Data Quality Review (DQR) Desk Review Tools and Methods Workshop

Session 4: Domain 2

Internal Consistency of Reported Data

World Health Organization

The Global Fund

JSI

USAID
FROM THE AMERICAN PEOPLE

MEASURE Evaluation

Learning Objectives

SESSION 4

Domain 2: Internal Consistency

- **Learning Objective** is to familiarize the workshop participants with the analysis of metrics for Domain 2–Internal Consistency. Specifically, by the end of the workshop the participants will;
- Understand the data requirements for the different metrics pertaining to Internal Consistency in Domain 2
- Understand mechanisms and challenges to accessing data for evaluating consistency, particularly program-specific data sets
- Learn to configure analyses and metrics within Domain 2 using the DHIS 2 app
- Learn to input and analyze data for internal consistency for the different metrics
- Learn to interpret findings and develop plans for improving internal consistency of data

Domain 2: Internal consistency

SESSION 4

Domain 2: Internal Consistency

Domains of Data Quality

1) Completeness & timeliness of data

2) Internal consistency of reported data

3) External comparisons/cross-checks (with other data sources, e.g. surveys)

4) External consistency of population data – review denominator data used to measure performance indicators

Domain 2: Internal consistency

SESSION 4

Domain 2: Internal Consistency

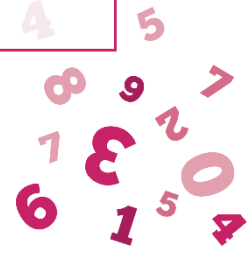
Domain 2: Internal Consistency of Reported Data

Focus

- Plausibility (apparent accuracy) of the data

Metrics

- a) Presence of "outliers" (suspicious values)
- b) Consistency from year to year
- c) Consistency of related indicators (e.g. Penta 1 vs Penta 3)
- d) Verification factor (a facility survey) -- consistency between source data (e.g. on clinic registers) and reported data



Outliers

SESSION 4

Domain 2: Internal Consistency

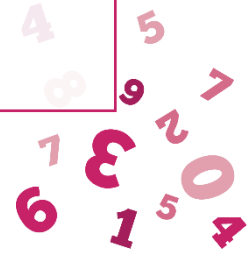
Domain 2: Internal Consistency of Reported Data

Focus

- Plausibility (apparent accuracy) of the data

Metrics

- a) Presence of “outliers” (suspicious values)
- b) Consistency from year to year
- c) Consistency of related indicators (e.g. Penta 1 vs Penta 3)
- d) Verification factor (a facility survey) -- consistency between source data (e.g. on clinic registers) and reported data



Internal Consistency of Reported Data—Outliers

Data Quality Metric	Severity	Definition	
		National Level	Subnational Level
Outliers (analyze each indicator separately)	**Extreme** (at least 3 standard deviations from the mean)	% of monthly subnational unit values that are extreme outliers	# (%) of subnational units in which ≥1 of the monthly subnational unit values over the course of 1 year is an extreme outlier value
	Moderate (between 2-3 standard deviations from the mean or >3.5 on modified Z-score method)	% of subnational unit values that are moderate outliers	# (%) of subnational units in which ≥2 of the monthly subnational unit values over the course of 1 year are moderate outliers

Outliers

SESSION 4

Domain 2: Internal Consistency

Example — Outliers in the Current Year

Months with at least 1 moderate outlier on district monthly report are shown in red.

