International Journal of Engineering Research and Modern Education (IJERME)
International Peer Reviewed - Refereed Research Journal, Website: www.rdmodernresearch.com
Impact Factor: 7.018, ISSN (Online): 2455 - 4200, Volume 9, Issue 2, July - December, 2024

ASSESSING ASSOCIATIONS BETWEEN ANTIMALARIAL TREATMENT REGIMENS, PARASITE SPECIES, REGIONAL SETTINGS, AND REPORTED MEDIAN TREATMENT FAILURE

Kannan Vishwanath* & Lloyd Zulu**
* Garwood Innovation Fellow, Garwood Center for Corporate Innovation,
Haas School of Business, Berkeley
** World Christian University, Lusaka, Zambia

Cite This Article: Kannan Vishwanath & Lloyd Zulu, "Assessing Associations Between Antimalarial Treatment Regimens, Parasite Species, Regional Settings, and Reported Median Treatment Failure", International Journal of Engineering Research and Modern Education, Volume 9, Issue 2, July - December, Page Number 88-96, 2024.
Copy Right: © R&D Modern Research Publication, 2024 (All Rights Reserved). This is an Open Access Article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract:

This analysis examines how aggregation and evidence composition shape international comparisons of reported antimalarial treatment failure. Using 244 World Health Organization summaries from 69 countries and territories covering 2010 to 2019, descriptive comparisons assess treatment regimens, parasite species and geographical settings, supplemented by exploratory sample restrictions. The mean reported failure median was 1.87 percent, while 62.30 percent of summaries had zero medians. However, 33.55 percent of zero median summaries contained positive failure maxima, revealing heterogeneity concealed by central values. The ordering of two commonly used regimens reversed across seven country pairs aligned by species, follow up duration and reported study interval bounds. Regional differences persisted within a shared species and treatment category, although means and medians produced different geographical orderings. Restricting observations to summaries representing multiple studies reduced dispersion while preserving a zero median. These findings refine surveillance interpretation by distinguishing favourable central performance from consistency across contributing evidence. International reporting should present distributions, study coverage and comparable analytical groups together to support targeted investigation of elevated failure observations.

Key Words: Aggregation Bias; Antimalarial Efficacy; Comparative Surveillance; Descriptive Analysis; Treatment Failure

1. Introduction:

Antimalarial surveillance can report a favourable median while concealing treatment failure among the studies contributing to that summary. This distinction matters because surveillance evidence informs decisions about where to investigate declining efficacy and whether treatment performance remains consistent across settings. A low central value describes the middle of an observed distribution; it does not establish uniform therapeutic success. The World Health Organization, 2020, reports treatment failure summaries across countries, parasite species and regimens, providing an international evidence base in which central outcomes can be examined alongside their dispersion. Therapeutic response arises from interactions between antimalarial exposure and parasite susceptibility. Regimens differ in their pharmacological properties, while parasite populations differ in their response to the administered compounds. White, 2013, explains how pharmacokinetic and pharmacodynamic relationships shape parasite elimination and the interpretation of treatment performance. These mechanisms justify comparing regimens, but they do not establish that a medicine has a fixed advantage across all populations. Differences between published outcomes can also reflect the countries, observation intervals and protocols represented in each treatment category. A defensible comparison must therefore distinguish therapeutic variation from variation in the composition of the evidence.

The closest clinical synthesis demonstrates both the value and the limits of aggregated efficacy evidence. Marwa et al., 2022, reviewed 52 studies conducted from 2010 to 2020 and reported high corrected cure rates for three widely used artemisinin combination therapies in sub-Saharan Africa. Their analysis establishes favourable pooled treatment performance within its clinical and geographical scope. It does not determine whether the ordering of country level failure summaries remains stable when the compared evidence shares species, follow up duration and study interval bounds. Nor does a pooled cure estimate reveal how frequently a zero-country level median coexists with positive failure outcomes among contributing studies. These are distinct questions about the interpretation of surveillance summaries. Outcome comparability also depends on the biological meaning and classification of recurrence. Jones et al., 2019, demonstrates that molecular correction procedures can influence estimated failure rates by changing the distinction between recrudescence and reinfection. Baird, 2019, further explains why clearance of blood stage infection differs from elimination of latent infection in relapsing malaria. These distinctions make parasite species essential to the interpretation of reported failure. A shared percentage scale does not ensure that every observation represents an equivalent biological event. The present analysis consequently preserves species classifications and examines shared treatment categories without interpreting species contrasts as isolated differences in drug efficacy.

Geographical comparisons introduce an additional layer of heterogeneity. Battle et al., 2014, documents systematic geographical variation in Plasmodium vivax relapse periodicity, demonstrating that location can structure recurrence patterns. However, WHO regions combine countries with different treatment histories, parasite populations and study coverage. Regional averages can therefore reflect both outcome variation and the mix of observations included. The general aggregation problem described by Kievit et al., 2013, provides a relevant interpretive principle: an ordering observed across pooled groups can change when comparisons are restricted. This analysis applies that principle to surveillance evidence without assuming that a regional label measures a specific environmental or institutional mechanism. The reviewed evidence leaves a bounded question unresolved: how much of the apparent variation in published treatment failure summaries remains visible when central values, upper outcomes and comparable categories are examined together? This analysis addresses that question using 244 WHO summaries from 69 countries and territories, covering observations from 2010 to 2019. Each record represents a country, parasite species and regimen combination over a reported study interval. The descriptive design examines distributions across 16 treatment

categories, five species and five WHO regions. Exploratory restrictions align selected comparisons by species, regimen, follow up duration and study interval bounds. This approach makes the composition of the evidence explicit while retaining the distinction between summary statistics and patient failure rates.

