

May 23, 2023

Documenting the evidence of routine DQAs on malaria data reporting accuracy in Zambia, 2015-2021

SMERG 2023



Why does understanding rDQAs and data reporting accuracy matter?

- No standard approach to measure impact of rDQAs over time
- Recognizing when data is of sufficient quality to guide decision making is critical
 - Different data quality cut-offs/thresholds might be needed for different decisions (e.g., outbreak response, subnational tailoring of interventions, implementing trainings and supervision)
- rDQAs are resource intensive!
- This work aimed to improve efficiencies in the rDQA process, and establish a simpler method for assessing trends in data reporting accuracy at the district level

Study objective and methods

Overall Objective: Evaluate trends in malaria data reporting accuracy using routine DQAs in Zambia from 2015 to 2021 to inform DQA practices and data quality improvement activities

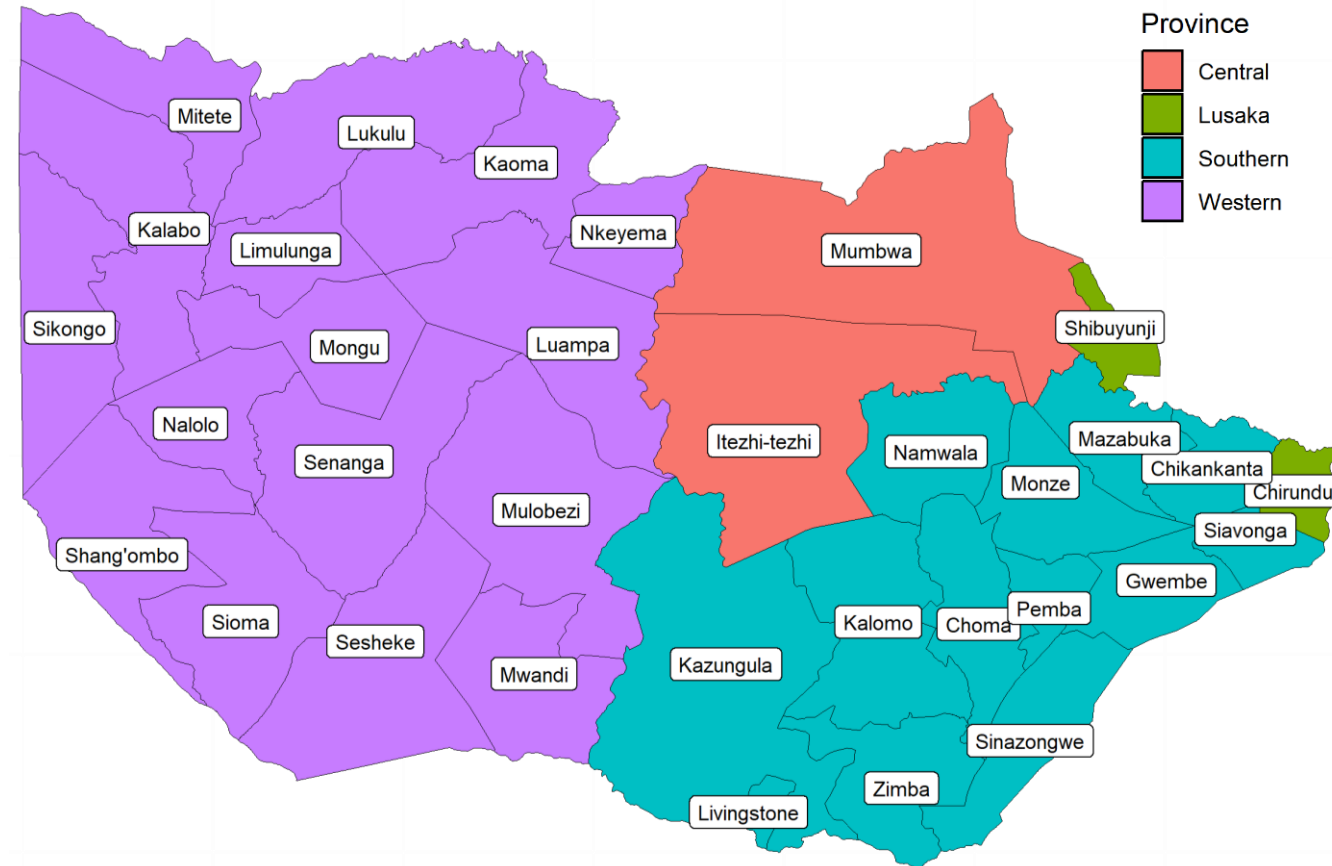
Methods and Analysis:

1. Weighted Average Percentage Error (WAPE) method was selected as the best method for measuring data reporting accuracy
2. Trends in accuracy over time for individual data elements (Total OPD attendance, RDT positive cases, RDT tests) and aggregate
3. Stratification of data reporting accuracy
 - a. Number of DQAs conducted at each health facility
 - b. Malaria incidence
4. Investigate if health facilities would “shift” the subnational tailoring (SNT) intervention response if reported data were accurate

For this study, we looked at DQA results in Zambia from 2015-2021 (aka – a very big and rich dataset!)

- Districts participating in DQAs expanded from 14 districts in 2015 to 33 districts in 2021, primarily in Southern and Western Provinces
- Health facilities are selected based on their previous DQA performance and other data quality indicators that are tracked each month via dashboard
- 155 facilities were audited in 2015. By 2021, at least 1 DQA had been performed at 645 health facilities
- Data are from DQAs of Malaria Rapid Reporting (MRR) system and 6 months of data are audited

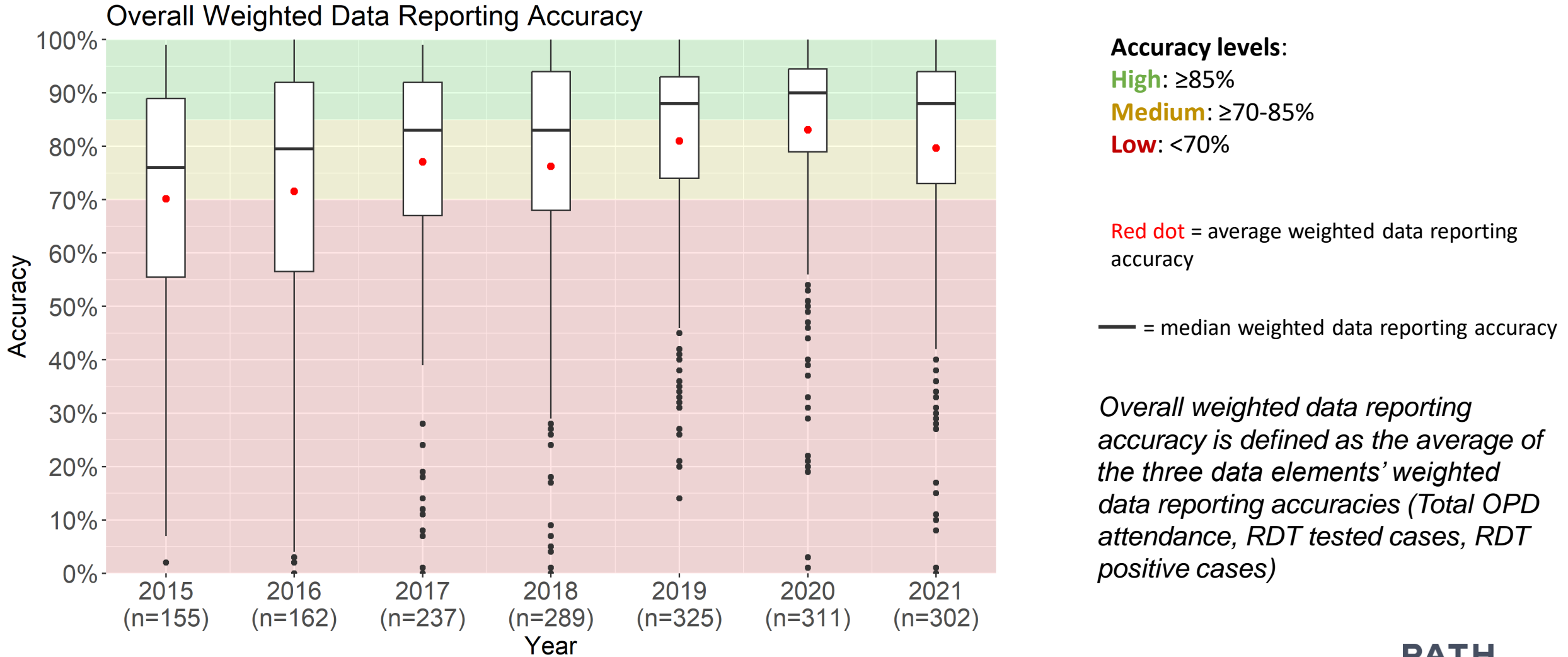
Participating Districts for DQAs, Zambia 2015-2021



