Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Restructuring national health information system software to improve malaria data quality: activation of zero values and ninety-nine validation rules in Cameroon's DHIS2

Norbert Nkapleweh Lemonge

lnorbert@clintonhealthaccess.org

Clinton Health Access Initiative

Claude Ngwayu Nkfusai
Clinton Health Access Initiative

Bernard Nsah
Clinton Health Access Initiative

Herve Bella
Clinton Health Access Initiative

Sirima Yaya
Clinton Health Access Initiative

Moïse Hugue
Ministry of Public Health

Jean Fosso
Ministry of Public Health

Wilfred Ekokobe
Community Health Association

Mbonda Aime
Red Cross

Fai Karl Njuwa
Community Health Association

Marcellin Joël Ateba
Ministry of Public Health

Yauba Saidu
Clinton Health Access Initiative

Research Article

Keywords: Malaria, Data quality, DHIS2, Validation rules, Health information system, Completeness, Internal consistency, Cameroon

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11014186/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The District Health Information System 2 (DHIS2) has been Cameroon's principal platform for routine health facility reporting since 2015. A 2023 assessment of malaria surveillance identified inaccurate French–English translation, misalignment between the paper Monthly Activity Report (RMA) and the DHIS2 form, and automated deletion of zero values, which made true zeros indistinguishable from missing values. The 2024 RMA revision also rendered the 52 existing malaria validation rules obsolete.

Methods

In this national before-and-after operational study, four linked changes were made to the malaria section of DHIS2 in 2024: variables were renamed and retyped as positive integer or zero; storage of zero values was enabled in April 2024; ninety-nine validation rules were configured and tested at a four-day workshop in May 2024; and variables and error messages were translated bilingually and aligned with the paper RMA. Completeness of malaria variables and internal consistency (five indicator-pair relationships) were compared between January–March and April–June 2024. Analyses were descriptive; no significance tests were applied.

Results

The restructured malaria section comprised seven sub-sections, 91 variables and 416 bilingual entry fields. Validation rules rose from 52 to 99, all configured and tested within the workshop. Mean completeness across the ten regions rose from 51.3% to 69.7%, an increase of 18.4 percentage points, and rose in every region, from + 7.6 percentage points in Far North to + 24.4 in Adamawa. Across the five indicator pairs, the proportion of reports free of the specified inconsistency rose from a mean of 88.2% to 92.6%.

Conclusions

Small, inexpensive configuration changes to a national reporting platform were followed by measurable improvements in the completeness and internal consistency of routine malaria data. The completeness gain must be interpreted with care: enabling storage of zero values changed what completeness measures as well as what facilities report, so part of the increase is definitional. The consistency gain is more directly attributable to the rules, which act at the point of data entry. Sustaining these gains will require replacing the annual deletion of zero values with archiving, differentiating rule importance levels, and adding visual alerts to the entry form.

BACKGROUND

Malaria remains one of the largest causes of preventable illness and death in sub-Saharan Africa. The World Health Organization (WHO) estimated 282 million malaria cases and 610,000 malaria deaths globally in 2024, with the WHO African Region accounting for 94% of cases and 95% of deaths [1]. Cameroon is among the eleven African countries carrying the highest burden [1]. Malaria accounts for 28% of outpatient consultations, 18% of hospital admissions and 7.3% of deaths in health facilities, making it the leading cause of morbidity in the country [2]. In the framework of the fight against malaria in high-burden countries, WHO has identified the use of strategic information to guide decision-making as one of the key response areas [3, 4].

Strategic information can only guide decisions if it exists and is credible [5]. In Cameroon, the Monthly Activity Report (Rapport Mensuel d'Activités, RMA) completed by each health facility is the integrated monthly reporting tool for services delivered, and it is the principal source that feeds the national health information system (Système National d'Information Sanitaire) and, through it, the DHIS2 [6]. DHIS2 has been used as the principal tool for routine data collection from both private and public health facilities in Cameroon since 2015. The RMA was last revised in 2020 and again in 2023, at a workshop held at Mfou from 24 to 29 July 2023 at which stakeholders across the health system validated the inputs for the revision of RMA I, II and III; the revised paper forms and their electronic versions in DHIS2 were updated accordingly [6].

The WHO Data Quality Assurance framework distinguishes four dimensions of routine facility data quality: completeness and timeliness of reporting; internal consistency, which covers outliers, trends over time and the coherence of related indicators; external comparison with other data sources; and the consistency of the population data used as denominators [7]. The first two dimensions are properties of the reporting system itself and are therefore the dimensions most directly exposed to the design of the software in which the data are entered. Evidence from countries using DHIS2 shows that both remain problematic at scale. A retrospective assessment of 265 facilities and 77 data elements in the Eastern Cape province of South Africa found completeness of 74.6% and internal consistency of 95.1%, with 121,199 missing values and 5,670 validation errors among 365,228 reported data element values [8]. In Kenya, a study of HIV indicator data extracted from the national DHIS2 found that 50.2% of the 93,179 facility-records extracted contained no data at all, and identified reporting rates exceeding 100%, duplicate facility records and reports submitted empty despite services having been provided [9].

What receives less attention is that the software itself can be the proximate cause of these defects. Three features of the Cameroonian DHIS2 instance illustrate the point. First, by default the platform did not store zero values: a facility that recorded a true zero for a malaria variable had that entry deleted, so a genuine zero could not be distinguished from a value that had never been entered. Completeness, defined as the proportion of expected values actually present, therefore could not be measured, and the reported figure necessarily understated what facilities had in fact reported. Second, the malaria section of the paper RMA and the corresponding DHIS2 form were misaligned, and translations of variable labels

and system messages between French and English were partial or inaccurate. In a country with two official languages this is not a cosmetic problem; a data clerk who cannot read the label of the field being completed cannot complete it correctly. Third, the revision of the RMA in 2024 disaggregated the malaria section further and thereby rendered the 52 validation rules previously configured for that section obsolete, leaving the newly disaggregated variables with no automated coherence checks at all.

An earlier evaluation of the national malaria surveillance system had already documented erroneous classification of malaria cases, poor completion of registers, weak mastery of the data collection tools and difficulties integrating community case management data into routine surveillance [10]. In 2023, CHAI conducted an assessment of the malaria surveillance system in Cameroon through a literature review, key informant interviews and focus group discussions. The assessment examined data flows and the limitations of existing databases, identified gaps in data collection and analysis processes, considered how collection tools could be improved with the contribution of frontline staff, explored the integration of community case management data into the national health information system, and identified opportunities for better use of data. Interviews with end users described the consequences of the design problems directly. One respondent explained that "the RMA doesn't match what's online and sometimes the indicators are in French — if it doesn't match, just type something in to move on to the next; you can't read or understand what it actually says." Another observed that "the DHIS2 interface is very complicated, which is why even those who can do it don't get it right easily." A third noted that "many things change and evolve daily on the DHIS2, except that when things change, nobody calls to tell you."

Following these findings, CHAI, the National Malaria Control Programme (NMCP) and the Health Information Unit of the Ministry of Public Health (Cellule des Informations Sanitaires, CIS) agreed a set of concrete changes to the functionality of the DHIS2, implemented in 2024 under the project Strengthening Surveillance and Data Use in Cameroon, funded by the Bill & Melinda Gates Foundation. This paper reports what was changed, how it was done, and what happened to the completeness and internal consistency of malaria data in the two quarters either side of the change. Figure 1 summarises the problems identified, the actions taken, and which dimension of data quality each action was intended to affect.