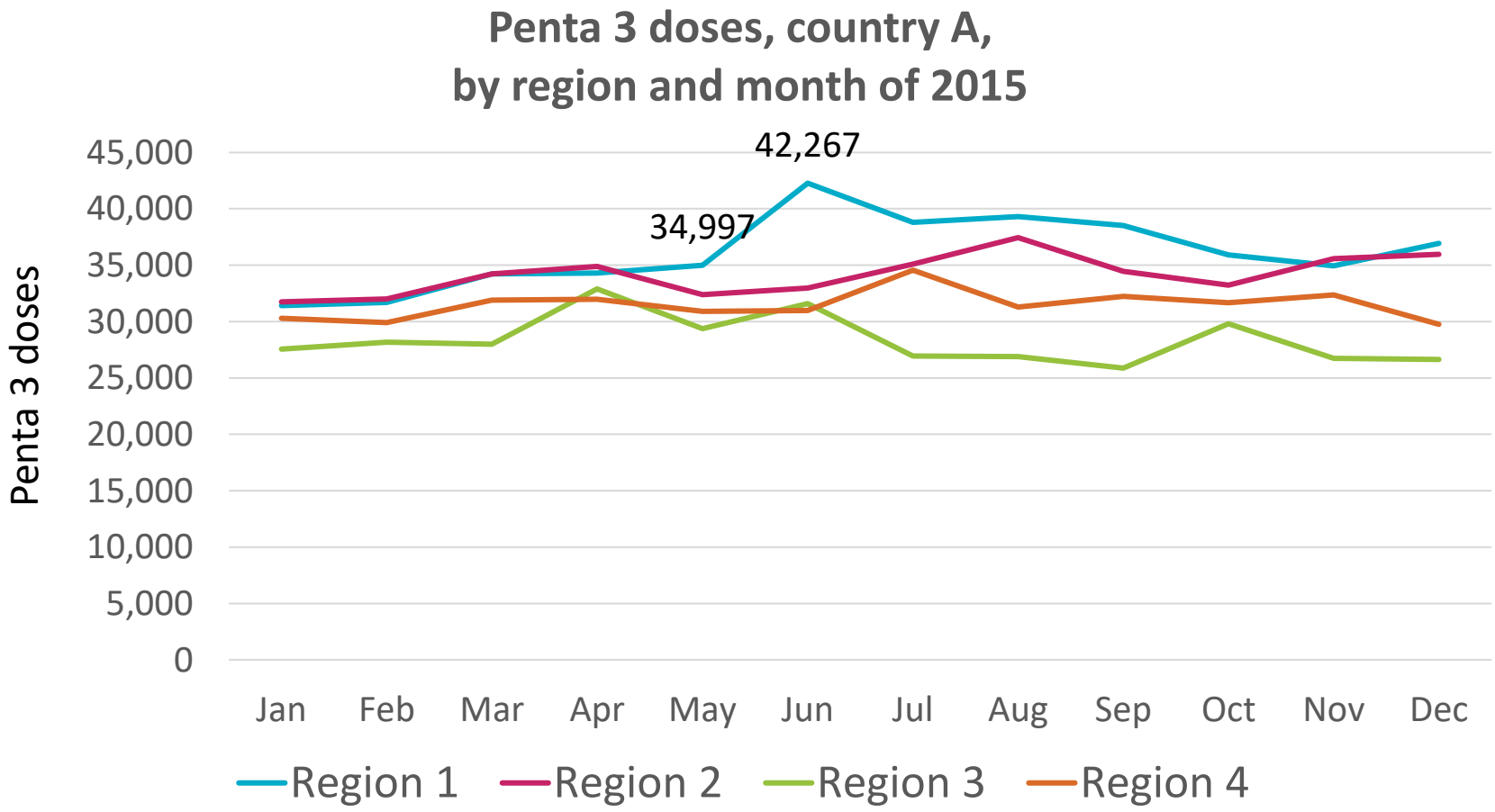
Dist	Month												Total Outliers	% Outliers
	1	2	3	4	5	6	7	8	9	10	11	12		
A	2543	2482	2492	2574	3012	2709	3019	2750	3127	2841	2725	2103	1	8.3%
B	1184	1118	1195	1228	1601	1324	1322	711	1160	1178	1084	1112	2	16.7%
C	776	541	515	527	857	782	735	694	687	628	596	543	0	0%
D	3114	2931	2956	4637	6288	4340	3788	3939	3708	4035	3738	3606	1	8.3%
E	1382	1379	1134	1378	1417	1302	1415	1169	1369	1184	1207	1079	0	0%
Nat'l	**0**	**0**	**0**	**0**	**2**	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**4**	**6.7%**

