeared to be similar to a method that requires much longer antecedent data. The LR method could be considered in those scenarios where retrospective data are missing, fragmented or non-comparable. It could be used when starting a new facility-based surveillance system, until enough historical data allows running a classic method. As with all surveillance tools, external and context-related information should be used to interpret the results in a holistic manner.

References


Contexte: Détecter des événements inhabituels en matière de paludisme, susceptibles de nécessiter une intervention opérationnelle constitue un défi surtout dans les zones d’endémie où la transmission est continue comme le Sud Soudan. Médécins Sans Frontières (MSF) utilise la méthode classique de moyenne et écart type (AV+SD) pour la surveillance du paludisme. De telles approches (comme les autres approches disponibles) dépendent de données passées qui sont souvent manquantes.

Objectif: Déterminer si une méthode basée sur la régression linéaire (LR) sur seulement huit semaines de données rétrospectives pourrait constituer une alternative à la méthode AV+SD.

Schéma: En l’absence de données historiques complètes relatives au paludisme au Sud Soudan, des rapports hebdomadaires relatifs à la grippe dans 19 contés de Norvège (2006–2015) ont été utilisés comme un ensemble de données test afin de comparer la performance des méthodes LR et AV+SD. La méthode *Moving epidemic method* a été utilisée comme étalon or. La méthode LR a ensuite été appliquée dans une étude de cas sur la survenue du paludisme dans les structures de MSF au Sud Soudan (2010–2016), observant la façon dont elle identifiait les événements relatifs au paludisme qui nécessitaient une riposte de MSF.

Résultats: En ce qui concerne les données norvégiennes relatives à la grippe, les méthodes LR et AV+SD ont eu une performance similaire (valeur de $P < 0.05$). En ce qui concerne les données relatives au paludisme au Sud Soudan, la méthode LR a identifié les périodes historiques auxquelles une riposte a été mise en œuvre.

Conclusion: La méthode LR semble être une alternative plausible à la méthode AV+SD quand les données rétrospectives sont manquantes.