METHODS

Study design

This was a descriptive, national before-and-after operational study. The unit of intervention was the national DHIS2 instance; the changes were therefore applied simultaneously to all ten regions and 202 health districts, and no contemporaneous control group was available. Two periods were compared: the first quarter of 2024 (January to March), before the changes were made, and the second quarter of 2024 (April to June), after them.

Setting

Cameroon has ten administrative regions and 202 health districts. Malaria transmission is perennial and hyperendemic, with intensity varying by ecological zone. Routine malaria data are recorded in facility registers, aggregated monthly onto the paper RMA, and entered into the national DHIS2 instance by facility or district staff. Both public and private facilities report through this channel, although private-sector participation is incomplete. The country is bilingual, and the DHIS2 instance must serve users working in French and in English.

The intervention

Four linked changes were made to the malaria section of the DHIS2 in 2024. All were configuration changes made within the existing platform; none required new software, new licences or new hardware.

Restructuring and renaming of variables. All variables in the malaria section were reviewed against the physical copies of the revised RMA. Partial translations were removed from the Translations module and every variable was retranslated into French and English from the paper form. Variables were renamed to an explicit nomenclature of the form PNLP_SubSectionNumber_VariableNumber.VariableName, so that the position of a variable in the paper form can be read directly from its name in the database. The variable type was changed to "positive integer or zero" in the Maintenance module. Indicators were renamed on the same principle and updated where the revised RMA required it. The resulting structure of the malaria section is described in Table 1.

Activation of zero values. For every variable in the malaria section the option "store zero values" was checked in the Maintenance module, so that a zero entered by a facility is retained in the database rather than deleted. This change took effect in April 2024. It makes a true zero distinguishable from a missing value and therefore makes completeness measurable. It was agreed with the programme that zero values for clinical variables would be removed from the DHIS2 in the first quarter of the following year to contain the growth of the database, and that the programme would need to establish a mechanism for archiving the validated annual database before that deletion occurs.

Development, configuration and testing of validation rules. A validation rule in DHIS2 is an expression defining a numerical relationship between data element values [11]. Each rule has a left side and a right side whose validity is compared using a comparison operator, and each carries an instruction that is displayed to the user when the relationship fails. The NMCP drafted 99 rules covering the variables of the revised malaria section, replacing the 52 rules that the revision had made obsolete. The rules were configured and tested at a four-day workshop held at the Hôtel Prodiges in Mfou from 7 to 10 May 2024 [12], chaired by a CIS official and bringing together representatives of the Secretariat General, the Cooperation Unit, the CIS and the NMCP, with CHAI as technical and financial partner.

For each rule the following fields were completed in the national DHIS2 instance: name and short name, following the same prefix convention as the data elements (PNLP2024_G[group number]_[rule name]); instruction, taken from the bilingual error message in the rule specification document [13]; importance, set to medium; period type, set to monthly; left side and right side, each with the missing value strategy set to "never skip" so that an absent value does not cause the rule to be bypassed; the comparison operator; and organisation unit level, set to the health facility. Rules were then grouped into validation rule groups by form, so that the scheduler can be configured to send notification e-mails when validation errors occur. Table 2 gives the anatomy of three representative rules.

Participants worked in four groups, each responsible for a contiguous block of rules and for a defined set of RMA sub-sections. Progress was reviewed in plenary at the end of each day. On day 4 the rules were tested by selecting a health facility that had entered no data since January (CSI Lugga Tapadji, in the Adamawa region), reactivating it for the purpose, and entering a January-to-April dataset; each group tested at least eight of its rules, comprising its first four and its last four. Discrepancies were discussed within groups and then in plenary, amendments were made, and all 99 rules were validated. One rule, stating that the number of cases not tested should be greater than the number of expected cases among people aged five years and over, was deleted as incorrect.

Bilingual translation and alignment. All variables, indicators and validation error messages were rendered in both French and English and aligned with the wording of the paper RMA, so that the label a user reads on screen matches the line being copied from the paper form. The bilingual error messages are displayed to the user at the point of entry. Visual alerts, such as colouring the fields implicated in a failed rule, were requested but were not implemented.

Outcome measures

Two outcomes were measured, corresponding to the first two dimensions of the WHO Data Quality Assurance framework [7]:

- **Completeness of malaria variables**: the proportion of expected malaria data element values that were present in the DHIS2. Following activation of zero values, an entered zero counts as a present value; before activation it did not. Completeness was computed for each of the ten regions for the quarter before and the quarter after activation, and monthly for the country as a whole.
- **Internal consistency**: the proportion of monthly facility reports in which a specified relationship between two related quantities held. Five relationships were examined, chosen because they trace the malaria case-management cascade: suspected cases not exceeding consultations; cases tested not exceeding suspected cases; cases confirmed not exceeding cases tested; cases treated not exceeding cases confirmed for uncomplicated malaria; and cases treated not exceeding cases confirmed for severe malaria. Reports entered between January and March 2024, before the updated validation rules existed, were compared with reports entered between April and June 2024. Because the rules were configured between 7 and 10 May 2024, reports entered in April were not

subject to them; the second period therefore mixes one month without the rules and two months with them, which biases the post-change estimate downwards.

Two process outcomes are also reported: the composition of the rule set by clinical domain and comparison operator, and the proportion of rules configured by each working group at the end of each workshop day.

Data sources and analysis

Data were extracted from the national DHIS2 instance. The rule specification document produced by the NMCP and the workshop report produced by the CIS and the NMCP were used to characterise the rule set and the configuration process; the exported custom data-entry form for the malaria section of the RMA was used to characterise the restructured section. Analyses were performed in Excel. Where more than one extraction of the same indicator was available, the extraction covering all ten regions was used, and the alternative values are noted in the limitations. Regional means are unweighted averages across regions and are not weighted by the number of reporting facilities. No statistical tests were applied, no confidence intervals were computed and no adjustment was made for secular trend or seasonality; all findings are therefore descriptive and no difference reported here should be read as statistically significant. The study is reported in accordance with the STROBE statement for observational research, and the completed STROBE checklist is submitted as a supplementary file. No questionnaire, survey instrument or other primary data collection tool was developed or administered for this study. Permission to access and analyse the national DHIS2 data was granted by the Ministry of Public Health of Cameroon

Ethical considerations

The analyses reported here used aggregate, non-identifiable routine data already held in the national health information system, together with programme documentation. No individual patient data were accessed and no human participants were enrolled for this analysis. The study was conducted in accordance with the principles of the Declaration of Helsinki. The wider surveillance strengthening study within which this work was conducted was reviewed and approved by the Cameroonian National Ethics Committee for Human Health Research (N°2023/04/1540CE/CNERSH/SP of 5 May 2023), by CHAI's Scientific and Ethical Review Committee and by an external United States institutional review board holding an active Federal Wide Assurance. Data were stored on encrypted, password-protected computers accessible only to authorised personnel. Ownership of the data belongs to the Government of Cameroon.

RESULTS

The restructured malaria section

The malaria section of the revised RMA as configured in the DHIS2 comprises seven sub-sections, 91 numbered variables, 101 distinct data elements and 416 data-entry fields, each field corresponding to one data element and one category option combination such as an age band or sex. Every field carries a paired French and English label. The distribution of variables across sub-sections is shown in Table 1: morbidity, mortality and case management by age accounts for the largest share, followed by the management of antimalarial commodities in health facilities.