Outliers

SESSION 4

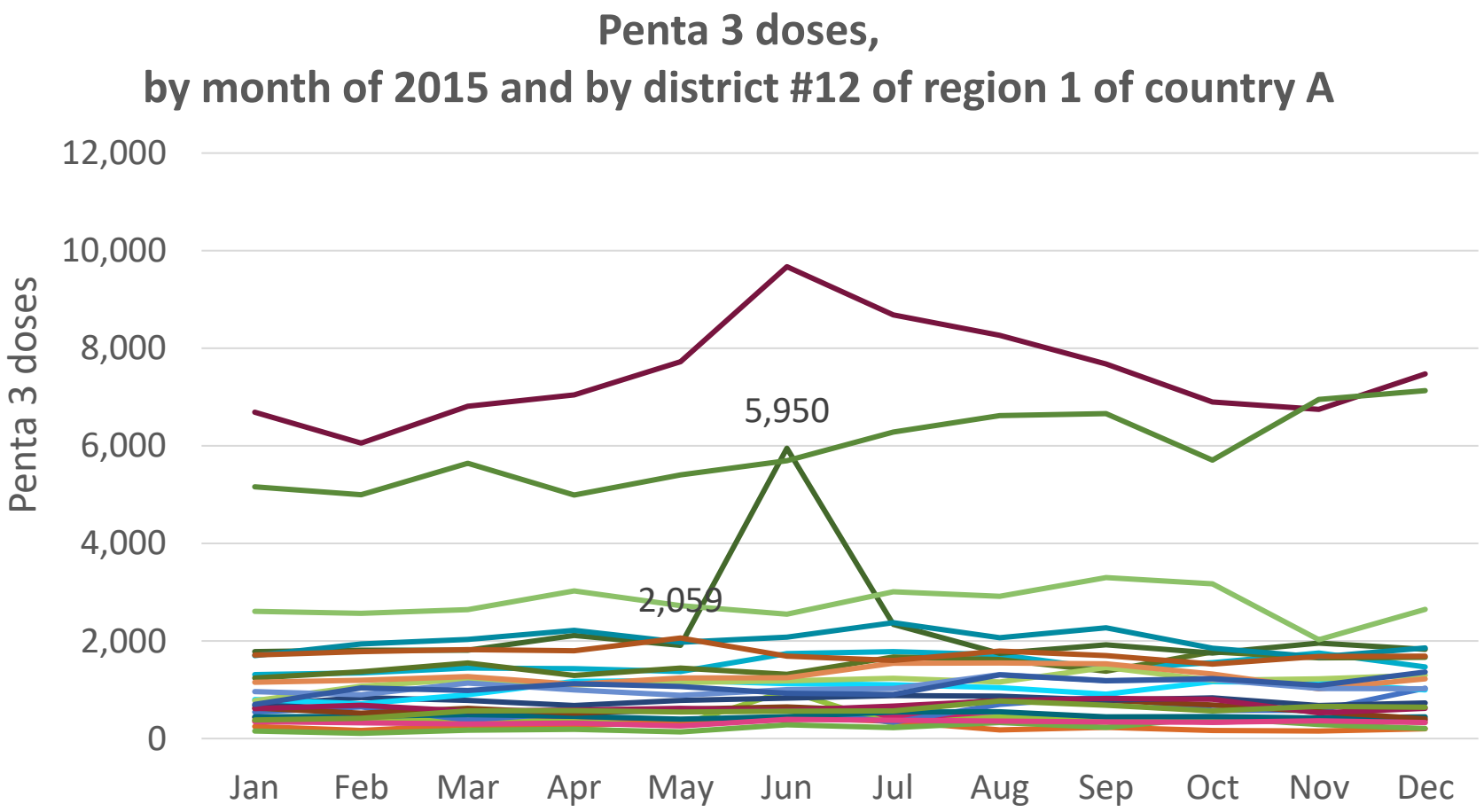
Domain 2: Internal Consistency

Domain 2a: "outliers" may be subtle with regional data



Outliers

Domain 2a: outliers become more apparent by "drilling down" to district-level data

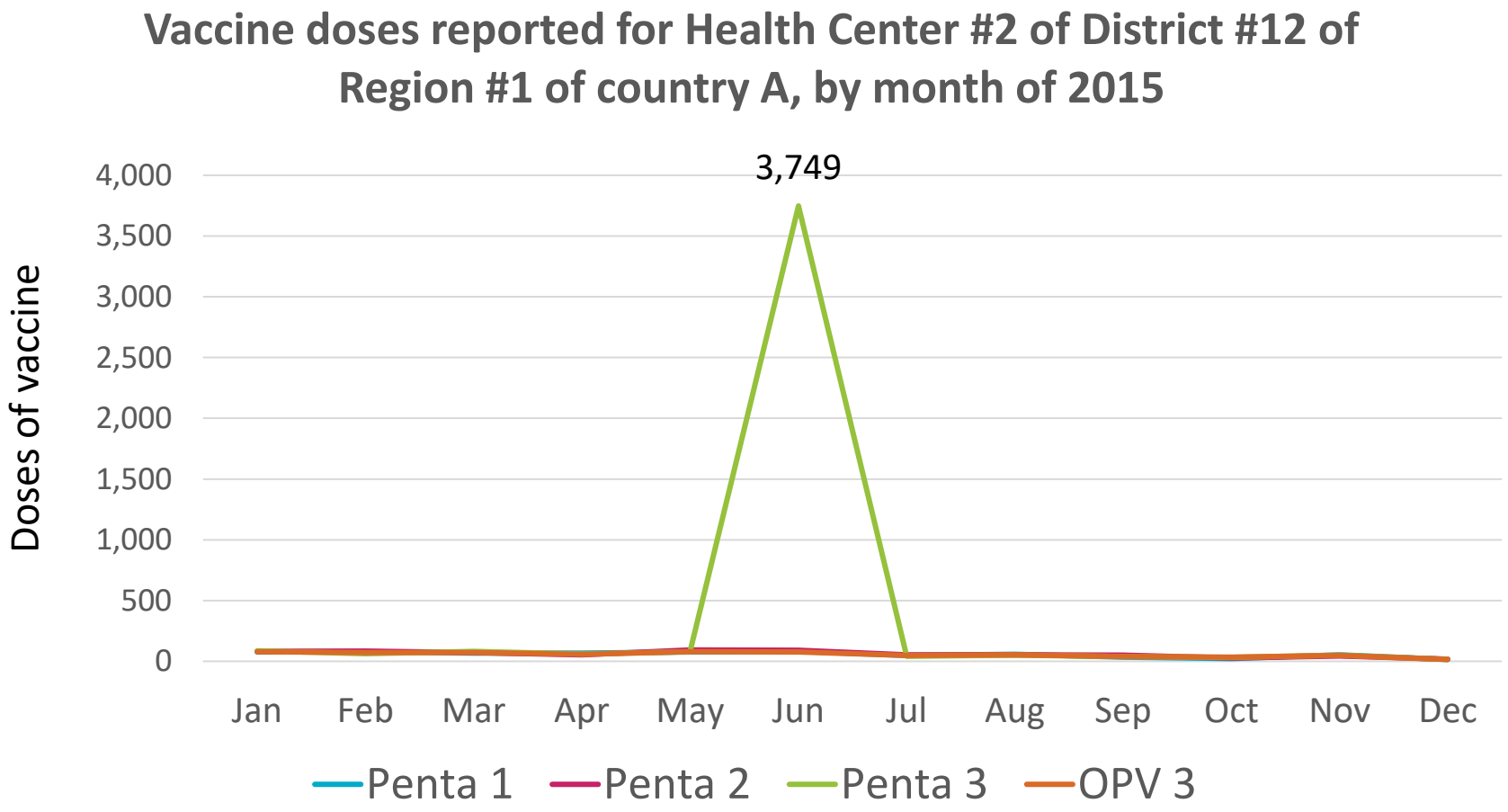


SESSION 4

Domain 2: Internal Consistency

Outliers

Domain 2a: by further "drilling down" to the facility-level data, we can often show that an extreme outlier is an error