Overview of rDQA in Zambia

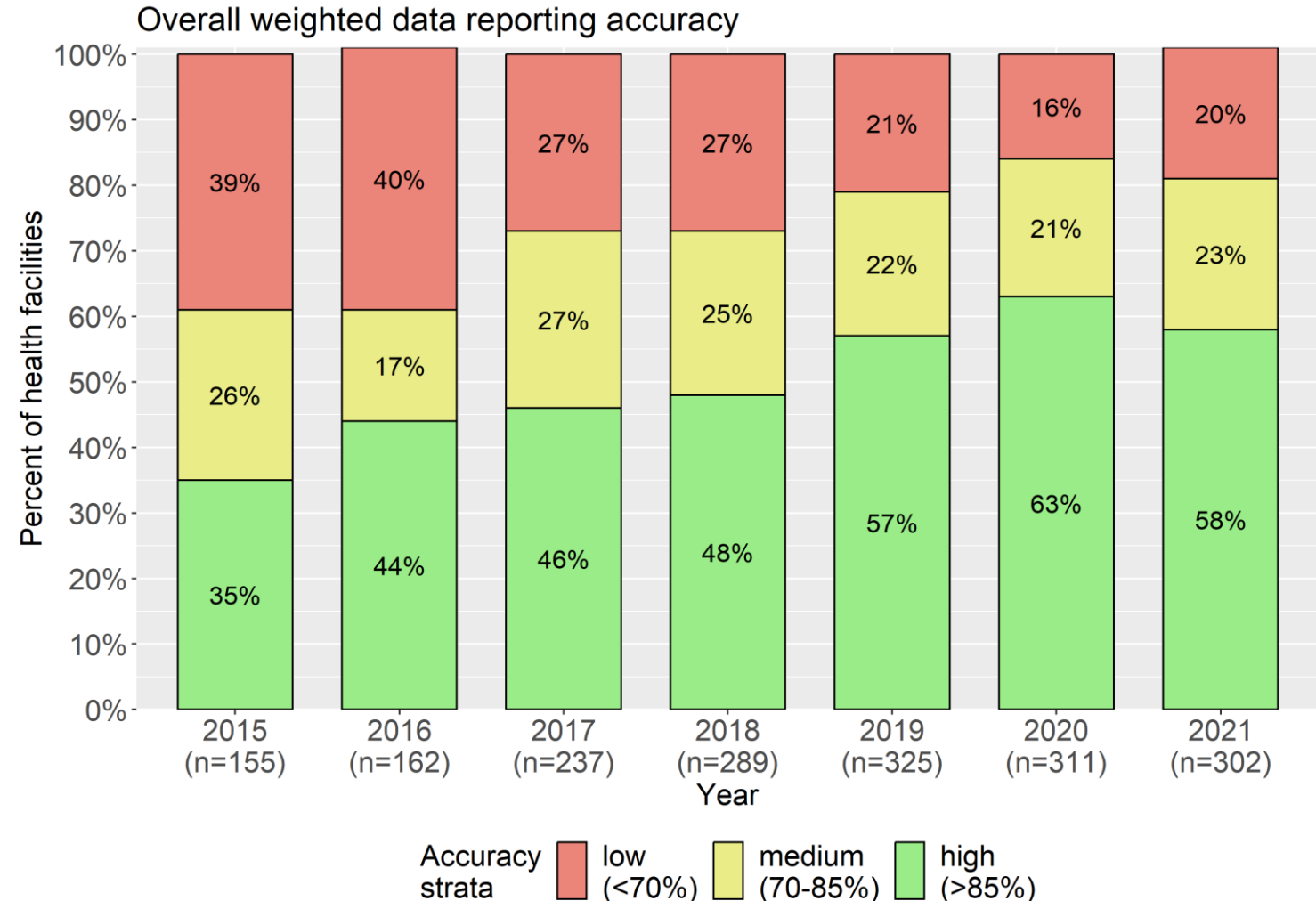
- Annual audit: 10 to 13 health facilities per district using purposive and random sampling methods
 - Every facility in a district is audited at least once every 2 years
 - 1 to 2 facilities in a district are audited per day
 - 6 months of MRR, HMIS, and CHW data are audited
- District audit teams are typically composed of 3 district staff plus 1-2 “outsiders” from another district or someone from the Province, NMEC, or supporting project
 - Prior to implementing DQAs, audit teams participate in a 1-2 day orientation meeting
 - DQA tools are pre-populated with reported data prior to DQA
- Teams travel to facilities and conduct audits on-site comparing reported data against data in register
- DQA findings are shared at the facility and district levels
 - Facility staff and CHWs: discuss discrepancies, provide technical mentorship, and develop action plans
 - District: Review meeting to develop action plan
- After all districts have completed DQAs, a 2-3 day meeting is conducted

The median overall weighted data reporting accuracy across all health facilities improved over time



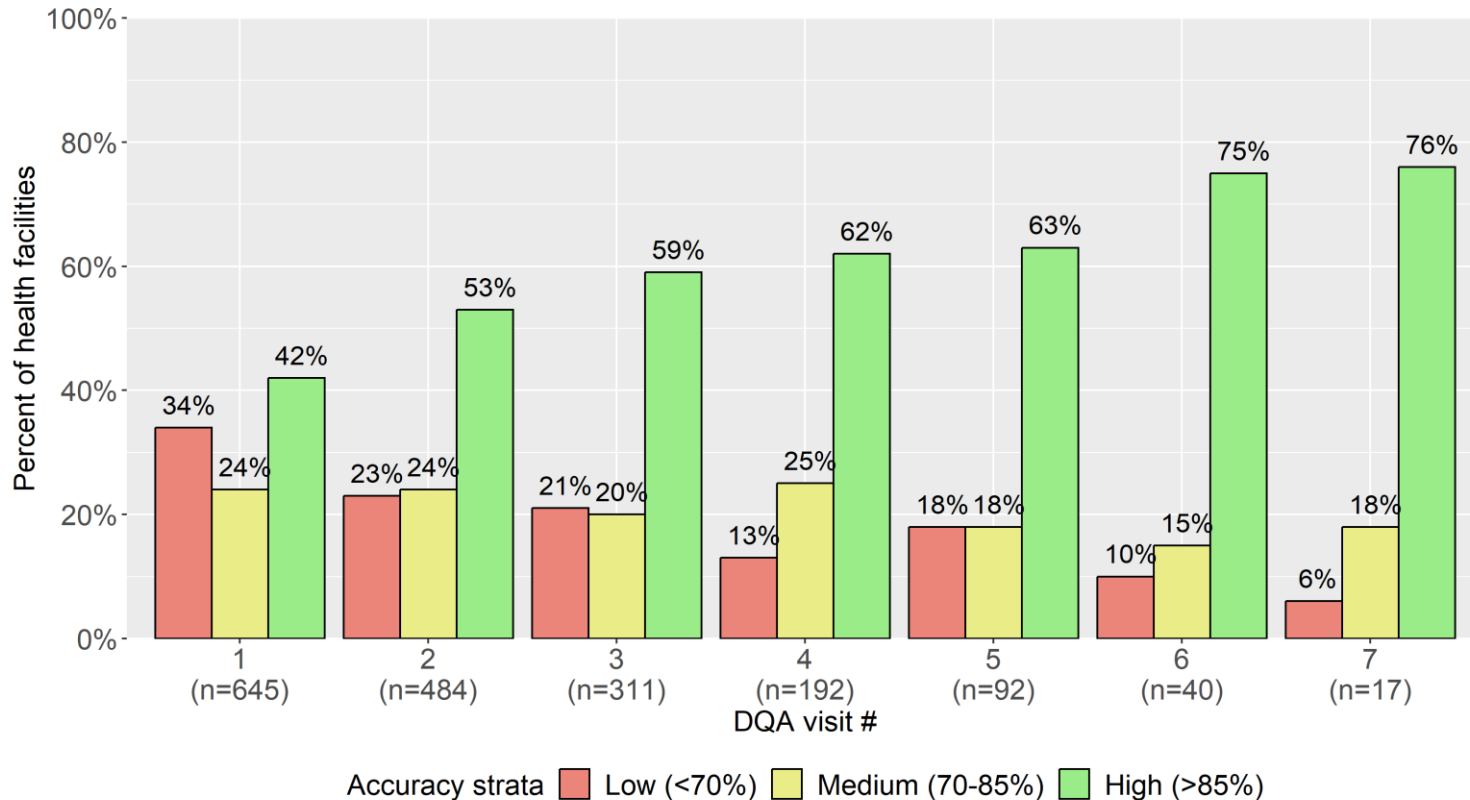
The percentage of health facilities with $\geq 85\%$ overall reporting accuracy increased over time

- Similarly, reduction observed in health facilities with low reporting accuracy from 39% in 2015 to 20% in 2021
- The percentage of health facilities in the medium accuracy strata is somewhat stable which may indicate similar rates of health facilities moving from **low** → **medium** accuracy strata and **medium** → **high** accuracy strata in a given year



Routine is important: reporting accuracy improves with repeated DQAs

Health facility overall weighted data reporting accuracy strata with each successive DQA conducted



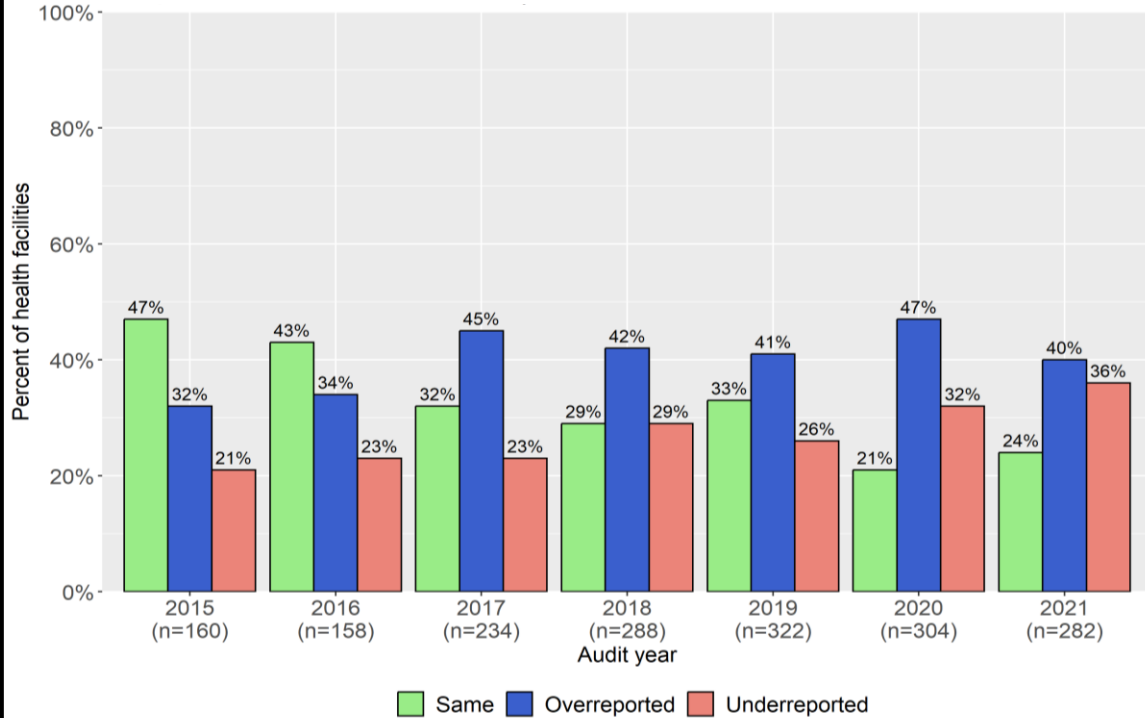
At the first DQA visit, overall reporting accuracies are mainly concentrated in the low (34%) and high (42%) accuracy strata. For those health facilities that eventually receive 6 or 7 DQAs, over 70% are in the high reporting accuracy strata while only about 10% are in the low accuracy strata

DQA visit	Accuracy strata	# HFs	DQA visit	Accuracy strata	# HFs
1	Low (<70%)	218	5	Low (<70%)	17
	Medium (70-85%)	155		Medium (70-85%)	17
	High (>85%)	272		High (>85%)	58
	Total	645		Total	92
2	Low (<70%)	113	6	Low (<70%)	4
	Medium (70-85%)	115		Medium (70-85%)	6
	High (>85%)	256		High (>85%)	30
	Total	484		Total	40
3	Low (<70%)	66	7	Low (<70%)	1
	Medium (70-85%)	63		Medium (70-85%)	3
	High (>85%)	182		High (>85%)	13
	Total	311		Total	17
4	Low (<70%)	25			
	Medium (70-85%)	48			
	High (>85%)	119			
	Total	192			