Table 1
Structure of the malaria section of the revised RMA as configured in the DHIS2

Section	Title	Numbered variables	Share of variables
IV.2.1	Intermittent preventive treatment in pregnant women	3	3%
IV.2.2	Routine distribution of LLINs to pregnant women	1	1%
IV.2.3	Communication for development	6	7%
IV.2.4	Morbidity, mortality and case management by sex and age	20	22%
IV.2.5	Morbidity, mortality and case management by age	32	35%
IV.2.6	Intermittent preventive treatment in infants in the community	8	9%
IV.2.7	Management of antimalarial commodities in health facilities	21	23%
Total	Seven sub-sections; 101 data elements; 416 bilingual data-entry fields	91	100%

The validation rule set

The number of validation rules for the malaria section rose from 52 before the revision to 99 after it, an increase that follows directly from the greater disaggregation of the revised RMA rather than from the addition of new logical checks to unchanged variables. Of the 99 rules, as specified by the programme [13], 87 are upper-bound rules using the operator "less than or equal to" and 12 are sum-equality rules using "equal to". The upper-bound rules enforce the arithmetic of the case-management cascade, in which a downstream quantity cannot exceed the upstream quantity from which it is drawn; the sum-equality rules enforce that disaggregations add to their totals, for example that cases tested by rapid diagnostic test plus cases tested by microscopy equal cases tested. Ninety-nine rules carry a French error message and 98 carry an English one.

The distribution of rules across clinical domains is shown in Fig. 2. Community case management by community health workers and testing and confirmation each account for 28 rules, and severe malaria and hospitalisation for a further 21; together these three domains hold 77 of the 99 rules. Preventive interventions in pregnancy are covered by a single rule each. Most rules are replicated across the

population strata into which the revised RMA disaggregates its variables, namely children under five years, people aged five years and over, and pregnant women, and by sex where the form disaggregates by sex.

Table 2
Anatomy of three representative validation rules

Rule	Left side	Operator	Right side and error message displayed
18	Uncomplicated malaria cases confirmed and treated with ACT, 0–5 years, female	≤	Suspected uncomplicated malaria cases testing positive, 0–5 years, female. "The number of simple malaria cases confirmed and treated by ACT (5) should be less than or equal to the number of suspected simple malaria cases tested positive (4) in female children under 5 years old."
62	Deaths from confirmed malaria, 0–5 years, female	≤	Deaths from illness, all causes, 0–5 years, female. "The number of deaths due to confirmed malaria (20) should be less than or equal to the number of deaths due to illness all causes (19) in female children under 5 years old."
70	Suspected malaria cases tested by RDT, 0–5 years, plus suspected malaria cases tested by microscopy, 0–5 years	=	Suspected malaria cases tested, 0–5 years, female plus male. "The sum of suspected malaria cases (simple and severe) tested by RDT and microscopy (23 + 24) should be equal to number of suspected malaria cases (simple or severe) tested (3) in children under 5 years of age (male and female)."

Configuration and testing of the rules

All 99 rules were configured and tested within the four days of the workshop. Progress was uneven and back-loaded: at the end of day 2 the four groups had configured 7 of 24, 11 of 24, 6 of 23 and 4 of 22 rules respectively, a total of 28 of 93 rules or 30% of the work then scoped. By the end of day 3 all four groups reported completion, including six rules added to the first group's block, giving 99 of 99 rules configured (Fig. 3). Day 4 was devoted to testing and amendment, after which all 99 rules were validated in plenary.

Completeness of malaria variables

Mean completeness of malaria variables across the ten regions was 51.3% in the quarter before activation of zero values and 69.7% in the quarter after it, an increase of 18.4 percentage points. Completeness increased in every region without exception (Fig. 4). The increase ranged from 7.6 percentage points in Far North, which had the highest baseline at 72.9%, to 24.4 percentage points in Adamawa and 23.8 in South. Far North was the only region above 70% before the change; after it every region was above 61%, nine of the ten were above 63% and four were above 74%. The rank order of regions changed, indicating that the change did not scale all regions equally: South rose from tenth place to seventh, while Far North, which began from the highest baseline, recorded the smallest absolute gain.

The monthly series locates the change in time (Fig. 5). National completeness was flat across the first quarter at 50.1%, 48.8% and 49.4%, rose to 57.6% in April, the month in which zero values were activated, reached 71.2% in May and was 71.0% in June. The same step is visible in each of the three regions for which a monthly series was available, with Adamawa rising from 51.4% in March to 81.0% in May, East from 54.7% to 77.7% and Centre from 42.1% to 60.7%. In all three the June value was marginally below the May value.

Completeness did not become complete. Missing values continued to be observed, and they were concentrated in variables corresponding to interventions that the reporting facility does not deliver, such as intermittent preventive treatment in infants and the disaggregation of test results by Plasmodium species, and in variables affected by major stock-outs of the corresponding commodity. In other words, after activation of zero values the residual missingness became interpretable: what remains blank now largely indicates an activity not performed or a commodity not available, rather than an unknown mixture of true zeros and unrecorded values.

Internal consistency of reported data

Before the updated validation rules were in place, the proportion of reports containing an inconsistency exceeded 10%. Across the five indicator pairs, the mean proportion of reports free of the specified inconsistency rose from 88.2% for reports entered between January and March 2024 to 92.6% for reports entered between April and June 2024, an average increase of 4.4 percentage points (Fig. 6). The rules were configured in early May, so April entries within the second period were not subject to them. Every pair improved. The improvement was largest for the pair with the worst baseline: reports in which treated cases did not exceed confirmed cases for uncomplicated malaria rose from 75% to 86%, a gain of 11 percentage points. The pairs already above 90% improved by between 1 and 4 percentage points, with suspected cases not exceeding consultations reaching 98% and cases tested not exceeding suspected cases reaching 96%.

These changes did not occur in isolation. Over the same period, quarterly district-level data review meetings were being held in three of the ten regions, and feedback on data quality was being provided to facilities through those meetings and through data quality reviews. Both were reported by the implementing team as having contributed to the improvement in data quality.

DISCUSSION

What the completeness gain does and does not show

The 18.4 percentage point rise in completeness is the largest number in this paper and the one most easily over-interpreted. Before April 2024 a zero entered by a facility was deleted by the platform, so it was counted as a missing value; after April 2024 the identical reporting behaviour produces a stored value and is counted as present. A substantial part of the increase is therefore a change in measurement rather than a change in what facilities do. This does not make the gain unimportant, but it changes what

it means: what improved first is the ability of the system to represent what facilities report, and only thereafter, if at all, the reporting itself. Any interpretation of the figure as an 18-point improvement in facility reporting behaviour would be unsupported by this design.

Two observations argue that the change is nonetheless substantive. First, completeness is now measurable at all, which it was not before; earlier spatiotemporal analyses of Cameroonian routine malaria data had to adjust reported case counts for incomplete reporting before incidence could be estimated [14]. The monitoring of completeness that the National Strategic Plan requires was previously impossible for the malaria section, because the denominator of expected values could not be separated from an unknown number of suppressed zeros. Second, the residual missingness has become diagnostic. Variables systematically left blank now correspond to interventions not implemented and to commodities out of stock, so a blank cell has become a programmatic signal rather than an artefact of the software. This is the mechanism by which a completeness metric becomes useful to a programme manager.

The magnitude of the gain also varied in an informative way. Far North, the only region with baseline completeness above 70%, gained least, and the regions with the lowest baselines gained most. This is what would be expected if the suppression of zeros had been depressing measured completeness most severely where facilities report many genuine zeros, which is to say in lower-volume settings, and it is inconsistent with a simple uniform inflation of all values.