SESSION 4

Domain 2: Internal Consistency

Outliers

Domain 2a: WHO's Excel-based DQR Tool

- identifies the districts with outliers; and
- invites users to plan follow-up actions

Indicator 2a.1: Extreme Outliers (>3 SD from the mean)		2014		
	National score	Districts with extreme outliers relative to the mean		
Program Area and Indicator	%	No.	%	Name
Maternal_Health - ANC 1st Visit	0.0%			-
Immunization - 3rd dose DPT-containing vaccine	0.1%	1	1.3%	District 5
General_Service_Statistics - OPD new curative consultations	0.0%			-
Maternal_Health - Institutional Deliveries	0.1%	1	1.3%	District 18
Immunization - 1st dose DPT-containing vaccine	0.0%			-
Maternal_Health - ANC 4th Visit	0.0%			-
Total (all indicators combined)	0.0%			

Interpretation of results - Indicator 2a1:

- DTP3 reported in August 2014 by District 5 (741) was 60% greater than the value reported in other months. DTP3 reported that month by Clinic #2 (363) of District #5 was 4 times the number reported by the same facility in other months and 4 times the number of doses of other vaccines (DTP1, DTP2, OPV1, OPV2, OPV3) reported in the same month. Contact the District Health Information Officer (Mr. X) and ask him to investigate.
- "Outlier" in District #18 reporting of institutional deliveries is due to the health facilities in that district failing to report any deliveries in 2014 or in 2015. Contact the District Medical Officer and the District Health Information Officer to ask them why almost no deliveries have been reported.

SESSION 4

Domain 2: Internal Consistency

Outliers

SESSION 4

Domain 2: Internal Consistency

Domain 2a: WHO's Data Quality Tool rapidly identifies extreme outliers in DHIS2 facility-level data and ranks the outliers in order of importance

WHO Data Quality Tool | Dashboard | Analysis | Annual Review | More

Unit	Data	Jan 14	Feb 14	Mar 14	Apr 14	May 14	Jun 14	Jul 14	Aug 14	Sep 14	Oct 14	Nov 14	Dec 14
Kawe dispensary	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	26.0	20.0	30.0	60.0	19.0	4647.0	24.0		7.0	18.0	20.0	26.0
Katesh Health Center	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	30.0	4345.0	53.0		54.0	63.0		32.0	37.0	36.0	34.0	38.0
ST. Aloyce Health Center	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	25.0	23.0	23.0		33.0	34.0	22.0	26.0	3432.0	35.0	21.0	49.0
RC/K/Ndege Dispensary	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	19.0	3021.0	18.0	15.0	28.0	35.0	22.0	42.0	36.0	32.0		39.0
Mlali Health Center	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	13.0	16.0	1710.0	17.0	13.0	17.0	9.0	12.0	6.0	12.0		
Kandashi Dispensary	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	17.0	13.0	24.0	15.0	14.0	20.0	16.0	1328.0	15.0	14.0	26.0	18.0
Balang`a Dispensary	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	1.0	6.0	13.0	5.0	3.0	11.0	9.0		1212.0	11.0	13.0	27.0
Ruanda Health Center	Penta vaccines given (KE, Under 1, Dose 3, Inside Service Area)	178.0	151.0	171.0	143.0	136.0	168.0	155.0	188.0	1110.0	121.0		169.0

Consistency over time

SESSION 4

Domain 2: Internal Consistency

Domain 2: Internal Consistency of Reported Data

Focus

- Plausibility (apparent accuracy) of the data

Metrics

- a) Presence of "outliers" (suspicious values)
- b) Consistency from year to year
- c) Consistency of related indicators (e.g. Penta 1 vs Penta 3)
- d) Verification factor (a facility survey) -- consistency between source data (e.g. on clinic registers) and reported data

Consistency over time

SESSION 4

Domain 2: Internal Consistency

Internal Consistency of Reported Data—2

Data Quality Metric	Definition	
	National Level	Subnational Level
Consistency over time (analyze each indicator separately)	Conduct 1 of the following based on expected trend of the indicator: ~ Indicators or programs with expected growth - Comparison of current year to the value predicted from the trend in the 3 preceding years ~ Indicators or programs expected to remain constant - Comparison of current year to the average of 3 preceding years	# (%) of districts whose current year to predicted value ratio (or current year to the average of the preceding 3 years) is at least ± 33% different from the national ratio
	Graphic depiction of trend to determine plausibility based on programmatic knowledge	