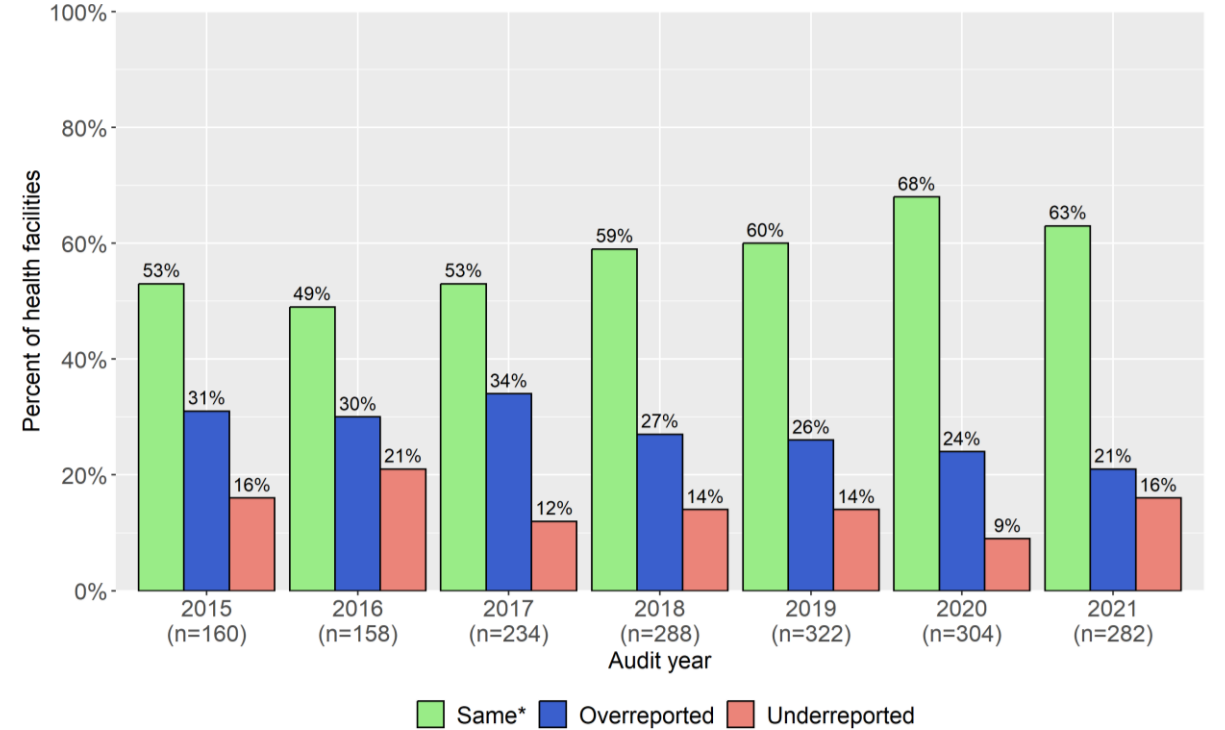
Do we need perfect reporting?

Reported incidence vs. register-based incidence each year

a) Unadjusted



b) Adjusted (allow ± 10% reporting error)

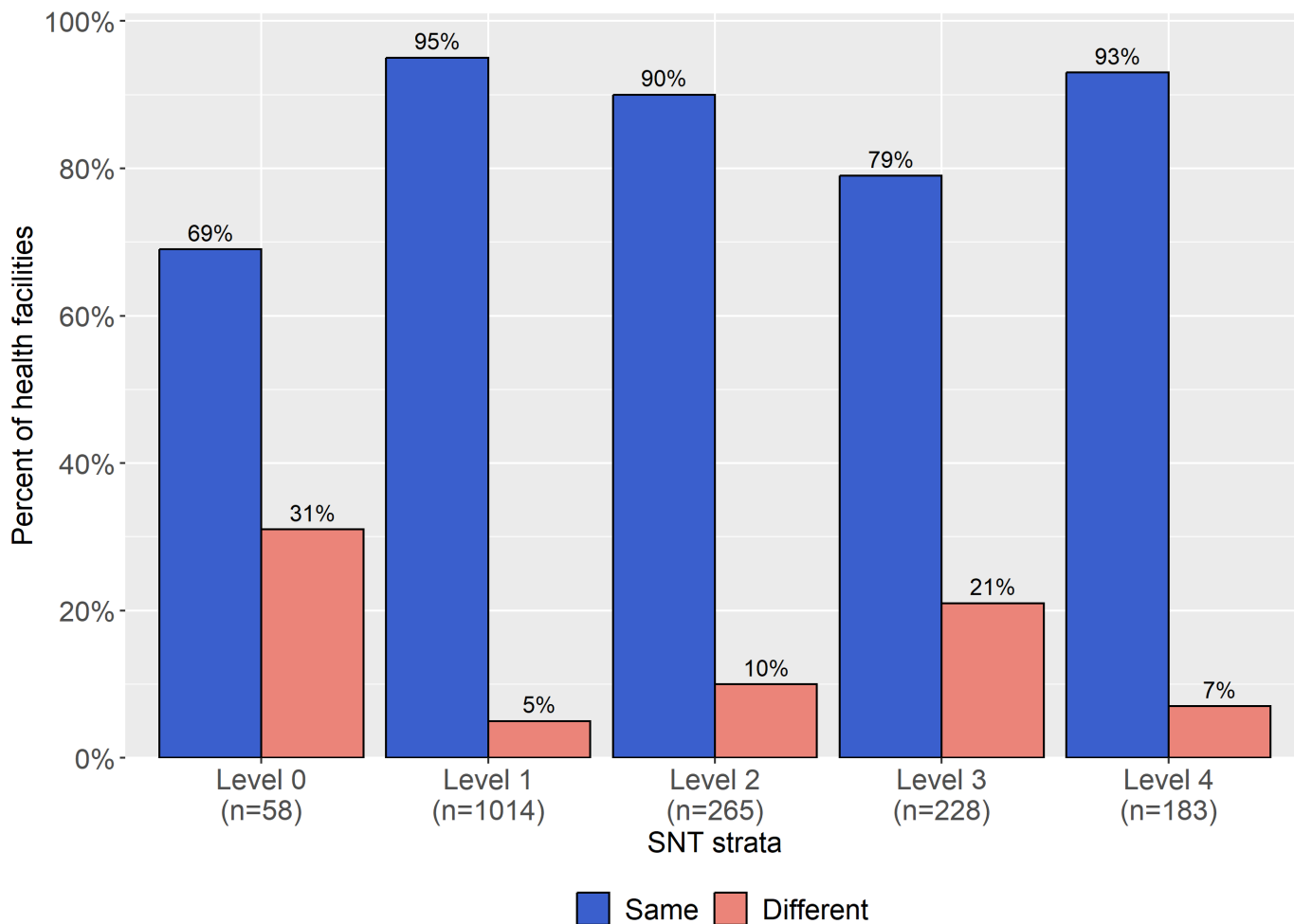


*Same when incidence_system is within +/- 10% of incidence_source

Most facilities appear to be under-reporting or over-reporting malaria cases (left figure). After adjusting to allow ± 10% error in reported incidence (right figure), most facilities appear "accurate enough".

Majority of health facilities assigned to SNT strata based on reported incidences are in the same SNT strata determined by register-based incidences

Comparison of SNT strata based on reported incidence to registered-based incidence by strata



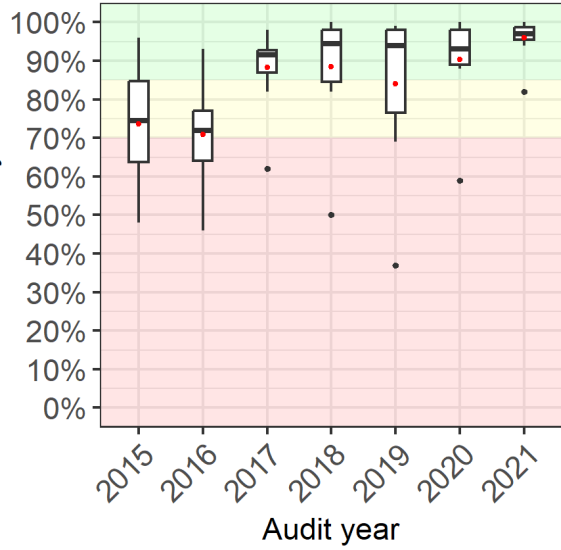
This visualization could be used to identify those “different” facilities and do additional follow-up or prioritize those facilities for the next DQA. For example, at Level 0, facilities had differing SNT strata when comparing reported incidence and register-based incidence. In an elimination setting, special attention should be given to these “different” health facilities.

The levels of malaria transmission intensity are stratified as “high” level 4 (above 500 cases per 1000 population/year), “moderate” level 3 (between 200 and 500 cases per 1000 population per year), “low” level 2 (between 50 and 200 cases per 1000 population/year), “very low” level 1 (between 0 and 50 cases per 1000 population/year), or “no malaria” level 0.