The validation rules and internal consistency

The consistency outcome can be attributed to the validation rules more directly, because the rules act at the moment and place where the error would be committed. The pattern of improvement supports this. The gain was concentrated in the single pair with the worst baseline, treated cases against confirmed cases for uncomplicated malaria, where 25% of reports were inconsistent before the rules and 14% afterwards. That relationship is precisely the one that a data clerk copying from a register is most likely to violate, because treatment and confirmation are recorded in different columns and at different times, and it is the relationship that rules 18 to 22 of the rule set were written to catch. The pairs that were already above 90% had little room to improve and improved little.

The effect is nevertheless modest, and the design of the rules explains why. All 99 rules were configured with importance set to medium and none blocks submission; the user is shown a bilingual message and may proceed. The rules therefore operate by prompting rather than by prevention. The workshop itself recognised this and recommended that the importance level be differentiated for key variables and that fields implicated in a failed rule be highlighted visually, neither of which had been implemented at the time of this analysis. A plausible reading of the 4.4 percentage point average gain is that a non-blocking prompt captures the errors a clerk is willing to correct when told, and that the remainder require either a hard constraint or a correction process outside the software.

By comparison, the Eastern Cape assessment in South Africa found internal consistency of 95.1% with validation rules long established [8], which is close to the 92.6% reached here in a single quarter. The Kenyan experience, where half of extracted facility-records contained no data and reporting rates above 100% were observed [9], illustrates the class of defect that upper-bound and sum-equality rules are designed to make impossible at entry rather than to detect at analysis.

The cost of a single click

The activation of zero values on the DHIS2 amounted, technically, to changing a variable type and checking a box in the Maintenance module for each variable. Achieving it required a national assessment, a series of formal and informal consultations with the stakeholders responsible for the management of the DHIS2 and with the NMCP, agreement that the instance would be restructured, and a policy decision about how long zero values would be retained. The 99 validation rules required a four-day residential workshop of four working groups drawn from four Ministry directorates and a technical partner. The distance between the technical triviality of the change and the institutional effort needed to make it is, in our view, the main transferable finding of this work. Programmes seeking to improve data quality through software configuration should budget for the consultation, not for the configuration.

The corollary is that the effort is small in absolute terms and non-recurrent. No new software, licence or hardware was required. Once configured, the rules apply to every facility in the country at every monthly entry, which is a different cost profile from interventions that scale with the number of facilities or the number of reporting periods.

Sustaining the gains

Three specific threats to these gains follow from the way the changes were implemented. First, it was agreed that zero values for clinical variables would be deleted from the DHIS2 in the first quarter of the following year in order to contain the growth of the database. Unless the validated annual database is archived before that deletion, the completeness series reported here will not be reproducible for 2024 and completeness will become unmeasurable again in retrospect. Establishing that archiving mechanism is, on the evidence of this study, the single most consequential outstanding action.

Second, the rules were configured at uniform medium importance with no visual alerts. Differentiating importance for the variables that matter most to the programme, and highlighting the implicated fields, were both recommended at the workshop and would be expected to increase the proportion of prompts acted upon.

Third, the changes were not accompanied by systematic communication to the users affected. The 2023 assessment recorded users complaining that "many things change and evolve daily on the DHIS2, except that when things change, nobody calls to tell you." The workshop recommended that the NMCP communicate about the new rules, and listed as next steps the production and circulation of an aide-mémoire, the revision of the RMA completion guide, and the training of field actors. A change in software behaviour that users do not understand is a change that will be worked around.

Position within a wider programme of work

The changes reported here were framed by the implementing partners as the first step in a sequence of DHIS2 optimisation, to be followed by improved geolocation of health facilities within the health pyramid, the establishment of a master facility list and a common facility reference, updating of the malaria dashboards, and the development of a malaria data repository on DHIS2 to support use cases including data quality review, on-site data quality audit, intervention mapping, epidemic alerts, monitoring of routine and campaign activities, and progress reporting. Software configuration addresses the quality of what enters the system; those subsequent components address what can be done with it once it is there.

Limitations

Several limitations qualify these findings.

- The design is a before-and-after comparison with no control group. The intervention was applied to the single national DHIS2 instance, so no region could serve as a comparator, and the observed changes cannot be separated from anything else that occurred nationally between the first and second quarters of 2024.
- The completeness outcome is partly definitional, as set out above under What the completeness gain does and does not show. The measure changed at the same time as the thing being measured.
- Two co-interventions were active over the same period. Quarterly district-level data review meetings were being held in three of the ten regions and data quality feedback was being provided to facilities; both were reported to have contributed to the improvement. That completeness rose in all ten regions, including the seven without district-level meetings, argues that the software change had an independent national effect, but the size of that effect cannot be separated from the co-interventions.
- Only one quarter of follow-up is available. May was the peak in every series examined and the June value was marginally below it, and it is not possible to say from two post-change months whether the gain is durable, whether it will continue to rise as users become familiar with the new form, or whether it will decay.
- Regional means are unweighted averages across regions and are not weighted by the number of reporting facilities, so a region with few facilities carries the same weight as one with many. Facility-weighted national estimates would differ.
- Two extractions of regional completeness were available and they disagree. The extraction used here, covering all ten regions, gives Adamawa 52.8% to 77.2%, East 57.2% to 74.1% and Centre 47.7% to 61.6%; a second extraction covering only those three regions gives 50.1% to 74.5%, 54.6% to 72.9% and 43.3% to 56.4% respectively. The direction and approximate magnitude of the change are the same in both, but the absolute values are not reproducible between them.

- The five indicator pairs used to measure internal consistency were selected because they trace the case-management cascade and were the pairs for which extracts were available; they are not a random or exhaustive sample of the relationships that the 99 rules enforce, and the average across them should not be read as an estimate of overall internal consistency.
- No statistical tests were applied and no confidence intervals are presented, so the precision of the estimates cannot be judged and none of the differences reported can be described as significant.
- Timeliness of reporting, the second component of the first WHO data quality dimension, was not analysed here, and the two remaining dimensions of the framework, external comparison and consistency of population data, were outside the scope of this study; the 2022 Malaria Indicator Survey would be the natural external comparator for such an analysis [15].
- The effect of the renaming of variables and of the bilingual retranslation was not measured. Both were implemented, and users had described translation as a direct cause of erroneous entry, but this study reports no outcome attributable to either.

CONCLUSIONS

Validation rules, the activation of zero values and other small changes in nationally used software such as the DHIS2 can be used to obtain higher-quality reported data from health facilities nationally. In Cameroon, enabling the storage of zero values was followed by a rise in the mean completeness of malaria variables from 51% to 70% across all ten regions, and the configuration of 99 new validation rules was followed by a rise in the proportion of reports free of a specified inconsistency from 88% to 93% on average, with the largest gain where the baseline was worst.

The two findings should be read differently. The completeness result is in the first instance a gain in the system's capacity to represent what facilities report, because a stored zero and a suppressed zero are not the same datum; its programmatic value is that residual missingness has become interpretable as an intervention not delivered or a commodity out of stock. The consistency result is more plausibly behavioural, because the rules act at the point of entry, and its size is consistent with what a non-blocking prompt can be expected to achieve.