Consistency over time

SESSION 4

Domain 2: Internal Consistency

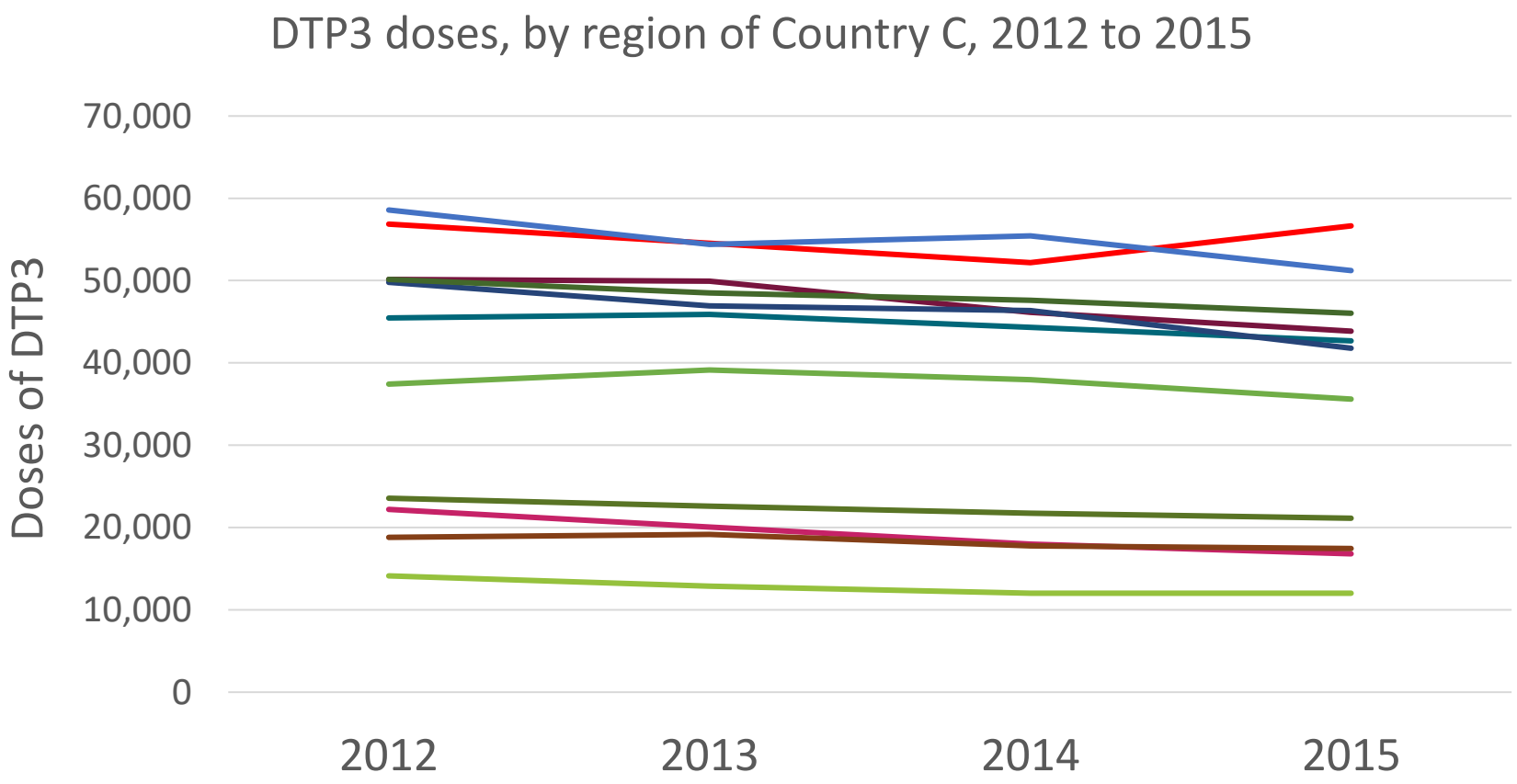
Example — Consistency Over Time

Consistency trend: Comparison of district ratios to national ratios. Any difference between district and national ratio that is ≥33% is highlighted in red.

District	Year				Mean of 2010-2012	Ratio of 2013 to Mean of 2010-2012	% Difference between National & District Ratios
	2010	2011	2012	2013			
A	30242	29543	26848	32377	28878	1.12	0.03
B	19343	17322	16232	18819	17632	1.07	0.08
C	7512	7701	7403	7881	7539	1.05	0.09
D	15355	15047	14788	25123	15063	1.67	0.44
E	25998	23965	24023	24259	24662	0.98	0.16
National	**98450**	**93578**	**89294**	**108459**	**93774**	**1.16**	