This study aims to determine how reported median treatment failure varies across antimalarial regimens, parasite species and geographical settings, and how these comparisons depend on aggregation and sample alignment. The objectives are to describe regimen differences within comparable species and follow up categories, examine species contrasts within shared treatment categories, and assess regional patterns after fixing species, regimen and follow up duration. The completed analysis contributes by showing that zero medians can conceal positive failure outcomes, that regimen ordering can reverse under tighter sample alignment, and that regional means and medians can support different interpretations. Its practical relevance lies in improving how international surveillance evidence is compared: treatment summaries should be interpreted with their observed ranges, contributing study counts and category coverage, rather than through central values alone.

2. Theory, Conceptual Framework and Research Questions Development:

This analysis adopts a pharmacological and evolutionary framework to explain variation in reported antimalarial treatment failure. Treatment outcomes emerge from the interaction between drug activity, parasite susceptibility and host clearance processes. Medicines reduce parasite biomass, while host responses contribute to the removal of infected cells. White, 2017, explains that parasite clearance reflects both drug action and physiological processes, so an observed therapeutic response cannot be attributed to the medicine alone. This framework directs attention toward treatment regimen, parasite species and regional setting as complementary dimensions for organizing observed outcomes. The conceptual framework links these three observed categorical dimensions to the reported median treatment failure rate. Treatment regimen identifies the administered medicine or combination. Parasite species identifies the organism associated with the treated infection. Regional setting identifies the geographical grouping used by the World Health Organization. The outcome represents the median of study failure percentages reported for each country, species and treatment combination. The analysis uses 244 published summaries covering observations from 2010 to 2019 across 69 countries and territories. These summaries provide the basis for descriptive comparisons within the broader eligibility period of 2010 to 2023. World Health Organization, 2020, supplies the corresponding treatment categories, geographical classifications and outcome summaries.

Treatment regimen captures differences in the pharmacological approach used to eliminate infection. Combination therapies bring together compounds with complementary activity, allowing one component to reduce parasite biomass while another clears remaining susceptible parasites. Evolutionary theory also identifies a competing mechanism: exposure to medicines can favour parasites that survive treatment, weakening the effectiveness of a previously successful regimen. Haldar, Bhattacharjee, and Safeukui, 2018, describe how resistance mechanisms differ across antimalarial compounds and parasite populations. Consequently, a regimen has no universally fixed relationship with failure across all settings. Its observed performance depends partly on the susceptibility of the parasites encountered. Empirical evidence supports examining regimen differences without imposing a universal ranking. Marwa et al., 2022, found high corrected cure rates for three widely used artemisinin combination therapies in sub-Saharan Africa, alongside differences between corrected and uncorrected outcomes. This evidence indicates that several regimens can retain substantial efficacy while outcome definitions influence apparent performance. The analysis therefore describes regimen patterns within parasite species and reported follow up categories. It examines the distribution of published medians rather than treating aggregated summaries as direct comparisons between randomly assigned treatments.

RQ1: How do reported median treatment failure rates vary across antimalarial regimens within parasite species and comparable follow up categories?

Parasite species introduces a biological distinction that changes the interpretation of recurrence after treatment. In Plasmodium vivax, dormant liver stages can reactivate after the initial blood infection has cleared. A recurrence can therefore arise through relapse rather than persistence of the original blood infection. White, 2011, explains how relapse periodicity depends on parasite biology and the persistence of antimalarial activity. This mechanism prevents a simple directional expectation that one species must consistently show higher treatment failure than another. Differences in reported outcomes may reflect distinct recurrence processes as well as differences in therapeutic response. The analysis accordingly treats species as an observed classification rather than an ordered measure of treatment difficulty. Comparisons describe the central tendency, spread and frequency of zero reported medians within each species. Where treatment categories overlap, the framework supports examining whether the observed distributions remain similar across species. These comparisons preserve the distinction between a published failure summary and the biological process producing recurrence.

RQ2: What differences in the distribution of reported median treatment failure rates appear across parasite species, including within shared treatment categories?

Regional setting locates therapeutic outcomes within geographically differentiated parasite populations and treatment environments. Regional patterns can emerge when neighbouring populations share transmission conditions, parasite lineages or treatment practices. However, geographical proximity does not ensure uniform outcomes. Battle et al., 2014, demonstrated geographical variation in Plasmodium vivax relapse periodicity, supporting the relevance of location while showing that recurrence patterns require biologically informed interpretation. Their findings justify geographical description without assuming that an administrative region represents a single epidemiological environment. The analysis uses the five observed WHO regional categories to describe geographical variation in reported medians. Comparisons retain species and regimen distinctions because regional differences can reflect the composition of available treatment summaries. The framework therefore examines whether regional patterns persist within shared species and treatment categories. Regional membership remains a geographical descriptor; it does not constitute a measurement of temperature, rainfall or institutional performance.

RQ3: How do reported median treatment failure rates vary across WHO regions within shared parasite species and treatment categories?

Together, the research questions establish a descriptive framework aligned with the available observations. Regimen, species and region remain separate categorical dimensions, while reported median treatment failure remains the single outcome.

The conceptual relationships represent comparisons across observed categories. Each question can be resolved through distributions, ranges and category specific summaries without requiring a predetermined direction or a claim of causation.

3. Data:

This analysis documents the provenance, sample construction and measurement of published antimalarial treatment failure summaries.It combines an audited analytical file with descriptive comparisons across treatment regimens, parasite species and geographical settings.