Although the activation of zeros on the DHIS2 seemed to be an easy click on the software, a good deal of stakeholder consultation and meetings were necessary to get it done. The technical cost of these changes was negligible and non-recurrent; the institutional cost was substantial. Programmes contemplating similar work should plan accordingly, and should secure three things at the outset: a mechanism for archiving the validated annual database before agreed deletions of zero values remove the ability to measure completeness retrospectively; differentiated importance levels and visual alerts, so that rules for the variables that matter most are harder to bypass; and a communication and training plan, because users who are not told what changed will find ways around it. Whether the gains reported here persist beyond one quarter, and how much of the completeness gain reflects a real change in reporting behaviour, are questions that a longer series and a facility-level analysis could answer.

Abbreviations

ACT Artemisinin-based combination therapy

ANC Antenatal care

BMGF Bill & Melinda Gates Foundation

CHAI Clinton Health Access Initiative

CHW Community health worker

CI Cellule de la Coopération

CIS Cellule des Informations Sanitaires (Health Information Unit)

DHIS2 District Health Information System 2

DQA Data quality assurance

DRM Data review meeting

FGD Focus group discussion

HMIS Health management information system

IPTi Intermittent preventive treatment in infants

IPTp Intermittent preventive treatment in pregnancy

IRB Institutional review board

KII Key informant interview

LLIN Long-lasting insecticidal net

MINSANTE Ministère de la Santé Publique (Ministry of Public Health)

NMCP National Malaria Control Programme

PNLP Programme National de Lutte contre le Paludisme

RDT Rapid diagnostic test

RMA Rapport Mensuel d'Activités (Monthly Activity Report)

SG Secretariat General

SNIS Système National d'Information Sanitaire

SOP Standard operating procedure

WHO World Health Organization

Declarations

Ethics approval and consent to participate

This analysis used aggregate, non-identifiable routine data held in the national health information system, together with programme documentation; no human participants were enrolled and no individual consent was required. The study was conducted in accordance with the principles of the Declaration of Helsinki. The wider surveillance strengthening study within which this work was conducted was approved by the Cameroonian National Ethics Committee for Human Health Research (N°2023/04/1540CE/CNERSH/SP of 5 May 2023), by the Clinton Health Access Initiative (CHAI) Scientific and Ethical Review Committee, and by an external United States institutional review board holding an active Federal Wide Assurance. Permission to access and analyse the national DHIS2 data was granted by the Ministry of Public Health of Cameroon.

Consent for publication

Not applicable. The manuscript contains no data, images or details relating to an identifiable individual.

Competing interests

NL, CNN, BN, SDN, HB, SY and YS are employed by the Clinton Health Access Initiative (CHAI), which provided technical and financial assistance to the Ministry of Public Health of Cameroon for the DHIS2 changes evaluated in this manuscript. MHRA, JF and MJA are staff of the National Malaria Control Programme, Ministry of Public Health of Cameroon, the institution responsible for the national malaria surveillance system described here. WE and MA reviewed the article for conformity and content strength. The work was funded by the Bill & Melinda Gates Foundation. Beyond these institutional roles, the authors declare that they have no financial or non-financial competing interests.

Funding

This work was carried out under the project Strengthening Surveillance and Data Use in Cameroon, funded by the Bill & Melinda Gates Foundation (grant ID BMGFMSURV2, project ID CHCAMEMELIM1). The funder had no role in the design of the study, in the collection, analysis or interpretation of the data, in the writing of the manuscript, or in the decision to submit it for publication.

Author Contribution

Conceptualisation: NL, CNN, YS, MJA. Methodology: NL, CNN, BN, MHRA. Software and system configuration: NL, SDN, HB, JF. Validation: BN, SY, MHRA, MJA, WE, MA. Formal analysis: NL, CNN, SDN. Investigation: NL, BN, HB, SY, JF. Data curation: SDN, HB, JF. Writing – original draft: NL, CNN. Writing – review and editing: all authors. Visualisation: NL, SDN. Supervision: YS, MJA. Project administration: NL, SY. Funding acquisition: YS. All authors read and approved the final manuscript.

Acknowledgement

We are grateful to the Ministry of Public Health of Cameroon, and in particular to the Secretariat General, the Cellule de la Coopération and the Cellule des Informations Sanitaires, whose collaboration made the changes to the DHIS2 possible. We thank the National Malaria Control Programme and the members of the Surveillance Technical Working Group. The data analysed here were generated by health facility staff who work relentlessly in the fight against malaria, and to whom we are grateful.

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

References

1. World Health Organization. World malaria report 2025. Geneva: World Health Organization; 2025.
2. National Malaria Control Programme. Annual report. Yaoundé: Ministry of Public Health, Cameroon; 2023.
3. World Health Organization. High burden to high impact: a targeted malaria response. Geneva: World Health Organization; 2018.
4. World Health Organization. Global technical strategy for malaria 2016–2030, 2021 update. Geneva: World Health Organization; 2021.
5. World Health Organization. Malaria surveillance, monitoring and evaluation: a reference manual. 2nd ed. Geneva: World Health Organization. 2025. Available from: https://www.who.int/publications/i/item/9789240112476. Accessed 16 Sept 2026.
6. Programme National de Lutte contre le Paludisme, Cellule des Informations Sanitaires. Report from the revision of the monthly progress reports (RMA) in Mfou from the 24th to the 29th July 2023. Yaoundé: Ministère de la Santé Publique; 2023.
7. World Health Organization. Data quality assurance. Module 1: framework and metrics. Geneva: World Health Organization. 2022. Available from: https://iris.who.int/handle/10665/366086. Accessed 16 Sept 2026.

8. Thabethe SS, Mathews VE. Quality of routine health data in DHIS2 in South Africa: Eastern Cape province from 2017–2020. S Afr J Inf Manag. 2024;26(1):a1903. 10.4102/sajim.v26i1.1903.
9. Gesicho MB, Were MC, Babic A. Data cleaning process for HIV-indicator data extracted from DHIS2 national reporting system: a case study of Kenya. BMC Med Inf Decis Mak. 2020;20:293. 10.1186/s12911-020-01315-7.
10. PMI Measure Malaria. Évaluation du système de surveillance du paludisme au Cameroun: rapport final. TR-22-500. Chapel Hill (NC): MEASURE Malaria, Carolina Population Center, University of North Carolina at Chapel Hill. 2022. Available from: https://measuremalaria.cpc.unc.edu/wp-content/uploads/2022/12/TR-22-500-ESSP-Cameroon-Rapport-508.pdf. Accessed 16 Sept 2026.
11. DHIS2. Data quality: implementing data quality functionality and procedures. DHIS2 implementation guide. Oslo: University of Oslo. 2024. Available from: DHIS2 Documentation. https://docs.dhis2.org/en/implement/data-quality/implementation-of-funtionality-and-procedures.html. Accessed 11 Sept 2026.
12. Programme National de Lutte contre le Paludisme, Cellule des Informations Sanitaires. Rapport de l'atelier d'actualisation des règles de validation de la section lutte contre le paludisme du rapport mensuel d'activités dans le logiciel DHIS2. Mfou, 7–10 mai 2024. Yaoundé: Ministère de la Santé Publique; 2024.
13. Programme National de Lutte contre le Paludisme. Règles de validation dans DHIS2 des données de lutte contre le paludisme collectées en routine dans les formations sanitaires à travers le RMA révisé de 2024. Yaoundé: Ministère de la Santé Publique; 2024.
14. Danwang C, Khalil É, Achu D, Ateba M, Abomabo M, Souopgui J, et al. Fine scale analysis of malaria incidence in under-5: hierarchical Bayesian spatio-temporal modelling of routinely collected malaria data between 2012–2018 in Cameroon. Sci Rep. 2021;11(1):11408. 10.1038/s41598-021-90997-8.
15. Institut National de la Statistique, Programme National de Lutte contre le Paludisme, ICF. Enquête sur les Indicateurs du Paludisme au Cameroun 2022: rapport final. Yaoundé (Cameroun) and Rockville (MD): INS, PNLP and ICF. 2023. Available from: https://dhsprogram.com/pubs/pdf/MIS43/MIS43.pdf. Accessed 16 Sept 2026.