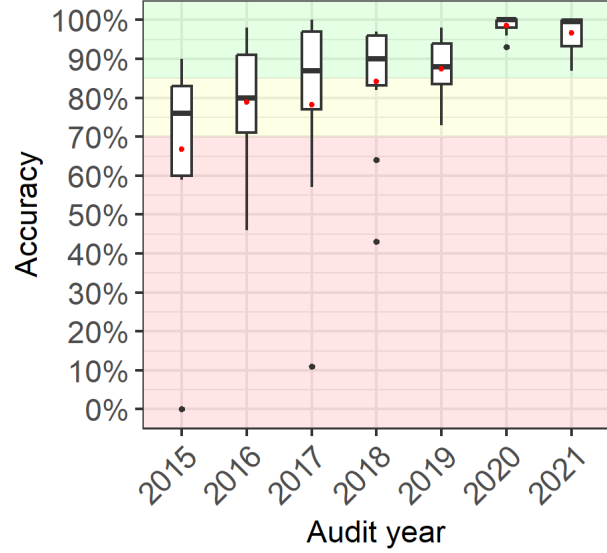
District level findings

Mazabuka District

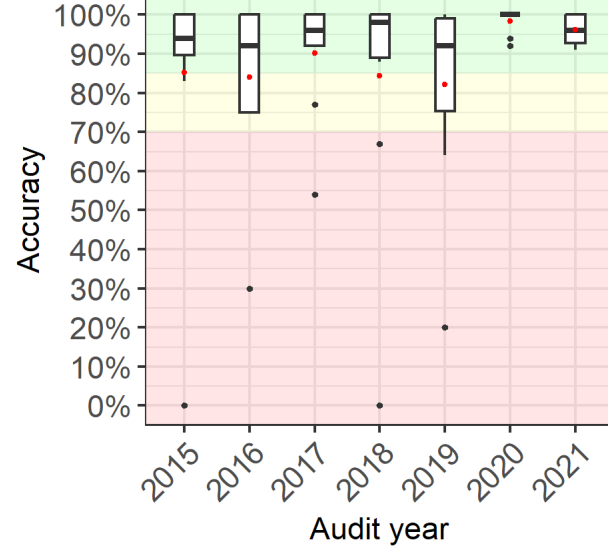
Total OPD attendance



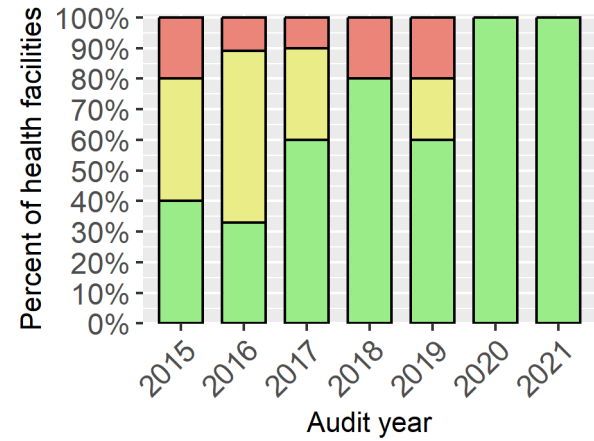
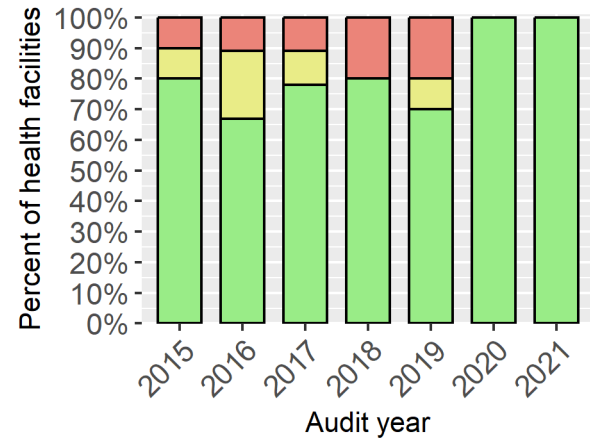
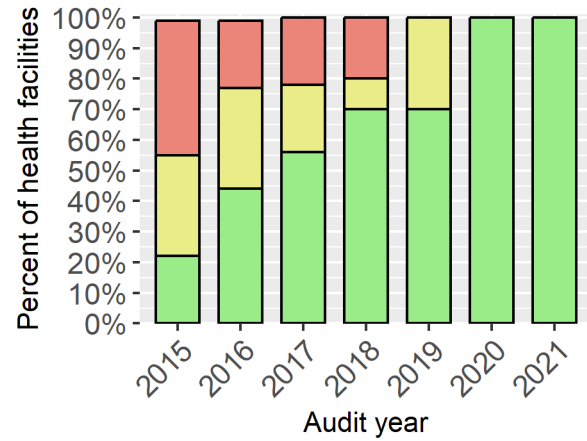
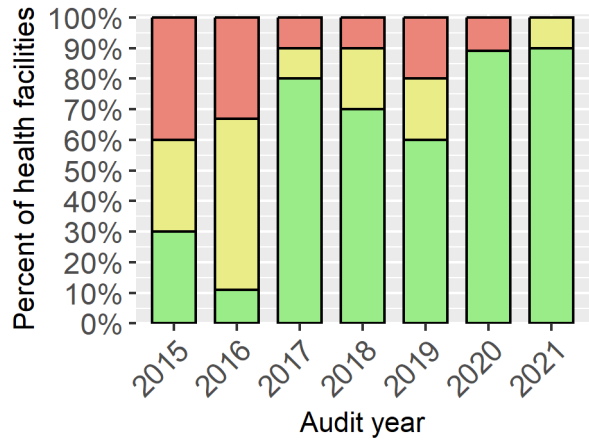
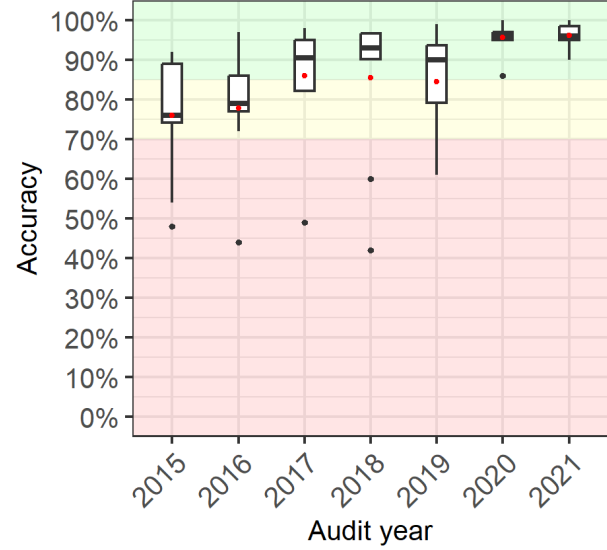
RDT tested cases



RDT positive cases



Overall



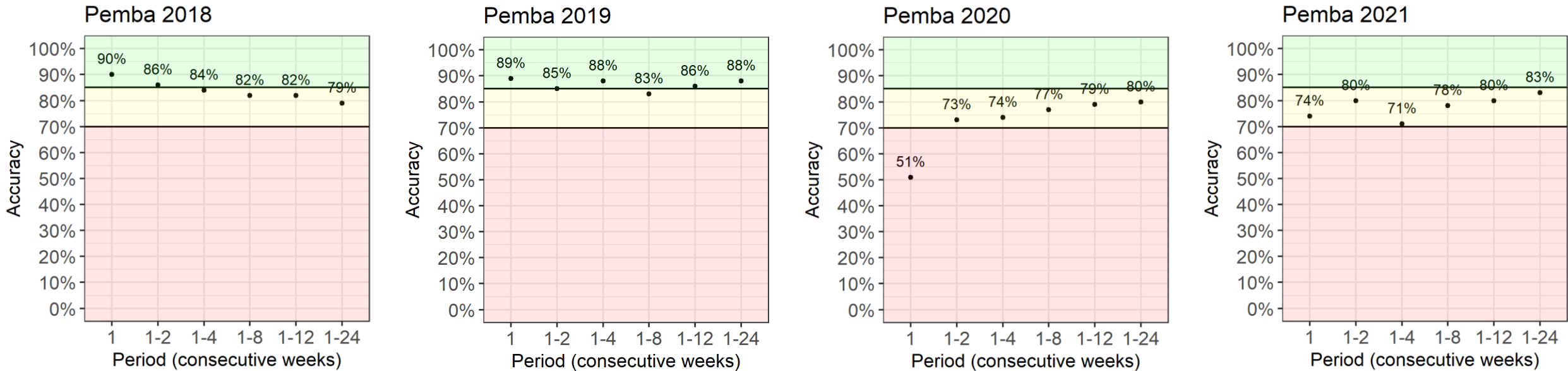
Accuracy strata: low (<70%), medium (70-85%), high (>85%)

Accuracy strata: low (<70%), medium (70-85%), high (>85%)

Accuracy strata: low (<70%), medium (70-85%), high (>85%)

Accuracy strata: low (<70%), medium (70-85%), high (>85%)

For each district and year, overall weighted data reporting accuracy was measured for subsets of consecutive weeks: 1, 1-2 weeks, 1-4 weeks, 1-8 weeks, 1-12 weeks vs. 1-24 weeks (current practice from January to June)



- Auditing fewer number of consecutive weeks, particularly weeks 1-8 and 1-12, will likely produce similar or equivalent data reporting accuracy when compared to weeks 1-24 (current practice)
- Health facilities that show poor accuracy (<70%) at weeks 1-8 and 1-12 can be further investigated by auditing the remaining weeks (up to 24 weeks)

Recommendations for Zambia NMEC Program

- The WAPE-based method should be used to measure aggregate data reporting accuracy
- The median should be used to understand aggregate data reporting accuracy instead of the average
- Routine DQAs should be maintained in Zambia to continue to increase the number of health facilities with $\geq 85\%$ reporting accuracy
- Develop a guidance document outlining use cases for DQAs at different health levels to better track the impact of DQAs on decision-making
- Consider conducting DQAs of fewer consecutive weeks to effectively and efficiently use resources
- Visualization (boxplots and stacked bar charts) of district level trends may help district health officers more clearly observe temporal trends in health facility data reporting accuracy across key malaria data elements and aggregate; May also help district officers identify consistently low performing/outlier facilities and direct additional resources.

For more
information
contact:

Smita Das

sdas@path.org

Michael Hainsworth

mhainsworth@path.org

Arantxa Roca Feltrer

afeltrer@path.org

PATH
▶◊::▲○◆//☹◻⊕

Supplementary

Conclusions

- Using WAPE-base method to measure reporting accuracy resolves many of the issues encountered with the current method by using the realized value of observations as the weighting factor
- Median is the best measure to determine the center of the distribution for data reporting accuracies; it is much less influenced by outliers
- From 2015 to 2021, the median weighted data reporting accuracy of the three data elements, individually and aggregate, increased and the data became less dispersed (smaller length of box). This change was observed to a lesser extent for RDT positives.
- Overall, by 2021, there were more health facilities with $\geq 85\%$ reporting accuracy and fewer health facilities with $\leq 70\%$ accuracy compared to previous audit years.
- With each successive DQA conducted, overall weighted reporting accuracy increased with most health facilities in the medium and high accuracy strata.
- Even health facilities that received a total of 2 DQAs benefited and showed improvement in weighted reporting accuracy. Considerable improvements by the last visit occurred in health facilities that received at least four total DQAs.
- Majority of underreported and overreported system-based incidences do not result in incorrect assignment to SNT intervention strata but there may be other malaria program activities that are more sensitive by inaccurate system and register values.
- Weighted data reporting accuracy trends (individual data elements and aggregate) for each district have been visualized.
- Performing audits for weeks 1-8 or 1-12 result in similar or equivalent data reporting accuracies as weeks 1-24.

Next steps

- Determine rate of corrections in MRR pre- and post-DQA
- Explore the influence, if any, of district level staff turnover on data reporting accuracy
- Examine if data reporting accuracy is affected by the number of total health facilities in each district
- Further investigate health facilities reporting <70% overall accuracy in at least three DQAs
- Continue to develop recommendations and present this work to Zambia NMEC Program for feedback and discussion on potential applications of learnings from this study to national and subnational level audit practices

The current method for measuring data reporting accuracy could be further improved to help prioritize health facilities using limited resources for data quality improvement actions

- For each week, the accuracy score is measured as: **Min/Max**

- Aggregate (total) accuracy: $1 - \frac{|\sum Source - \sum System|}{\sum Source}$

Example 1 – Errors between source and system values cancel each other when measuring overall accuracy

Health facility #1

Week	Source	System	Difference (Source-System)	Accuracy
1	1	4	-3	25%
2	1	2	-1	50%
3	2	1	+1	50%
4	0	0	0	100%
5	2	0	+2	33%
6	2	1	+1	50%
Total	8	8	0	100%

Accuracy for Weeks 1-6 determined by Min/Max measure

Aggregate accuracy

Example 2 – Equal importance given to health facilities with low reporting accuracies regardless of source values

Health facility #1

Week	Source	System	Accuracy
1	1	0	50%
2	2	1	50%
3	0	0	100%
4	2	1	50%
5	0	0	100%
6	3	2	67%
Total	8	4	50%

Health facility #2

Week	Source	System	Accuracy
1	103	53	51%
2	110	42	38%
3	101	65	64%
4	98	50	51%
5	100	50	50%
6	88	40	45%
Total	600	300	50%

Weighted Average Percentage Error (WAPE) was selected as the best method of measuring data reporting accuracy

$$WAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n y_i} * 100$$

where:

- n : represents the number of observations
- y_i : represents the realized value
- \hat{y}_i : represents the predicted value

For each health facility, WAPE was applied to three data elements (Total OPD attendance, RDT tested cases, RDT positive cases) to determine individual element accuracies. Overall accuracy was measured by taking the average of the three data elements' accuracies.