3.1 Data Sources, Population, and Sample Construction:

The numerical source is Annex 2 of the World Health Organization report on antimalarial drug efficacy, resistance and response, published in 2020. The source identifier is ISBN 9789240012813. Records originate from printed pages 52 to 62, corresponding to PDF pages 64 to 74. The accompanying workbook retains the source URL, page locations, original extracted text and retrieval metadata. The report permits attributed noncommercial reuse and adaptation under the Creative Commons Attribution Non Commercial Share Alike 3.0 IGO licence. World Health Organization, 2020, provides the published treatment failure summaries used throughout the analysis. The unit of analysis is one country or territory, parasite species and treatment regimen summary. Each observation describes the distribution of failure percentages across one or more therapeutic efficacy studies conducted during a reported interval. The dataset forms a cross section of aggregated historical summaries. Observation intervals extend from 2010 to 2019 within the eligibility window of 2010 to 2023. They do not represent annual observations. The outcome is a median across study results rather than a flow, stock or patient population average. The target population comprises antimalarial therapeutic efficacy evidence across geographical settings. The accessible sampling frame comprises the numerical summaries published in Annex 2. Selection retained every available row in that annex. The resulting sample contains 244 observations from 69 countries and territories, five parasite species, 16 treatment categories and five WHO regions. The sample construction rules were documented during workbook preparation; the audit confirmed their application.

Table 1: Sample construction

Stage	Countries or Territories	Observations	Reason or Rule
Accessible published frame	69	244	All numerical summaries in Annex 2
Duplicate exclusions	0 removed	0 removed	Unique country, species and regimen combinations
Missing analytical field exclusions	0 removed	0 removed	Complete outcome, regimen, species, region and study interval
Harmonized analytical sample	69	244	Labels standardized while original source entries were retained
Final descriptive sample	69	244	All observations retained for distributional summaries

The empirical setting is therapeutic efficacy surveillance. Treatment guidance, terminology and global surveillance reports establish the clinical context without supplying additional numerical observations. World Health Organization, 2015, distinguishes therapeutic treatment recommendations; World Health Organization, 2021, standardizes malaria terminology; and World Health Organization, 2023, situates efficacy monitoring within international malaria surveillance. Marwa et al., 2022, further demonstrates the relevance of distinguishing corrected from uncorrected treatment outcomes when comparing published evidence.

3.2 Data Integration, Cleaning, and Quality Assurance:

Integration linked the analytical and source extraction worksheets through 244 unique record identifiers. The audit confirmed matching country labels, species, study intervals, study counts, failure statistics, page references and source URLs across all corresponding rows. No unmatched identifiers, duplicate analytical keys or conflicting numerical entries were identified. Different regimens or species within the same country represent distinct observations. Four regimen entries required label harmonization. Two explicit 28-day annotations were separated from regimen names and retained in the follow up field. One missing combination separator was restored, and one redundant space was removed. These changes preserved treatment identity and outcome values. The original labels remain available in the source extraction worksheet.

Every analytical field was complete across the retained sample. No deletion, imputation, interpolation or external matching was applied. The audit checked that study intervals remained within the eligibility window and that reported percentages satisfied the ordered relationship from minimum through lower quartile, median, upper quartile and maximum. All 244 observations satisfied these checks. Each calculated failure proportion also matched its source percentage divided by 100. Valid extreme observations were retained without trimming or winsorization. The cleaning process preserved published outcome classifications. Felger et al., 2020, explains why distinguishing recrudescence from reinfection matters for efficacy measurement, while Jones et al., 2019, shows that classification algorithms can change estimated failure rates. Plucinski et al., 2015, and Jones et al., 2020, demonstrate the dependence of recurrence classification on genotyping procedures. Flegg et al., 2011, separately distinguishes parasite clearance measurement from later treatment outcomes. These sources clarify the interpretation of retained measurements; their algorithms were not applied to the aggregated records. The file preserves historical summaries as published in the fixed source release. Study intervals were not expanded into annual records, and observations were not assigned to years without corresponding evidence. Completeness therefore describes the published analytical frame, rather than comprehensive surveillance coverage across all countries and years.

3.3 Variable Measurement, Construction, and Validation:

The outcome is the published median treatment failure percentage for each country, species and regimen combination. For analytical storage, the workbook converts this percentage into a proportion using Equation 1.

$$Y_i = \frac{P_i}{100}$$

Here, Y_i denotes the failure proportion for summary i, and $P_i$denotes its published median failure percentage. Higher values indicate higher reported failure. The transformation preserves zero values and does not change the ordering of observations. Treatment regimen, parasite species and WHO region are observed categorical dimensions. Their numerical codes identify categories without imposing order or distance. Study interval, study count and follow up duration describe coverage and protocol characteristics. Minimum, maximum and quartile values retain information about variation among the studies represented by each summary.

Table 2: Variable dictionary

Variable or Indicator	Analytical Role	Definition and Construction	Source and Identifier	Unit and Frequency	Transformation and Direction	Missing Value Rule	Coverage and Limitation
Median treatment failure	Outcome	Published median of study failure percentages within each source summary	WHO2020, Annex 2	Percent per historical summary	Divided by 100 for proportional storage; higher means greater reported failure	None missing; no replacement	244 observations; not a pooled patient failure rate
Treatment regimen	Categorical grouping dimension	Reported medicine or combination, with documented label harmonization	WHO2020, treatment label	Category per summary	Nominal code; no numerical direction	None missing	16 categories with unequal representation
Parasite species	Categorical grouping dimension	Reported Plasmodium species	WHO2020, species heading	Category per summary	Nominal code; no numerical direction	None missing	Five species with different recurrence processes
WHO region	Geographical grouping dimension	Regional classification printed in the source	WHO2020, regional heading	Category per summary	Nominal code; no numerical direction	None missing	Five regions; not a climatic measure
Study interval	Coverage metadata	First and last study years reported for the summary	WHO2020, study years	Calendar year bounds	Retained without annual expansion	None missing	Intervals fall within 2010 to 2019
Study count	Evidence coverage metadata	Number of studies represented by the source row	WHO2020, number of studies	Count per summary	Untransformed; higher means more contributing studies	None missing	Not a patient denominator
Follow up duration	Protocol classification	Duration assigned from source footnotes and explicit annotations	WHO2020, Annex 2 footnotes	Days per summary	Categories of 28 and 42 days	None missing	Protocol classification rather than individual observation time
Minimum, maximum and quartiles	Distributional metadata	Published distribution of contributing study failure percentages	WHO2020, distribution columns	Percent per summary	Retained as published	None missing	Describe study variation within each summary