Figures

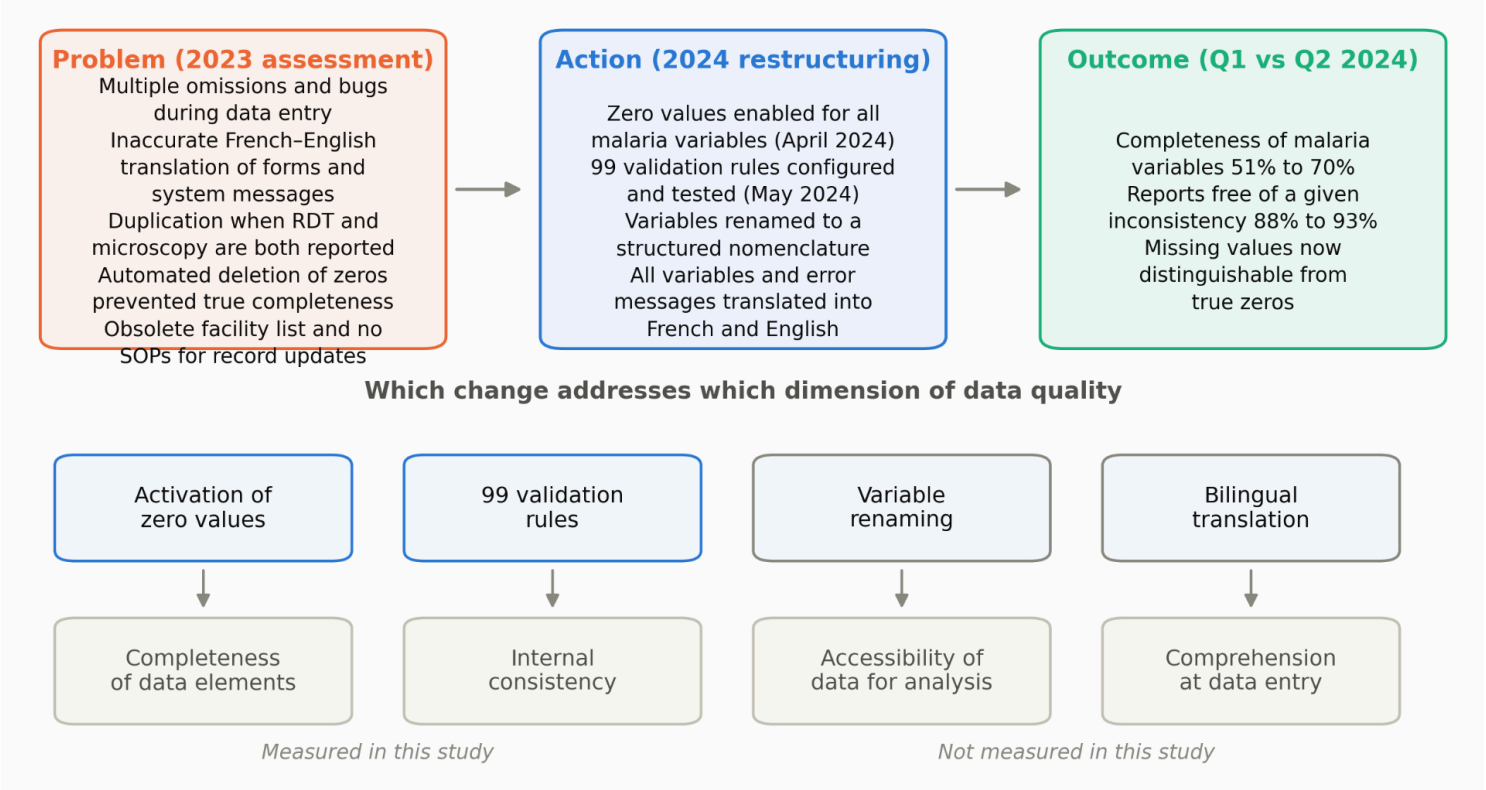


Figure 1

DHIS2 changes made in 2024 and the data quality dimensions they address. Problems identified in the 2023 assessment are shown alongside each change made to the DHIS2 in 2024 and the dimension of data quality that it addresses.

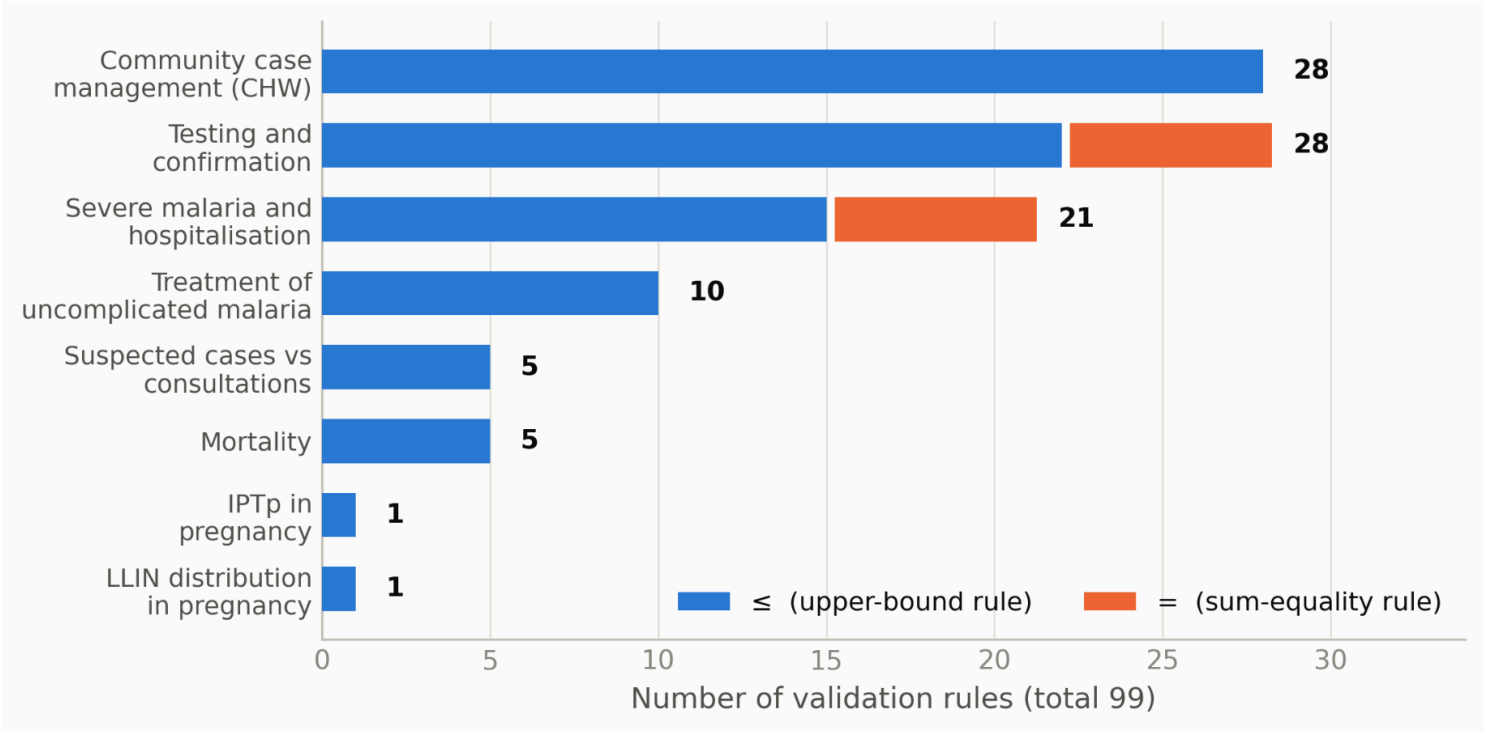


Figure 2

Distribution of the 99 malaria validation rules by clinical domain and comparison operator