- Extension of the Mean Average Percentage Error (MAPE) – a common method for reporting accuracy which takes the sum of each week's PE and divides it by the total number of observations (aka - average PE)
- Instead of weighing the sum of the absolute error by the total number of observations, **WAPE uses the source value of each observation as the weighting factor**
- **Gives more importance to the errors of observations with a high source value – it is a convenient metric to prioritize observations**
- Equally penalizes overreporting and underreporting
- Accuracy = 1 – WAPE

Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value A_t . The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points, n .

- Relative error measure that uses absolute values to keep the positive and negative errors from cancelling one another out and uses relative errors to enable comparison of forecast accuracy between time-series models
- Asymmetric and puts heavier penalty on negative error (when forecasts are higher than actuals) than on positive errors. This is caused by the fact that the percentage error cannot exceed 100% for forecasts that are too low. While there is no upper limit for forecasts which are too high. As a result, MAPE will favor models that under-forecast rather than over-cast.
- Degenerates into positive infinity as soon as any of the actual values is zero; even relatively small actual values can easily send MAPE toward infinity
- Large percentage errors can drive most of the overall error in a group
- Unweighted and averages percentages – assumes absolute error on each item is equal

For these reasons, we do not recommend using MAPE for determining data reporting accuracy

Example 1 - MAPE vs. WAPE methods: Nakatindi Urban Health Centre, RDT+ cases in 2020

Period	Date	Value (Source)	Value (System)	Abs diff	PE
1	12/29/2019	1	1	0	0
2	1/5/2020	2	2	0	0
3	1/12/2020	1	1	0	0
4	1/19/2020	1	1	0	0
5	1/26/2020	1	1	0	0
6	2/2/2020	1	1	0	0
7	2/9/2020	2	2	0	0
8	2/16/2020	3	3	0	0
9	2/23/2020	2	1	1	50
10	3/1/2020	1	1	0	0
11	3/8/2020	1	1	0	0
12	3/15/2020	1	1	0	0
13	3/22/2020	16	1	15	93.75
14	3/29/2020	11	0	11	100
15	4/5/2020	1	1	0	0
16	4/12/2020	1	2	1	100
17	4/19/2020	1	1	0	0
18	4/26/2020	1	1	0	0
19	5/3/2020	3	1	2	66.66667
20	5/10/2020	1	1	0	0
21	5/17/2020	1	1	0	0
22	5/24/2020	1	2	1	100
23	5/31/2020	2	1	1	50
24	6/7/2020	1	2	1	100
	Total	57		33	

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

vs.

$$WAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n y_i} * 100$$

	Percentage error (PE)	Accuracy (100-PE)
WAPE	57.9%	42.1%
MAPE	27.5%	72.5%

Weeks 13 and 14 make up most cases at this health facility, but are not being reported in DHIS2 correctly, which is concerning. The WAPE recognizes this and indicates a large error that should be further examined.

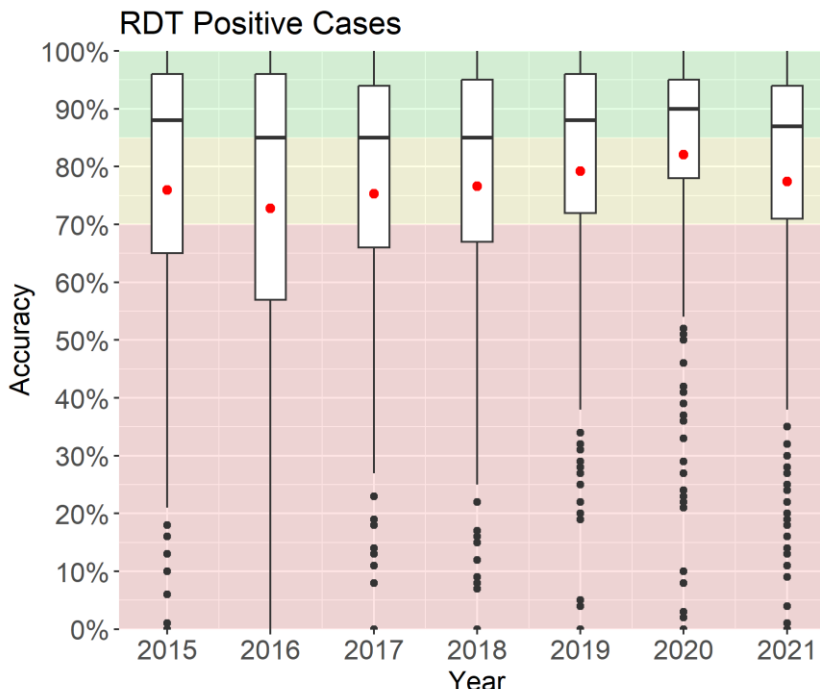
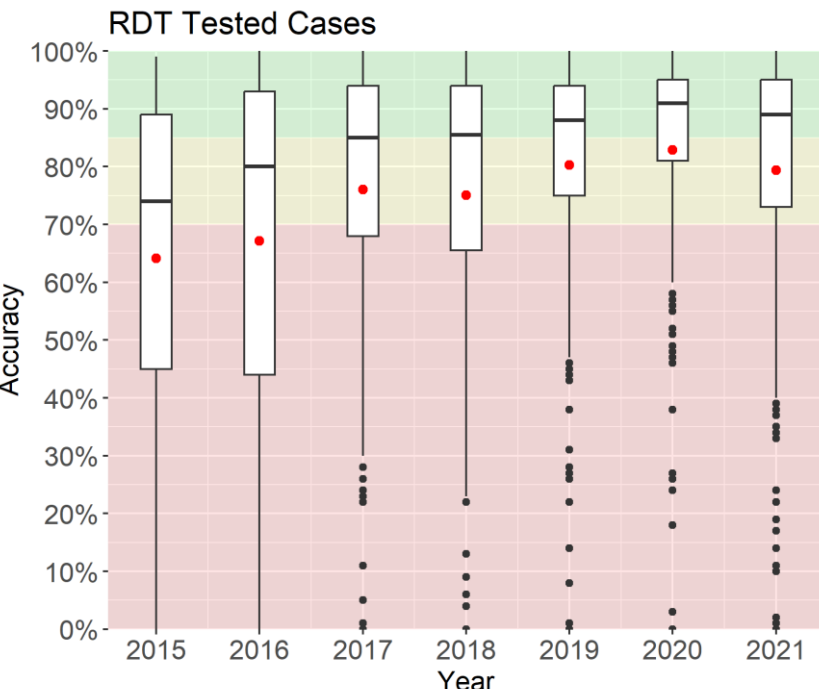
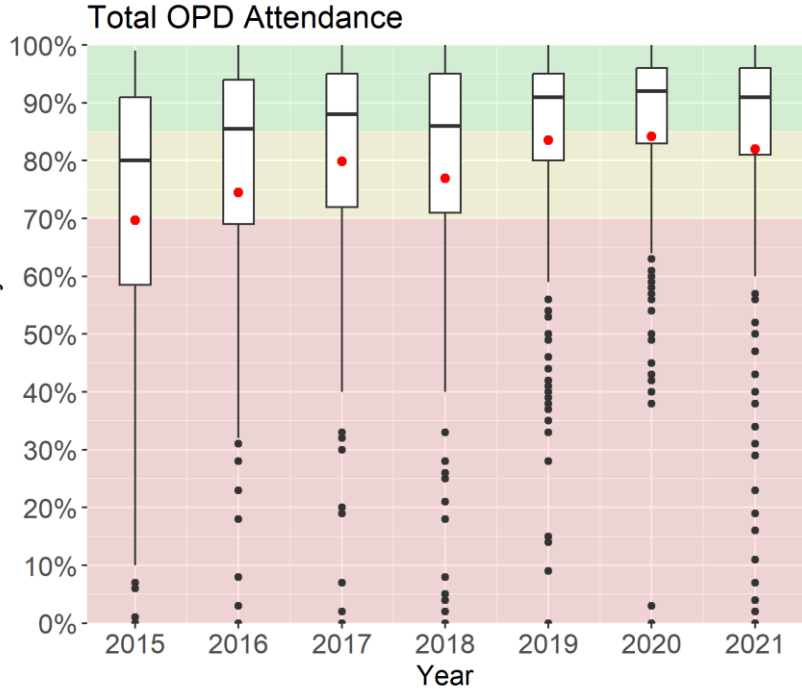
Example 2 - MAPE vs. WAPE methods: Mambolomoka RHC, RDT+ cases in 2020

Period	Date	Value (Source)	Value (System)	Abs diff	PE
1	12/29/2019	6	6	0	0.0
2	1/5/2020	1	1	0	0.0
3	1/12/2020	4	5	1	25.0
4	1/19/2020	8	8	0	0.0
5	1/26/2020	15	14	1	6.7
6	2/2/2020	18	16	2	11.1
7	2/9/2020	4	4	0	0.0
8	2/16/2020	3	5	2	66.7
9	2/23/2020	1	7	6	600.0
10	3/1/2020	2	0	2	100.0
11	3/8/2020	14	14	0	0.0
12	3/15/2020	11	9	2	18.2
13	3/22/2020	49	52	3	6.1
14	3/29/2020	57	55	2	3.5
15	4/5/2020	47	42	5	10.6
16	4/12/2020	51	50	1	2.0
17	4/19/2020	31	31	0	0.0
18	4/26/2020	53	51	2	3.8
19	5/3/2020	57	56	1	1.8
20	5/10/2020	36	53	17	47.2
21	5/17/2020	56	50	6	10.7
22	5/24/2020	36	34	2	5.6
23	5/31/2020	39	39	0	0.0
24	6/7/2020	56	56	0	0.0
	Total	655		55	

	Percentage error (PE)	Accuracy (100-PE)
WAPE	8.4%	91.6%
MAPE	38.3%	61.7%

The large PE in Week 9 contributes to the high overall MAPE-based PE, resulting in a lower grade of accuracy than WAPE-based accuracy. This is an example where one week's over-reporting drives the overall accuracy when applying MAPE.