Measurement interpretation follows the biological distinction between clearance, recrudescence and relapse. White, 2017, describes parasite clearance as a process involving drug activity and host responses. White, 2011, explains how dormant liver stages generate recurrent Plasmodium vivax infection. Battle et al., 2014, documents geographical variation in relapse periodicity. These distinctions require species specific interpretation of the recorded outcome. Reported failure is not itself a direct measurement of resistance. Haldar et al., 2018, describes mechanisms that establish resistance at the parasite level, while Price et al., 2014, identifies the difficulty of separating resistance from recurrence in Plasmodium vivax efficacy evidence. Accordingly, the analysis retains the source outcome label and validates transcription, scale and internal consistency.

Table 3: Descriptive statistics for numerical measures

Measure	Valid Observations	Mean	Standard Deviation	Median	Minimum	Maximum
Reported median failure, percent	244	1.87	5.39	0.00	0.00	61.90
Reported median failure, proportion	244	0.0187	0.0539	0.0000	0.0000	0.6190
Contributing study count	244	5.35	7.24	3.00	1.00	56.00

The outcome distribution contains 152 zero medians, representing 62.30 percent of summaries. A zero median does not establish that every contributing study recorded zero failure. Species coverage comprises 153 Plasmodium falciparum, 62 Plasmodium vivax, 14 Plasmodium malariae, 12 Plasmodium ovale and three Plasmodium knowlesi summaries. Regional coverage comprises 124 African, 46 Western Pacific, 32 South East Asian, 24 Eastern Mediterranean and 18 American summaries. Follow up classifications comprise 198 observations at 28 days and 46 at 42 days.

3.4 Analytical Framework, Identification, and Reproducibility:

The implemented analysis is descriptive. Its object is the distribution of published summary medians across observed treatment, species and regional categories. Counts establish evidence coverage, while means, standard deviations, medians and ranges describe outcome magnitude and dispersion. Each published summary receives equal weight. Study counts describe evidence coverage and do not substitute for patient denominators. This formulation separates the analytical objective from a causal treatment comparison. Hernán, 2018, emphasizes that the intended estimand must be explicit because descriptive associations and causal effects require different analytical assumptions. The present summaries characterize the accessible published evidence without identifying the effect of assigning a treatment or changing regional conditions.

Reporting prioritizes numerical magnitude and distributional shape. Cumming, 2014, supports framing research questions around quantitative estimates. Greenland et al., 2016, distinguishes statistical evidence from claims about scientific hypotheses. Wasserstein and Lazar, 2016, further explains that statistical significance does not measure effect size or practical importance, while Wasserstein et al., 2019, advocates interpretation beyond significance thresholds. The descriptive outputs therefore report observed patterns without assigning significance labels. Reproducibility rests on the workbook's analytical data, variable dictionary and source extraction worksheets. Record identifiers connect each observation to its original text, source page and URL. Formula checks reproduce the percentage conversion, while category codes preserve treatment, species and regional

classifications. Consistent units and retained study intervals align the descriptive outputs with the published evidence. Unequal category coverage and differences in recurrence processes define the interpretation of comparisons across groups.

4. Methods:

4.1 Research Design and Analytical Scope:

This analysis uses a descriptive comparative design to examine published antimalarial treatment failure summaries. The analytical unit is a country or territory, parasite species and treatment regimen combination over a reported study interval. The objective is to characterize the distribution of reported failure medians and determine how comparisons change when observations share selected clinical and geographical characteristics. The design describes the accessible surveillance evidence rather than estimating the consequences of assigning patients to treatment. The useful variation arises across treatment categories, species, regions and contributing study intervals. Comparisons proceed within increasingly specific groups to distinguish broad distributional patterns from differences associated with evidence composition. This distinction follows Pollet et al., 2015, who explains why relationships observed in aggregated data cannot automatically be transferred to individuals. The reported comparisons therefore concern published summaries, with country coverage, protocol differences and recurrence classification defining their interpretive boundaries. The sample construction table and variable dictionary establish the eligible observations and measurement definitions. The full descriptive analysis retains 244 summaries from 69 countries and territories, covering observations from 2010 to 2019. Treatment regimen, species and region remain categorical classifications. Their numerical codes serve as identifiers and are not interpreted as ordered scores.

4.2 Descriptive Specification and Measurement Alignment:

The outcome is the published median failure percentage for each analytical observation. Equation 1 defines its conversion into a proportion. Calculations retain numerical precision, while presentation uses percentages for outcome levels and percentage points for differences and standard deviations. Each published summary receives equal weight because contributing study counts do not provide patient denominators. For each comparison group, the analysis calculates the observation count, arithmetic mean, sample standard deviation, median, minimum and maximum. Equation 2 defines the group mean.

$$\overline{P}_g = \frac{1}{n_g} \sum_{i \in G_g} P_i$$

Here, $\overline{P}_g$ is the mean of published failure medians in group g, G_g is the set of observations assigned to that group, $n_g$is its observation count, and $P_i$is the published failure median expressed as a percentage for observation i. This quantity is an average of source medians, not a pooled patient failure percentage. Equation 3 defines dispersion within each group.

$$s_g = \sqrt{\frac{\sum_{i \in G_g} \left[P_i - \overline{P}_g\right]^2}{n_g - 1}}$$

Here, s_g is the sample standard deviation in percentage points. The remaining symbols retain their definitions from Equation 2. Standard deviations are calculated only for groups containing at least two observations. For an even group size, the median is the arithmetic average of the two central ordered values. The analysis also counts zero medians and identifies summaries whose median equals zero while their reported maximum exceeds zero. This comparison assesses whether a favourable central value conceals positive failure outcomes among contributing studies. It preserves the distinction between variation across published summaries and variation within the studies represented by each summary.