Consistency over time

Domain 2b: inconsistency from year to year may be subtle with regional data

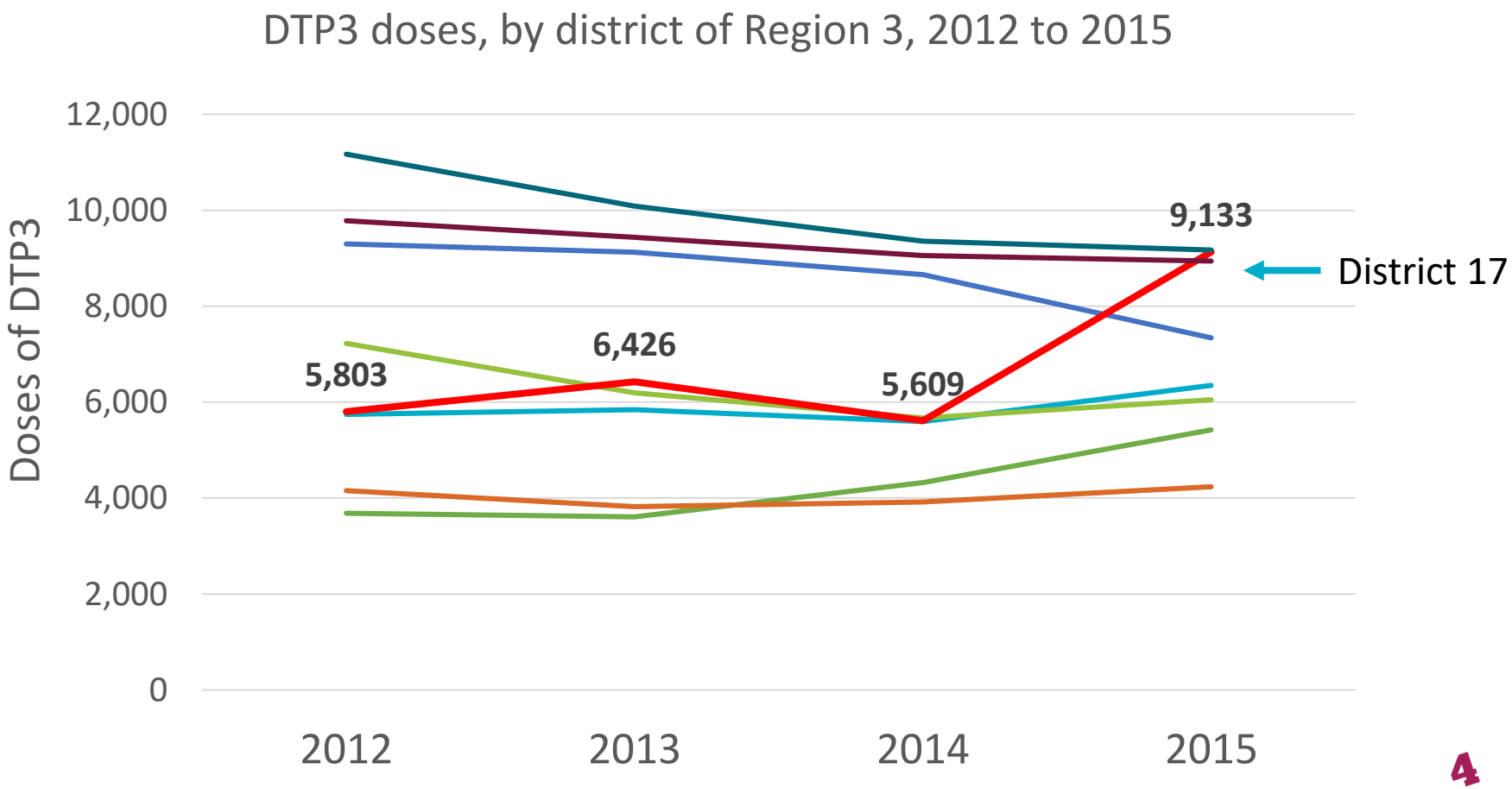


SESSION 4

Domain 2: Internal Consistency

Consistency over time

Domain 2b: inconsistency from year-to-year is usually more apparent with district-level data