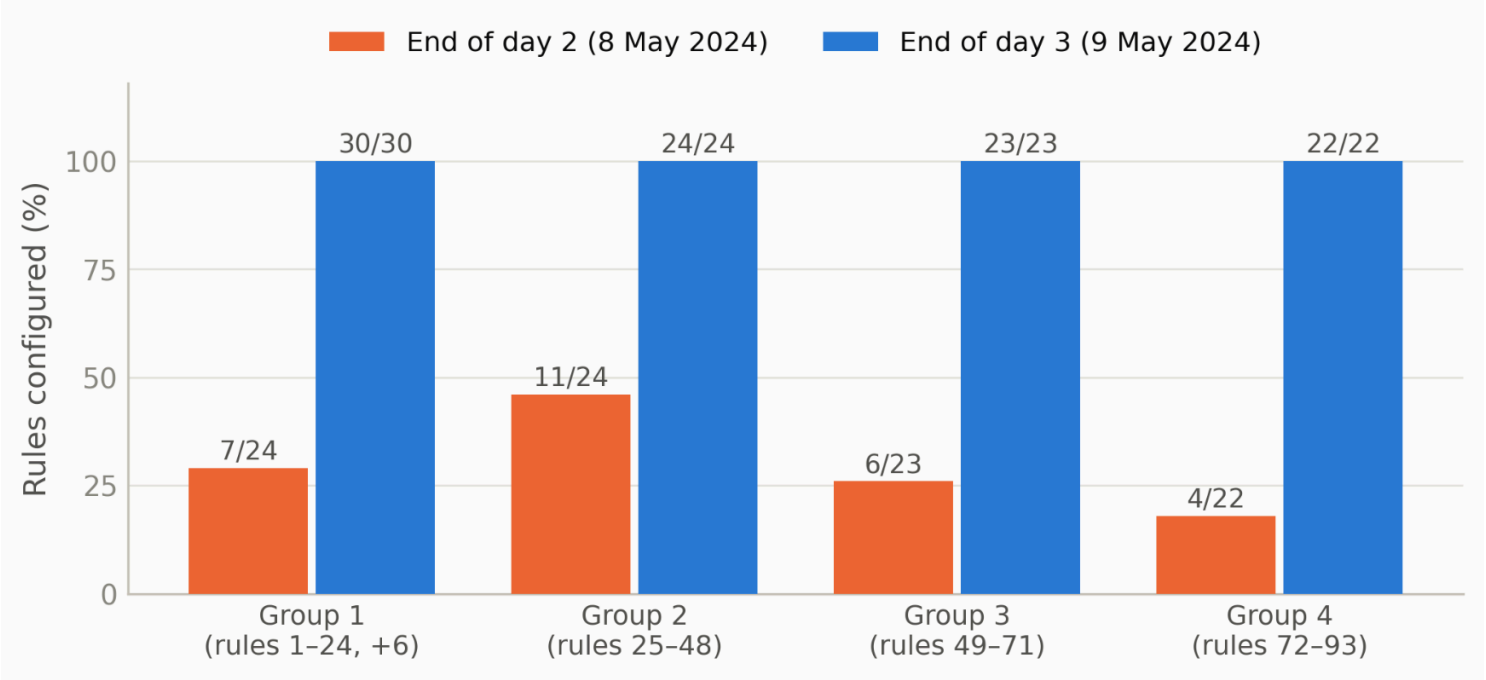


Figure 3

Validation rules configured by each working group, days 2 and 3. Proportion of rules configured by each working group at the end of day 2 and at the end of day 3 of the Mfou workshop, 7–10 May 2024.

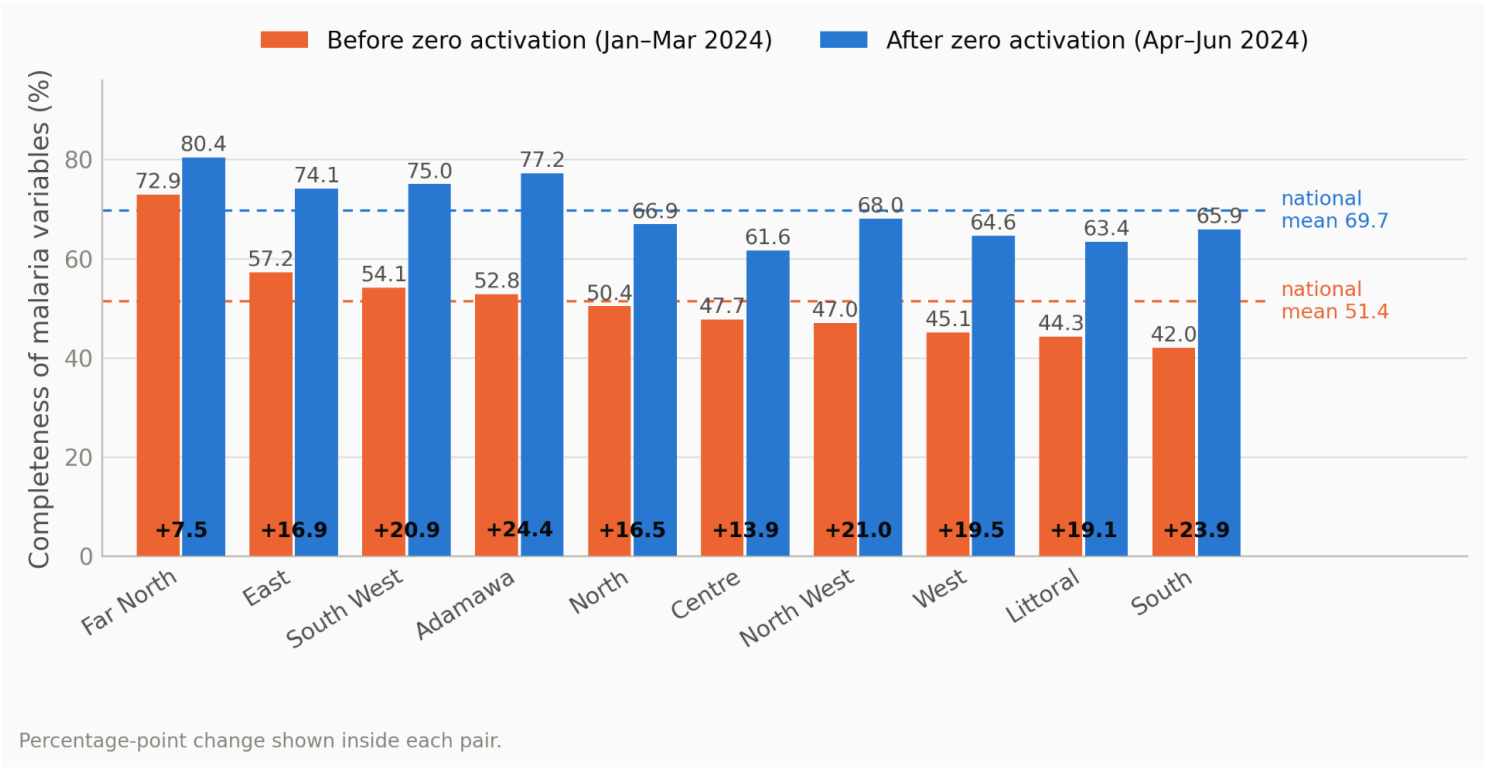


Figure 4

Completeness of malaria variables by region before and after activation of zero values