4.3 Comparative Procedures and Exploratory Sensitivity Analyses:

The analytical sequence begins with full sample distributions and category coverage. Treatment comparisons then restrict observations to Plasmodium falciparum and separate the recorded 28 day and 42 day follow up categories. The principal presentation covers artemether plus lumefantrine and artesunate plus amodiaquine at 28 days, together with artesunate plus mefloquine, artesunate plus pyronaridine and dihydroartemisinin plus piperaquine at 42 days. These groups retain the source classifications rather than combining outcomes observed over different follow up periods. Species comparisons first describe all five species. A shared treatment comparison then retains artemether plus lumefantrine observations with 28 day follow up. Regional comparisons similarly begin with the complete dataset before restricting observations to Plasmodium falciparum treated with artemether plus lumefantrine at 28 days. These restrictions align selected recorded characteristics without implying equivalence in patient populations or study procedures.

An exploratory comparison aligns artemether plus lumefantrine and artesunate plus amodiaquine summaries by country, species, follow up duration and both reported study interval bounds. Seven countries satisfy these conditions. Each country contributes one summary for each regimen. The analysis calculates the mean for each treatment within this common set and counts countries in which the first treatment has a higher, lower or equal reported median. Identical interval bounds establish temporal alignment at the summary level, not patient pairing or common treatment assignment. This restriction examines whether descriptive ordering changes with sample composition. Kievit et al., 2013, establishes the relevance of comparing aggregated and restricted relationships when group composition can alter their direction. The comparison is explicitly exploratory and does not constitute a clinical superiority test. A second exploratory sensitivity analysis retains summaries representing at least two contributing studies. It compares the resulting 177 observations with the complete sample using the same descriptive statistics. This restriction assesses sensitivity to summaries supported by one study. It changes the evidence base rather than the outcome definition, and excluded observations remain part of the principal analysis.

4.4 Computational Verification, Transparency, and Interpretation:

Calculations were implemented in Python version 3.12.14 using the standard libraries zipfile, xml.etree. Element Tree, collections and statistics. Workbook contents were read directly from the underlying spreadsheet files. The statistics. mean, statistics. stdev and statistics. median functions produced the principal numerical summaries. Group membership and exploratory restrictions used exact comparisons of the cleaned categorical fields and reported interval bounds. Verification linked the

analytical and source extraction worksheets through record identifiers. Checks assessed identifier alignment, uniqueness of country, species and regimen combinations, completeness of analytical fields, and agreement between corresponding source and analytical values. Regimen differences were reviewed separately against documented label harmonization. Percentage checks required values between zero and 100 and the ordered sequence of minimum, lower quartile, median, upper quartile and maximum. Conversion checks compared every stored proportion with its source percentage divided by 100. No missing value replacement, interpolation or outcome trimming was applied. Valid extreme observations remained in the full sample. Calculations used unrounded values, with rounding applied only when presenting results. The procedures are deterministic and require no random seed.

The supplied workbook preserves analytical records, measurement definitions, original extracted entries and source locations. These materials support reconstruction of the reported comparisons. The methodological contribution lies in combining distributional description with transparent comparison restrictions, rather than introducing a new estimator. Consistent with Kass et al., 2016, analytical choices follow the scientific question and the information contained in the observations. The resulting evidence characterizes the published surveillance frame without treating descriptive differences as causal treatment effects.

5. Findings:

This analysis examines variation in reported treatment failure across regimens, parasite species and WHO regions. The findings answer the three research questions through descriptive comparisons of published summary medians. Exploratory restrictions assess how sample composition influences the observed patterns. The analytical sample contains 244 country, species and regimen summaries from 69 countries and territories. Observations cover 2010 to 2019 and include 16 treatment categories, five parasite species and five WHO regions. All records contain complete analytical fields, and no observations were removed for missingness or duplicate identifiers. Each observation summarizes one or more therapeutic efficacy studies, following the reporting structure of World Health Organization, 2020. The numerical results below therefore describe published summaries rather than individual patients. Across all observations, the mean reported failure median was 1.87 percent, with a standard deviation of 5.39 percentage points. The median was zero, and values ranged from zero to 61.90 percent. This difference between the mean and median indicates concentration at zero alongside a smaller group of elevated observations. Specifically, 152 summaries, representing 62.30 percent of the sample, reported zero medians.

A zero median did not consistently indicate an absence of failure among contributing studies. Of the 152 summaries with zero medians, 51 had a positive reported maximum. Thus, 33.55 percent of zero median summaries contained at least one contributing study with recorded failure. Central tendency alone therefore concealed variation within a substantial share of apparently favourable summaries. The first research question concerns regimen differences within species and comparable follow up categories. Among Plasmodium falciparum summaries with 28 day follow up, artemether plus lumefantrine had a mean reported median of 1.40 percent, compared with 2.13 percent for artesunate plus amodiaquine. Both distributions had zero medians, but their upper values differed: 8.50 percent for artemether plus lumefantrine and 18.20 percent for artesunate plus amodiaquine. The difference therefore appeared mainly in the positive tail rather than in the central observation.