Generally, the median of the weighted data reporting accuracy improves each year across the three data elements for all health facilities



Accuracy levels:

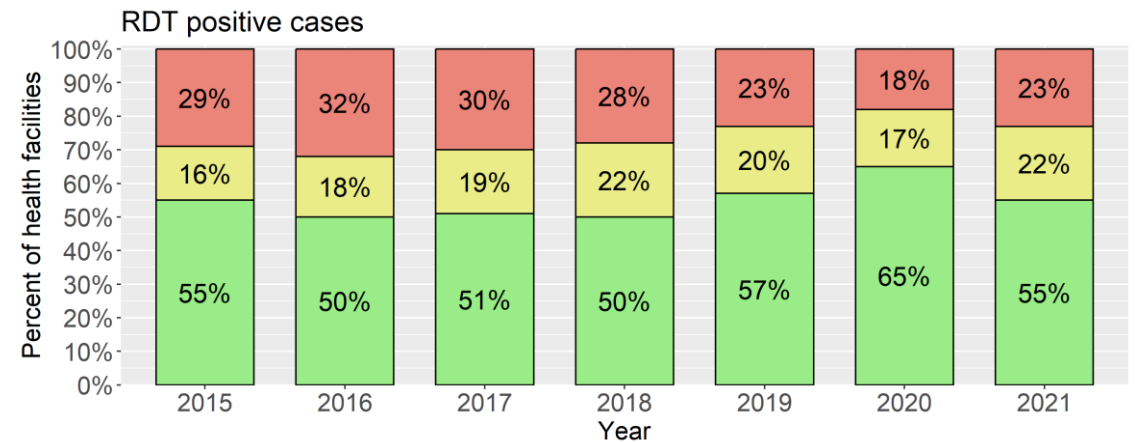
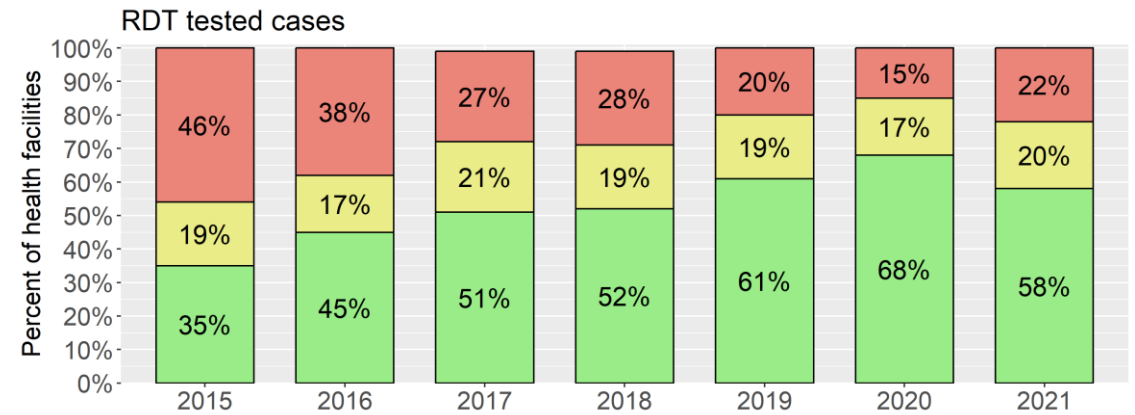
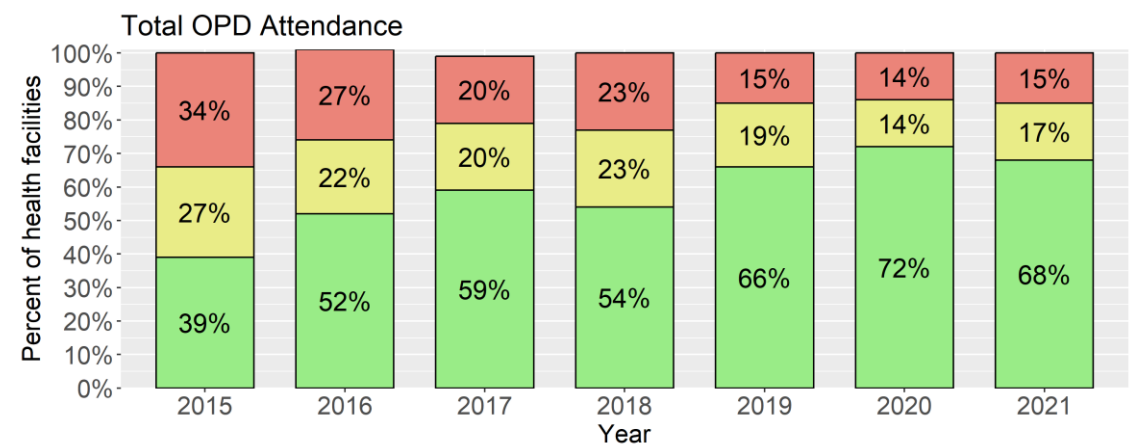
- High:** ≥85%
- Medium:** ≥70-85%
- Low:** <70%

Red dot = average weighted data reporting accuracy

— = median weighted data reporting accuracy

The percentage of health facilities with high reporting accuracy for each data element increases from 2015 to 2020

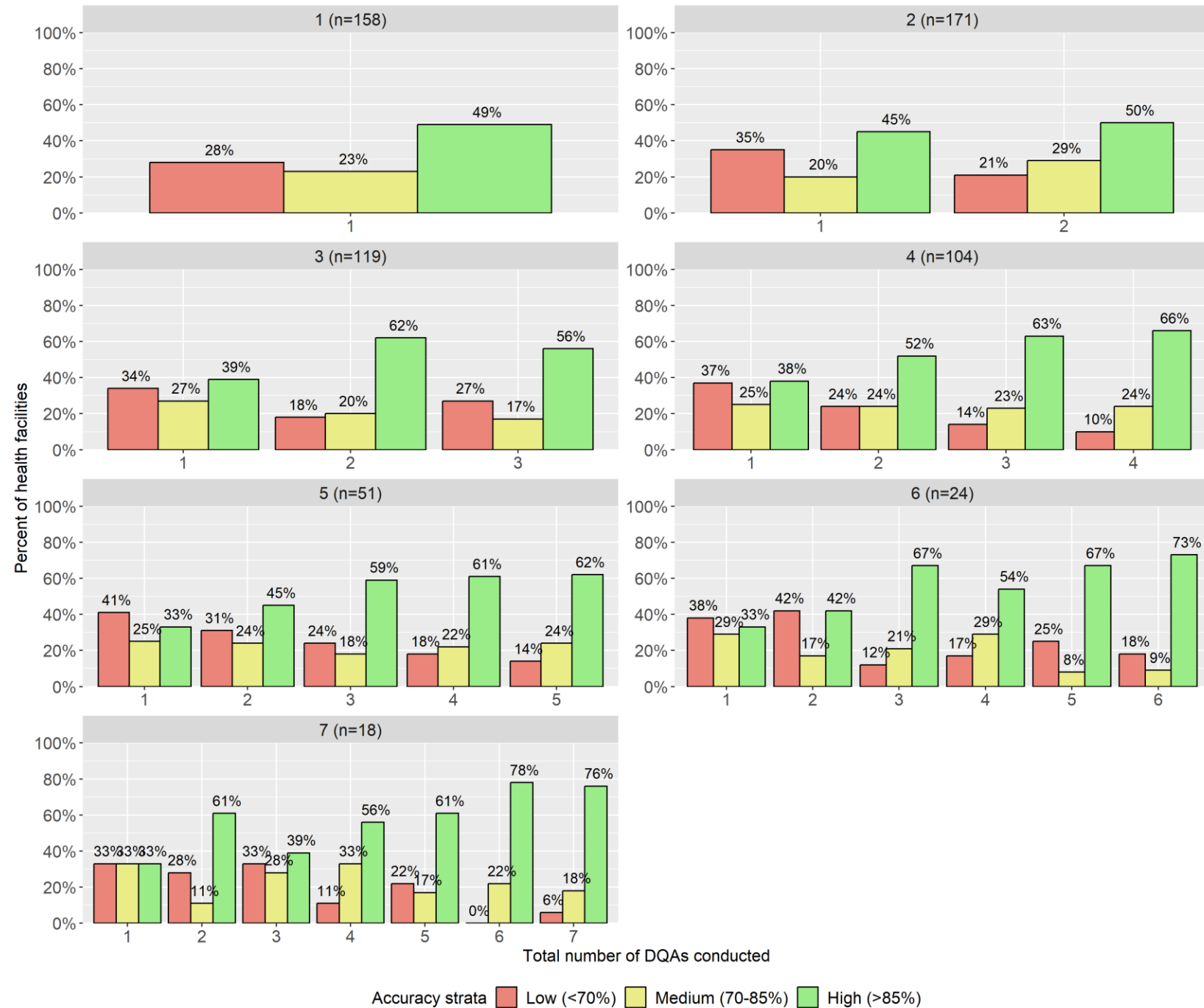
- The percentage of health facilities in the low accuracy strata also decreases over time across data elements
- The percentage of health facilities in the medium accuracy strata is somewhat stable which may indicate similar rates of health facilities moving from **low** → **medium** accuracy strata and **medium** → **high** accuracy strata in a given year



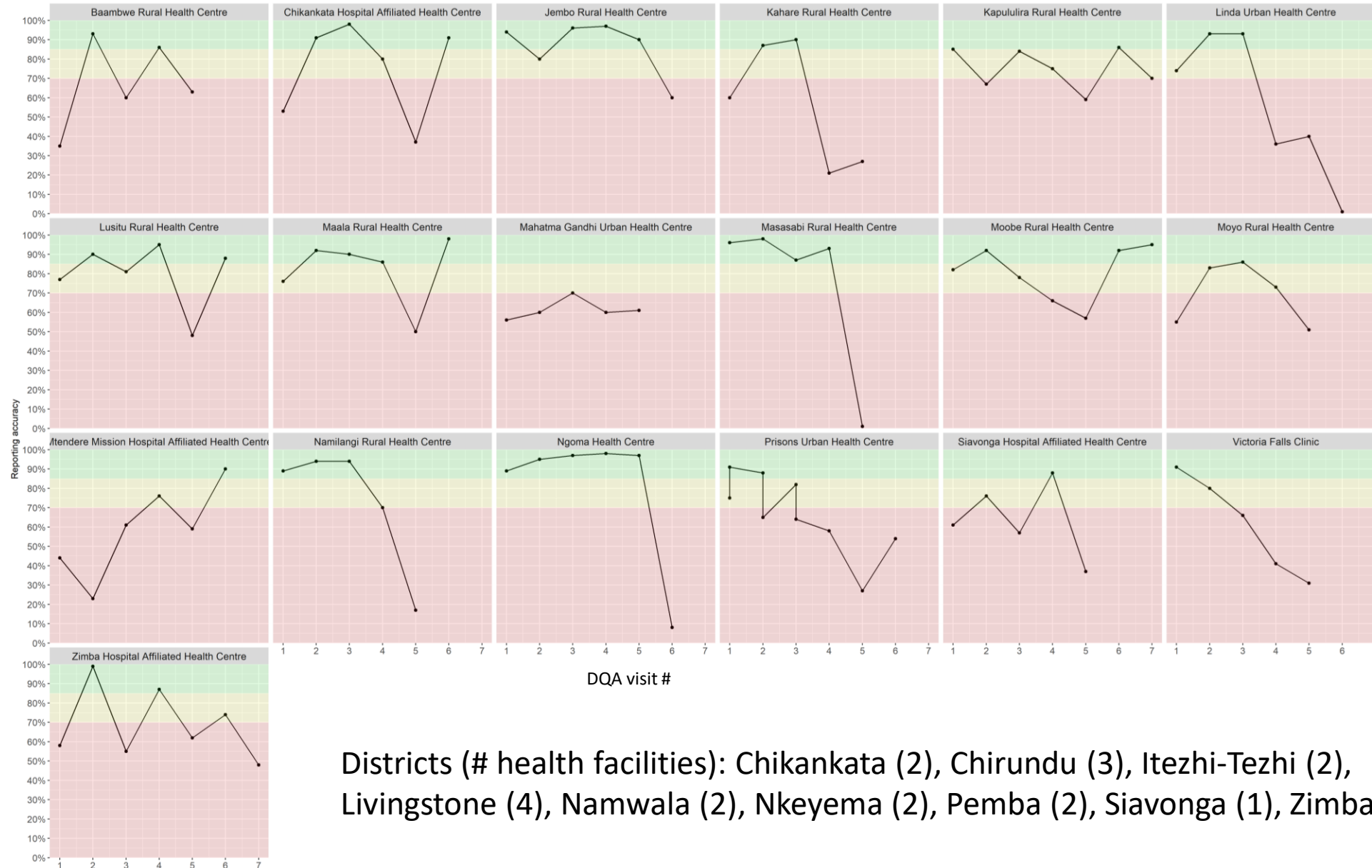
Accuracy strata: Low (<70%), Medium (70-85%), High (>85%)