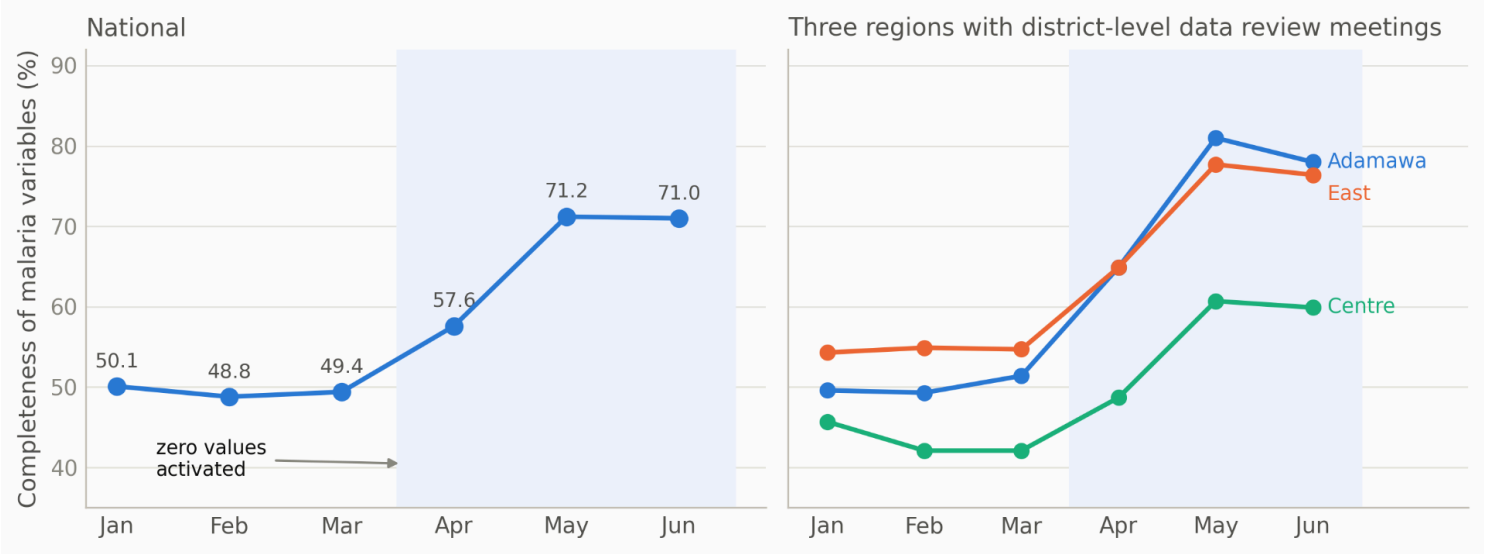
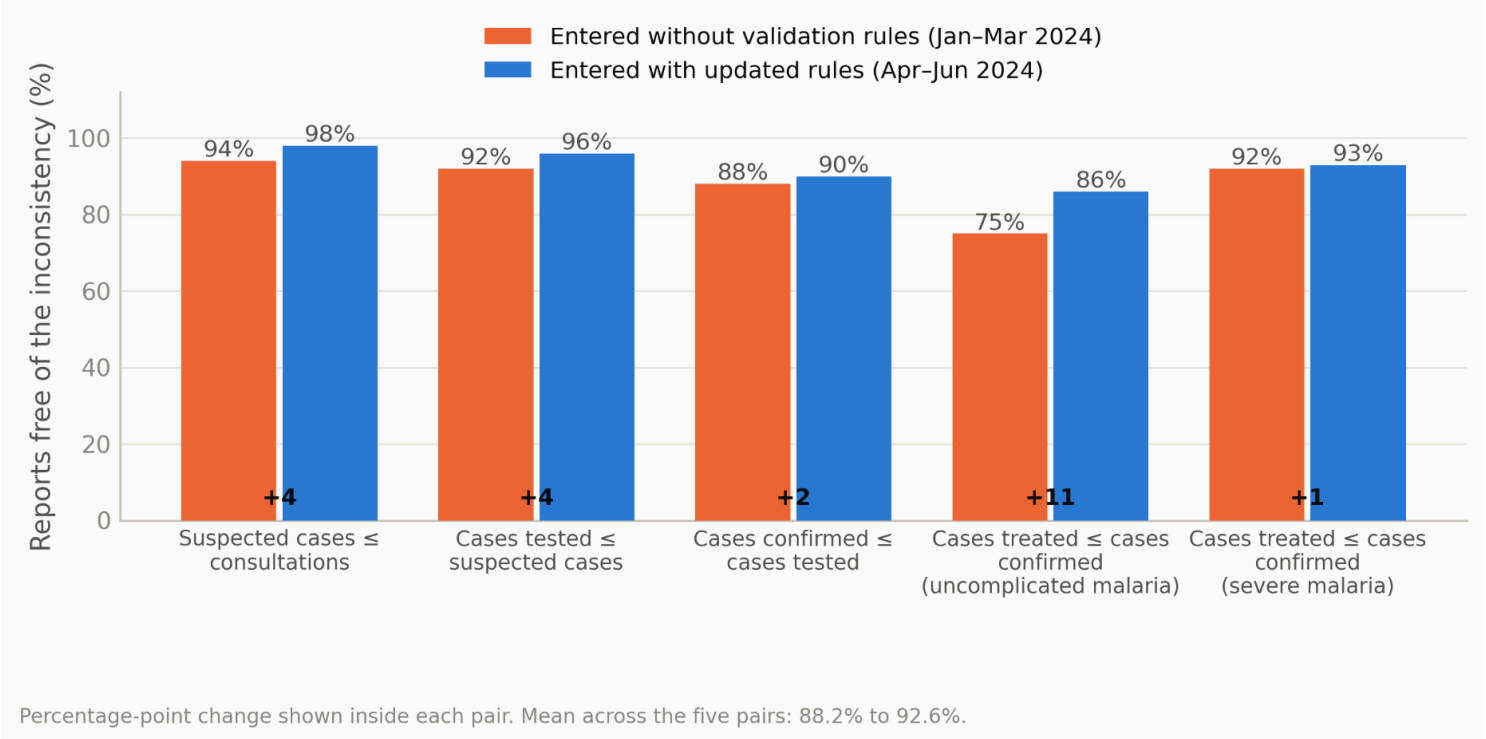


Figure 5

Monthly completeness of malaria variables, January to June 2024, nationally and in three regions



Percentage-point change shown inside each pair. Mean across the five pairs: 88.2% to 92.6%.

Figure 6

Facility reports free of five specified inconsistencies, before and after. Proportion of monthly facility reports free of each of five specified inconsistencies, for reports entered before the updated validation rules (January–March 2024) and after them (April–June 2024).