Treatment Comparisons within Plasmodium Falciparum:

Regimen	Follow Up, Days	Summaries	Mean Failure Median, Percent	Standard Deviation, Percentage Points	Median, Percent	Observed Range, Percent
Artemether plus lumefantrine	28	57	1.40	1.99	0.00	0.00 to 8.50
Artesunate plus amodiaquine	28	31	2.13	3.97	0.00	0.00 to 18.20
Artesunate plus mefloquine	42	7	2.46	3.73	0.00	0.00 to 8.30
Artesunate plus pyronaridine	42	8	1.29	1.88	0.70	0.00 to 5.60
Dihydroartemisinin plus piperaquine	42	31	2.18	6.28	0.00	0.00 to 32.40

Among the 42-day summaries, artesunate plus pyronaridine had the lowest mean, at 1.29 percent, but its median of 0.70 percent exceeded the zero medians of the other two regimens. Dihydroartemisinin plus piperaquine showed the greatest dispersion, with a standard deviation of 6.28 percentage points and a maximum of 32.40 percent. Consequently, the regimen with the lowest median was not necessarily the regimen with the lowest mean or narrowest observed distribution. These generally low central values are consistent with the high corrected cure rates reported by Marwa et al., 2022, although their pooled clinical estimates and the present averages of published medians measure different quantities. An exploratory restriction examined artemether plus lumefantrine and artesunate plus amodiaquine in countries with identical reported study interval bounds, the same species and the same follow up duration. Seven country pairs met these conditions. Within this restricted set, the respective means were 0.86 and 0.46 percent. The descriptive difference therefore changed from 0.73 percentage points lower for artemether plus lumefantrine in the unrestricted comparison to 0.40 percentage points higher in the restricted comparison. Four pairs had equal medians, two had higher values for artemether plus lumefantrine, and one had a higher value for artesunate plus amodiaquine. Matching interval bounds does not establish that the underlying studies enrolled comparable patients, but the reversal demonstrates that the aggregate regimen ordering depends on sample composition. The second research question examines species differences. Plasmodium vivax had the highest mean reported median, at 2.95 percent, compared with 1.76 percent for Plasmodium falciparum. Nevertheless, every species had a zero median. The contrast therefore concerns the frequency and magnitude of positive observations rather than a separation in median outcomes.

Species Distribution of Reported Failure Medians:

Parasite Species	Summaries	Mean, Percent	Standard Deviation, Percentage Points	Median, Percent	Observed Range, Percent	Zero Medians
Plasmodium falciparum	153	1.76	3.76	0.00	0.00 to 32.40	84
Plasmodium vivax	62	2.95	8.81	0.00	0.00 to 61.90	40
Plasmodium malariae	14	0.00	0.00	0.00	0.00 to 0.00	14
Plasmodium ovale	12	0.31	1.07	0.00	0.00 to 3.70	11
Plasmodium knowlesi	3	0.00	0.00	0.00	0.00 to 0.00	3

The species contrast remained visible within a shared treatment and follow up category. For artemether plus lumefantrine at 28 days, the mean was 4.90 percent across nine Plasmodium vivax summaries and 1.40 percent across 57 Plasmodium falciparum summaries. Both medians remained zero. The higher Plasmodium vivax mean therefore reflected elevated observations within a small and unevenly represented category. White, 2011, explains that relapse can generate recurrent Plasmodium vivax infection after clearance of the initial blood infection. Accordingly, the observed contrast does not isolate differences in blood stage drug efficacy. The third research question concerns regional variation within shared species and treatment categories. Across the complete sample, the Western Pacific region had the highest mean reported median, at 4.27 percent, compared with 1.15 percent in Africa, 1.19 percent in the Americas, 1.52 percent in the Eastern Mediterranean and 1.83 percent in South East Asia. All five regional medians were zero. The aggregate regional contrast again arose from differences in positive observations and their magnitude. Restricting the comparison to Plasmodium falciparum treated with artemether plus lumefantrine at 28 days reduced variation attributable to species, regimen and follow up composition. The Western Pacific region retained the highest mean, at 2.56 percent. Africa and South East Asia had similar means of 1.43 and 1.49 percent, while the Eastern Mediterranean mean was 0.28 percent and the three American summaries were all zero.

Regional Comparison within Plasmodium Falciparum Treated with Artemether Plus Lumefantrine at 28 Days:

WHO Region	Summaries	Mean, Percent	Standard Deviation, Percentage Points	Median, Percent	Observed Range, Percent
Africa	35	1.43	1.47	1.20	0.00 to 5.00
Americas	3	0.00	0.00	0.00	0.00 to 0.00
Eastern Mediterranean	5	0.28	0.63	0.00	0.00 to 1.40
South East Asia	7	1.49	3.17	0.00	0.00 to 8.50
Western Pacific	7	2.56	3.29	0.00	0.00 to 7.60

Africa had the only positive median in this restricted comparison, despite having a lower mean than the Western Pacific region. This distinction indicates more consistently positive summaries in Africa and greater concentration of elevated values among fewer Western Pacific observations. Regional differences therefore depend on whether the comparison emphasizes central prevalence or the upper tail. Battle et al., 2014, similarly demonstrates the importance of geographical heterogeneity in malaria recurrence, although the present regional categories do not identify the biological or environmental processes producing these patterns. An additional exploratory sensitivity analysis retained only summaries representing at least two contributing studies. This restriction reduced the sample from 244 to 177 observations. The mean declined from 1.87 to 1.49 percent, the standard deviation declined from 5.39 to 3.18 percentage points, and the maximum declined from 61.90 to 18.80 percent. The median remained zero. Low central values therefore persisted, while the magnitude of the upper tail was sensitive to summaries based on one study.

The three research questions yield a consistent descriptive conclusion: treatment failure summaries are dominated by zero medians, while regimen, species and regional differences emerge principally through positive tails and unequal evidence coverage. Regimen ordering changes under tighter sample alignment, species contrasts remain visible within a shared treatment category, and regional contrasts persist after fixing species, regimen and follow up. These patterns characterize published surveillance evidence. Their interpretation remains sensitive to recurrence classification, which Jones et al., 2019, shows can influence estimated treatment failure rates.