Grouping facilities by the total number of DQAs received further strengthens the finding that repeated DQAs are associated with improved reporting accuracy

Total DQAs conducted	Percentage point change per strata, first to last DQA visit		
	Low (<70%)	Medium (70-85%)	High (>85%)
1	-	-	-
2	-14	+8	+5
3	-6	-9	+15
4	-26	-1	+27
5	-28	-2	+30
6	-20	-20	+40
7	-27	-15	+43



There are health facilities that show erratic or inconsistent weighted accuracy levels despite receiving at least 5 DQA visits, suggesting other influencing factors that will be critical to understand

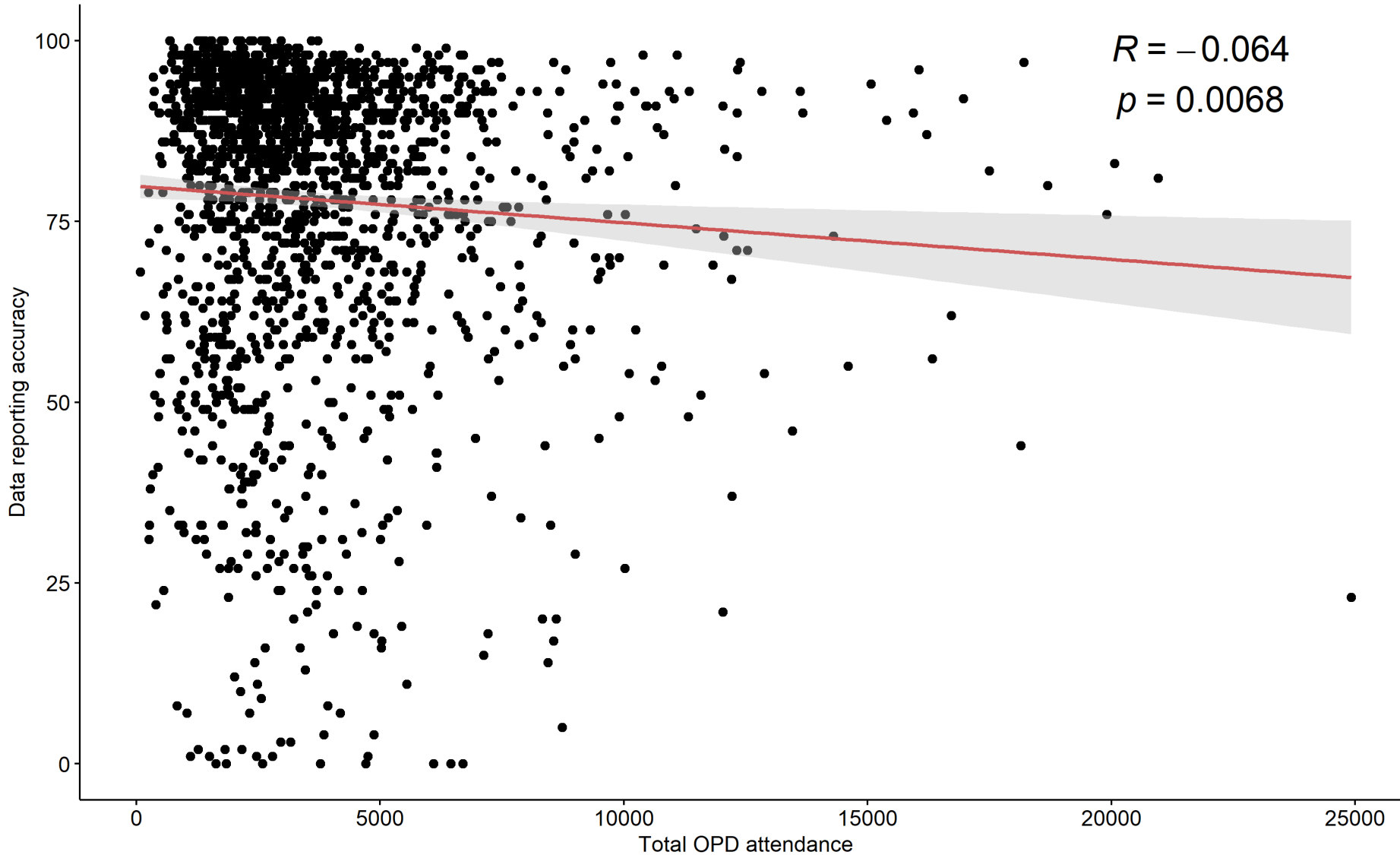


Accuracy levels:
High: ≥85%
Medium: ≥70-85%
Low: <70%

Districts (# health facilities): Chikankata (2), Chirundu (3), Itezhi-Tezhi (2), Livingstone (4), Namwala (2), Nkeyema (2), Pemba (2), Siavonga (1), Zimba (1)

There is a *very weak* negative correlation between health facility size and overall weighted data reporting accuracy

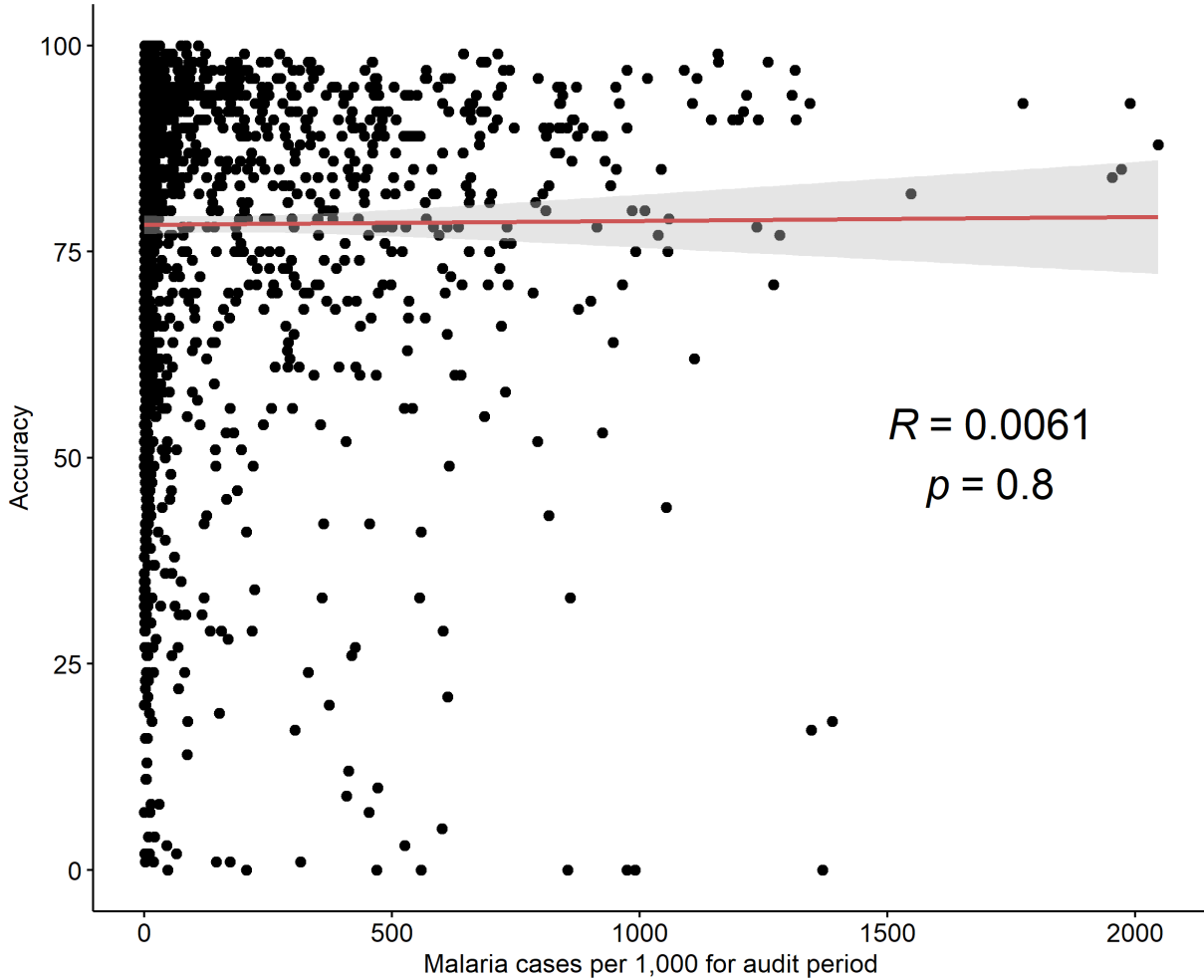
Pearson correlation analysis between health facility size and overall weighted data reporting accuracy