6. Discussion:

A zero median can conceal treatment failure rather than establish its absence. The contrast between the mean and median in Table 3 becomes more informative when considered alongside the positive maxima recorded in one third of summaries with zero medians. This analysis identifies an aggregation problem in the surveillance evidence: favourable central values coexist with elevated outcomes among contributing studies. Equation 1 preserves this distribution because converting percentages into proportions changes scale without recovering information lost through aggregation. The contribution is therefore a distinction between low central failure and consistency across the underlying evidence. Pollet et al., 2015, explains how aggregation can obscure heterogeneity and encourage inappropriate inference across analytical levels. Here, retaining distributional information changes the interpretation of apparently favourable surveillance summaries. Treatment comparisons further show that an apparent regimen advantage depends on the composition of the evidence being compared. The ordering of artemether plus lumefantrine and artesunate plus amodiaquine reversed when the analysis retained countries with identical study interval bounds, species and follow up duration. This exploratory result does not identify therapeutic superiority, but it demonstrates that aggregate ordering is unstable under tighter alignment. Kievit et al., 2013, describes how differences in group composition can produce reversals between aggregated and disaggregated comparisons. The present result illustrates why a surveillance ranking should be treated as

conditional on coverage. Its substantive contribution is to make comparability part of the interpretation of treatment performance, rather than treating every difference between published averages as a property of the medicines.

Species comparisons reveal another boundary to interpretation: a shared treatment category does not ensure a shared biological outcome. The higher mean for Plasmodium vivax remained visible within artemether plus lumefantrine summaries at the same follow up duration, while both species retained zero medians. This pattern locates the contrast in elevated observations rather than a uniform displacement of the outcome distribution. Baird, 2019, distinguishes treatment of latent infection from clearance of blood stage infection, showing why recurrence in relapsing malaria requires separate interpretation. The analysis therefore refines the conceptual framework by treating species as a condition governing the meaning of reported failure. A common numerical scale supports comparison, but biological equivalence cannot be inferred from that scale alone. The sensitivity analysis shows that evidence depth affects the apparent magnitude of exceptional outcomes. Restricting the sample to summaries containing multiple studies reduced the mean and dispersion while leaving the median unchanged. This does not invalidate observations drawn from one study; it distinguishes concentrated evidence from patterns reproduced across a broader source base. Together with the distribution reported in Table 3, the result indicates that surveillance interpretation requires both outcome magnitude and information about the evidence supporting it. Kass et al., 2016, emphasizes matching statistical interpretation to data collection and the scientific objective. The practical implication is specific: surveillance reporting should display contributing study counts and observed ranges alongside medians, so that exceptional outcomes remain visible without receiving an unsupported interpretation as broadly representative.

The international contribution lies in showing that regional differences persist within a shared species, regimen and follow up category, yet change meaning according to the distributional measure selected. Africa showed more consistently positive summaries, whereas the Western Pacific comparison combined a zero median with a higher mean and greater dispersion. These patterns complicate any single ordering of regional treatment performance. White, 2013, grounds therapeutic response in drug exposure and parasite susceptibility; the present comparisons complement that framework by demonstrating how the structure of published evidence shapes its apparent geographical expression. Decision makers should therefore compare like treatment categories, preserve species distinctions and examine elevated observations alongside central values. The theoretical extension is an evidence sensitive interpretation of efficacy surveillance: observed performance reflects both therapeutic outcomes and the way heterogeneous studies are summarized. Future comparisons can assess whether the same ordering persists when patient denominators, study protocols and recurrence definitions are aligned.

7. Conclusion and Implications:

Favourable surveillance summaries can conceal the failures that international malaria control needs to detect. This analysis shows that apparent treatment performance depends on how heterogeneous evidence is grouped, aligned and summarized. Low central outcomes coexist with elevated observations, while tighter comparability can reverse an apparent treatment ordering. The core contribution is an empirical demonstration that favourable central performance and consistency across contributing studies are distinct properties of surveillance evidence. This distinction refines therapeutic interpretation by making aggregation and evidence composition explicit conditions of comparison. Decision makers should therefore assess distributions and coverage before using published summaries to prioritize investigations or compare treatment options. Surveillance programmes can operationalize this approach by reporting central outcomes alongside observed extremes, contributing study counts and clearly defined comparison groups. International reporting guidance should encourage consistent recurrence definitions and transparent documentation of observation periods, enabling comparisons that preserve biological and methodological distinctions. These practices can improve the targeting of confirmatory surveillance without treating exceptional observations as representative of entire populations. Their potential social value lies in reducing misplaced reassurance and supporting better informed protection of effective treatment. The analysis advances a globally transferable principle: surveillance becomes more informative when it communicates where favourable summaries conceal heterogeneous evidence.

The aggregated historical design creates opportunities for more detailed validation using linked study records and patient denominators. Future analyses should examine whether the observed ordering persists after aligning recruitment periods, follow up procedures and recurrence classifications. Longitudinal surveillance could distinguish persistent geographical patterns from changes in evidence coverage. Patient data could assess clinical differences that published medians cannot identify, while linked laboratory measurements could examine the biological processes underlying elevated outcomes. Prospective protocols should specify comparison rules before examining results. These extensions would clarify transferability across treatment environments and establish which descriptive patterns warrant investigation through designs supporting causal inference.

References:

1. Baird, J. K. 2019. 8-Aminoquinoline therapy for latent malaria. Clinical Microbiology Reviews, volume 32, issue 4, Article e00011-19. https://doi.org/10.1128/CMR.00011-19
2. Battle, K. E., Karhunen, M. S., Bhatt, S., Gething, P. W., Howes, R. E., Golding, N., Van Boeckel, T. P., Messina, J. P., Shanks, G. D., Smith, D. L., Baird, J. K., & Hay, S. I. 2014. Geographical variation in Plasmodium vivax relapse. Malaria Journal, volume 13, issue 1, Article 144. https://doi.org/10.1186/1475-2875-13-144
3. Cumming, G. 2014. The new statistics: Why and how. Psychological Science, volume 25, issue 1, pages 7-29. https://doi.org/10.1177/0956797613504966
4. Felger, I., Snounou, G., Hastings, I., Moehrle, J. J., & Beck, H. P. 2020. PCR correction strategies for malaria drug trials: Updates and clarifications. The Lancet Infectious Diseases, volume 20, issue 1, pages e20-e25. https://doi.org/10.1016/S1473-3099(19)30426-8
5. Flegg, J. A., Guerin, P. J., White, N. J., & Stepniewska, K. 2011. Standardizing the measurement of parasite clearance in falciparum malaria: The parasite clearance estimator. Malaria Journal, volume 10, issue 1, Article 339. https://doi.org/10.1186/1475-2875-10-339

6. Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. 2016. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. European Journal of Epidemiology, volume 31, issue 4, pages 337-350. https://doi.org/10.1007/s10654-016-0149-3
7. Haldar, K., Bhattacharjee, S., &Safeukui, I. 2018. Drug resistance in Plasmodium. Nature Reviews Microbiology, volume 16, issue 3, pages 156-170. https://doi.org/10.1038/nrmicro.2017.161
8. Hernán, M. A. 2018. The C-word: Scientific euphemisms do not improve causal inference from observational data. American Journal of Public Health, volume 108, issue 5, pages 616-619. https://doi.org/10.2105/AJPH.2018.304337
9. Jones, S., Kay, K., Hodel, E. M., Chy, S., Mbituyumuremyi, A., Uwimana, A., Menard, D., Felger, I., & Hastings, I. 2019. Improving methods for analyzing antimalarial drug efficacy trials: Molecular correction based on length-polymorphic markers msp-1, msp-2, and glurp. Antimicrobial Agents and Chemotherapy, volume 63, issue 9, Article e00590-19. https://doi.org/10.1128/AAC.00590-19
10. Jones, S., Plucinski, M., Kay, K., Hodel, E. M., & Hastings, I. M. 2020. A computer modelling approach to evaluate the accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of antimalarial drug efficacy. Antimicrobial Agents and Chemotherapy, volume 64, issue 4, Article e01517-19. https://doi.org/10.1128/AAC.01517-19
11. Kass, R. E., Caffo, B. S., Davidian, M., Meng, X. L., Yu, B., & Reid, N. 2016. Ten simple rules for effective statistical practice. PLOS Computational Biology, volume 12, issue 6, Article e1004961. https://doi.org/10.1371/journal.pcbi.1004961
12. Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. 2013. Simpson's paradox in psychological science: A practical guide. Frontiers in Psychology, volume 4, Article 513. https://doi.org/10.3389/fpsyg.2013.00513
13. Marwa, K., Kapesa, A., Baraka, V., Konje, E., Kidenya, B., Mukonzo, J., Kamugisha, E., & Swedberg, G. 2022. Therapeutic efficacy of artemether-lumefantrine, artesunate-amodiaquine and dihydroartemisinin-piperaquine in the treatment of uncomplicated Plasmodium falciparum malaria in Sub-Saharan Africa: A systematic review and meta-analysis. PLOS ONE, volume 17, issue 3, Article e0264339. https://doi.org/10.1371/journal.pone.0264339
14. Plucinski, M. M., Morton, L., Bushman, M., Dimbu, P. R., & Udhayakumar, V. 2015. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrobial Agents and Chemotherapy, volume 59, issue 10, pages 6096-6100. https://doi.org/10.1128/AAC.00072-15
15. Pollet, T. V., Stulp, G., Henzi, S. P., & Barrett, L. 2015. Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. American Journal of Primatology, volume 77, issue 7, pages 727-740. https://doi.org/10.1002/ajp.22405
16. Price, R. N., von Seidlein, L., Valecha, N., Nosten, F., Baird, J. K., & White, N. J. 2014. Global extent of chloroquine-resistant Plasmodium vivax: A systematic review and meta-analysis. The Lancet Infectious Diseases, volume 14, issue 10, pages 982-991. https://doi.org/10.1016/S1473-3099(14)70855-2
17. Wasserstein, R. L., & Lazar, N. A. 2016. The ASA statement on p-values: Context, process, and purpose. The American Statistician, volume 70, issue 2, pages 129-133. https://doi.org/10.1080/00031305.2016.1154108
18. Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. 2019. Moving to a world beyond $p < 0.05$. The American Statistician, volume 73, supplement 1, pages 1-19. https://doi.org/10.1080/00031305.2019.1583913
19. White, N. J. 2011. Determinants of relapse periodicity in Plasmodium vivax malaria. Malaria Journal, volume 10, issue 1, Article 297. https://doi.org/10.1186/1475-2875-10-297
20. White, N. J. 2013. Pharmacokinetic and pharmacodynamic considerations in antimalarial dose optimization. Antimicrobial Agents and Chemotherapy, volume 57, issue 12, pages 5792-5807. https://doi.org/10.1128/AAC.00287-13
21. White, N. J. 2017. Malaria parasite clearance. Malaria Journal, volume 16, issue 1, Article 88. https://doi.org/10.1186/s12936-017-1731-1
22. World Health Organization. 2015. Guidelines for the treatment of malaria. Third edition. https://www.afro.who.int/sites/default/files/2017-06/9789241549127_eng.pdf
23. World Health Organization. 2020. Report on antimalarial drug efficacy, resistance and response: 10 years of surveillance, 2010-2019. https://cdn.who.int/media/docs/default-source/malaria/9789240012813-eng.pdf
24. World Health Organization. 2021. WHO malaria terminology, 2021 update. https://www.who.int/publications/i/item/9789240038400
25. World Health Organization. 2023. World malaria report 2023. https://www.who.int/publications/i/item/9789240086173