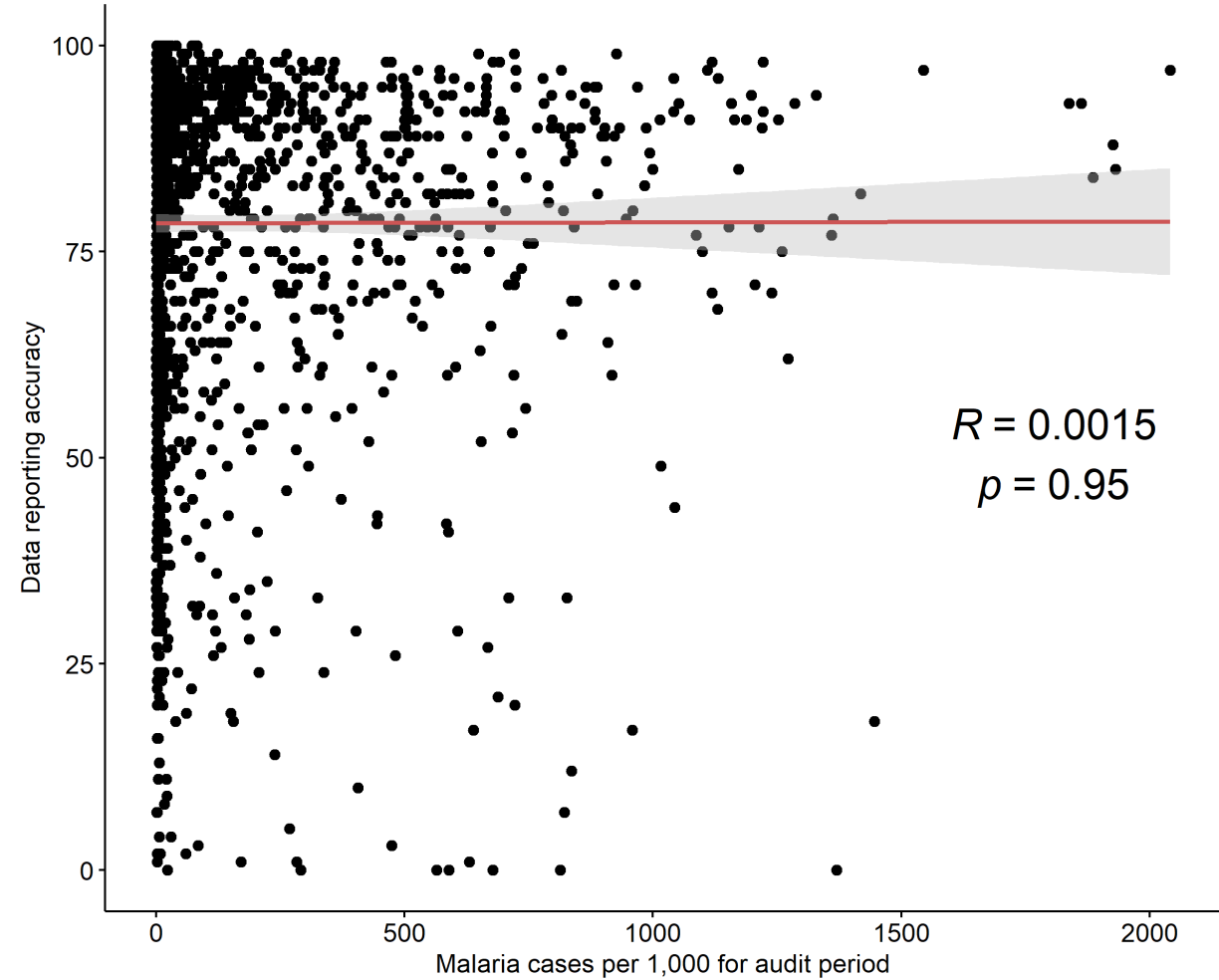
Because Pearson correlation coefficient is so close to zero, we should interpret it as no association between health facility size and data reporting accuracy.

There is no correlation between audit period malaria incidence and overall weighted data reporting accuracy

Pearson correlation analysis between malaria incidence (based on source values) and overall weighted data reporting accuracy



Pearson correlation analysis between malaria incidence (based on system values) and overall weighted data reporting accuracy



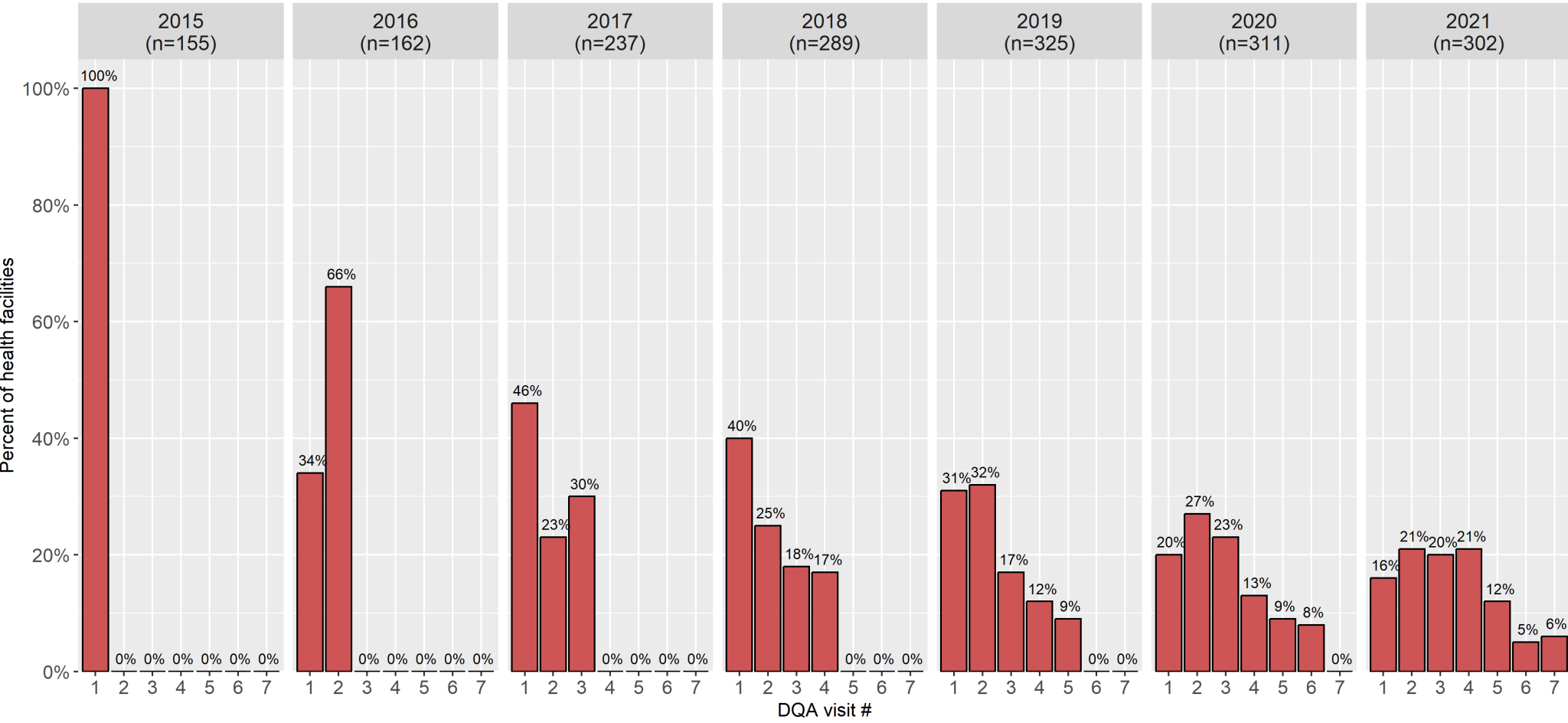
Sample sizes: Comparison of SNT strata based on reporting incidence to register-based incidence by year

Audit year	Concordance of SNT incidence strata, source vs. system	Number of health facilities (unadjusted)	Number of health facilities (adjusted)
2015 (n=160)	Same	142	143
	Different	18	17
2016 (n=158)	Same	151	152
	Different	7	6
2017 (n=234)	Same	215	215
	Different	19	19
2018 (n=288)	Same	255	257
	Different	33	31
2019 (n=322)	Same	295	297
	Different	27	25
2020 (n=304)	Same	279	287
	Different	25	17
2021 (n=282)	Same	261	265
	Different	21	17

Sample sizes: Comparison of SNT strata based on reported incidence and register-based incidence by level

SNT level	Concordance of SNT incidence strata, source vs. system	Number of health facilities (unadjusted)	Number of health facilities (adjusted)
Level 0 No malaria (n=58)	Same	40	40
	Different	18	18
Level 1 Very Low (n=1014)	Same	968	970
	Different	46	44
Level 2 Low (n=265)	Same	238	241
	Different	27	24
Level 3 Moderate (n=228)	Same	181	191
	Different	47	37
Level 4 High (n=183)	Same	171	174
	Different	12	9

Fewer new facilities were added in 2020 and 2021 compared to most previous years

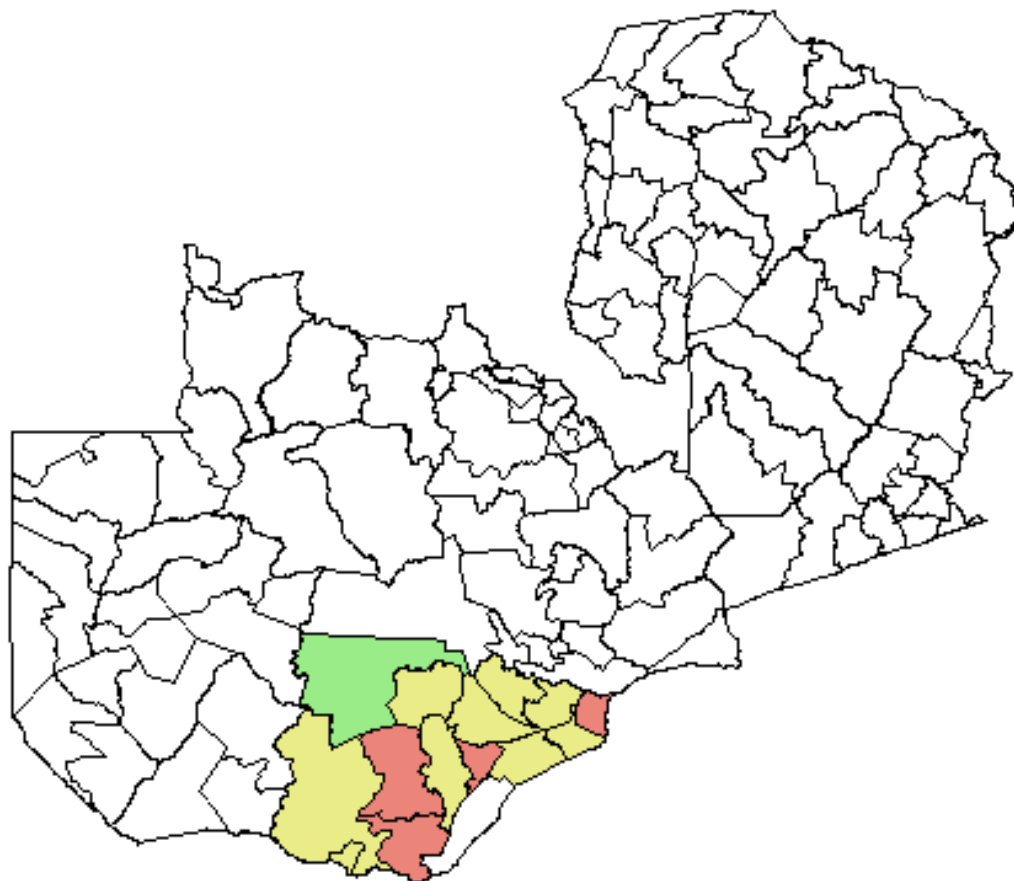


Audit year	# HF's receiving their first DQA
2015	155
2016	55
2017	110
2018	116
2019	100
2020	62
2021	47

Median overall weighted data reporting accuracy* by district from 2015-2021

* based on Total OPD attendance, RDT tested cases, RDT positive cases

2015



Median reporting accuracy strata

low (<70%)	medium (70-85%)	high (>85%)
------------	-----------------	-------------

Overall Data Reporting Accuracy by District, 2021



Reporting accuracy strata based on auditing weeks 1-8 and 1-12 is notably comparable to auditing weeks 1-24