SESSION 4

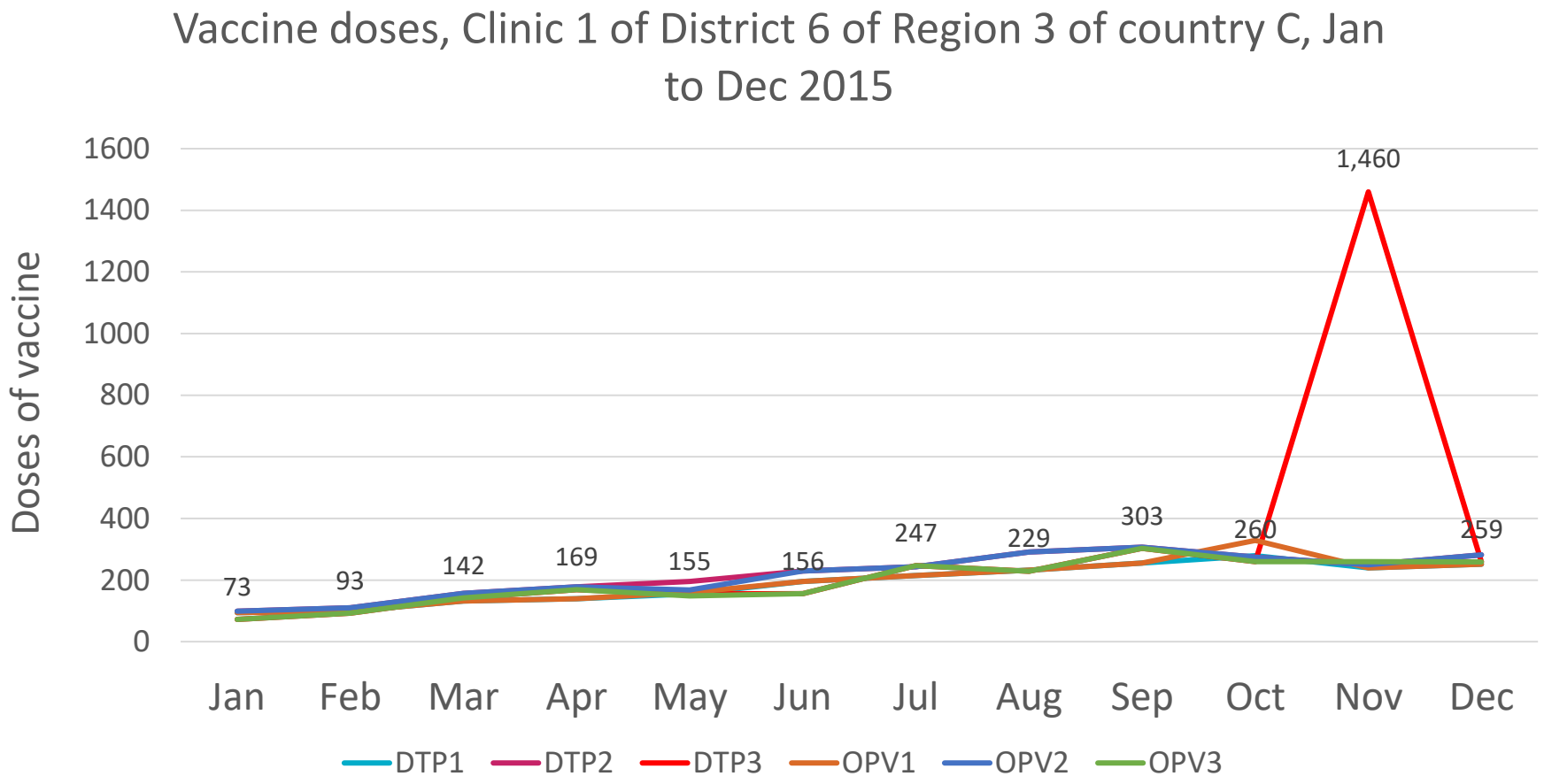
Domain 2: Internal Consistency

Consistency over time

SESSION 4

Domain 2: Internal Consistency

Domain 2b: "drilling down" to the facility-level data can reveal that a marked year-to-year inconsistency is an error



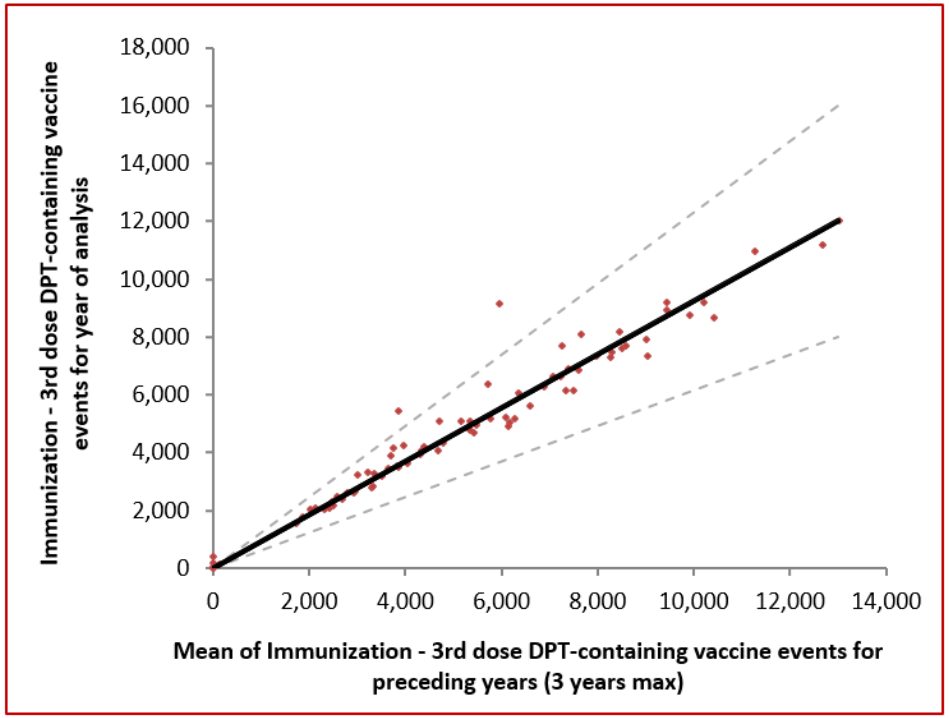
Consistency over time

SESSION 4

Domain 2: Internal Consistency

Domain 2b: WHO's Excel-based DQR Tool identifies the districts with inconsistency from year-to-year

2b2: Consistency of 'Immunization - 3rd dose DPT-containing vaccine' over time	
Year	2015
Expected trend	Constant
Compare districts to:	national result
Quality threshold	33%
National score (%)	93%
Number of districts with divergent scores	2
Percent of districts with divergent scores	3%
Names of districts with divergent scores:	
	District 12, District 17



Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

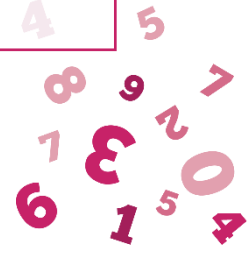
Domain 2: Internal Consistency of Reported Data

Focus

- Plausibility (apparent accuracy) of the data

Metrics

- a) Presence of "outliers" (suspicious values)
- b) Consistency from year to year
- c) Consistency of related indicators (e.g. Penta 1 vs Penta 3)
- d) Verification factor (a facility survey) -- consistency between source data (e.g. on clinic registers) and reported data



Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

Internal Consistency of Reported Data—3

Data Quality Metric	Definition	
	National Level	Subnational Level
Consistency between related indicators	**Maternal Health**: ANC1 - IPT1 or TT1 (should be roughly equal)	# (%) of subnational units where there is an extreme difference (≥ ± 10%)
	Immunization: DTP3 dropout rate = (DTP1 - DTP3)/DTP1 (should not be negative)	# (%) of subnational units with # of DTP3 immunizations > DTP1 immunizations (negative dropout)
	HIV/AIDS: ART coverage - HIV coverage (should be <1)	# (%) of subnational units where there is an extreme difference (≥ ± 10%)
	TB: TB cases notified - TB cases on treatment (should be roughly equal)	# (%) of subnational units where there is an extreme difference (≥ ± 10%)
	Malaria: # confirmed malaria cases reported - cases testing + (should be roughly equal)	# (%) of subnational units where there is an extreme difference (≥ ± 10%)

Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

Domain 2c:
Some examples of related indicators

- DTP1 & DTP3 (dropout)
- DTP1 & OPV1
- ANC1 & IPTp1
- Confirmed malaria cases (from the OPD report) & positive tests for malaria (from the lab report)
- ANC1 & DTP1 (when ANC coverage and immunization coverage are both > 90%)

Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

Example — Consistency Between Related Indicators

% difference between ANC1 and IPT1 by district. Districts with % difference ≥10% are flagged in red.

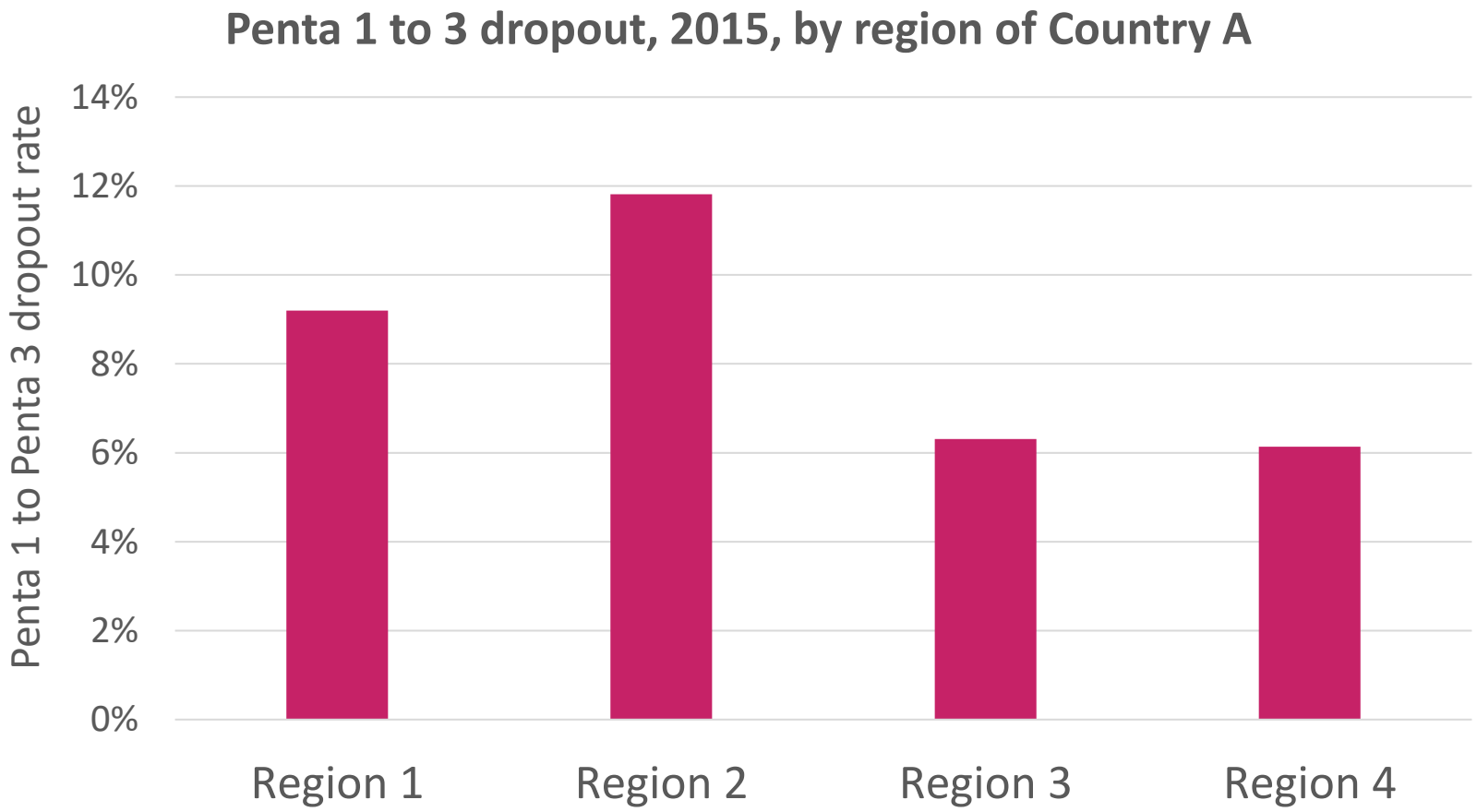
District	ANC1	IPT1	Ratio of ANC1 to IPT1	% Difference between National & District Ratios
A	20995	18080	1.16	0.02
B	18923	16422	1.15	0.02
C	7682	6978	1.10	0.07
D	12663	9577	1.32	0.12
E	18214	15491	1.18	0
National	**78477**	**66548**	**1.18**	

Consistency between related indicators

SESSION 4

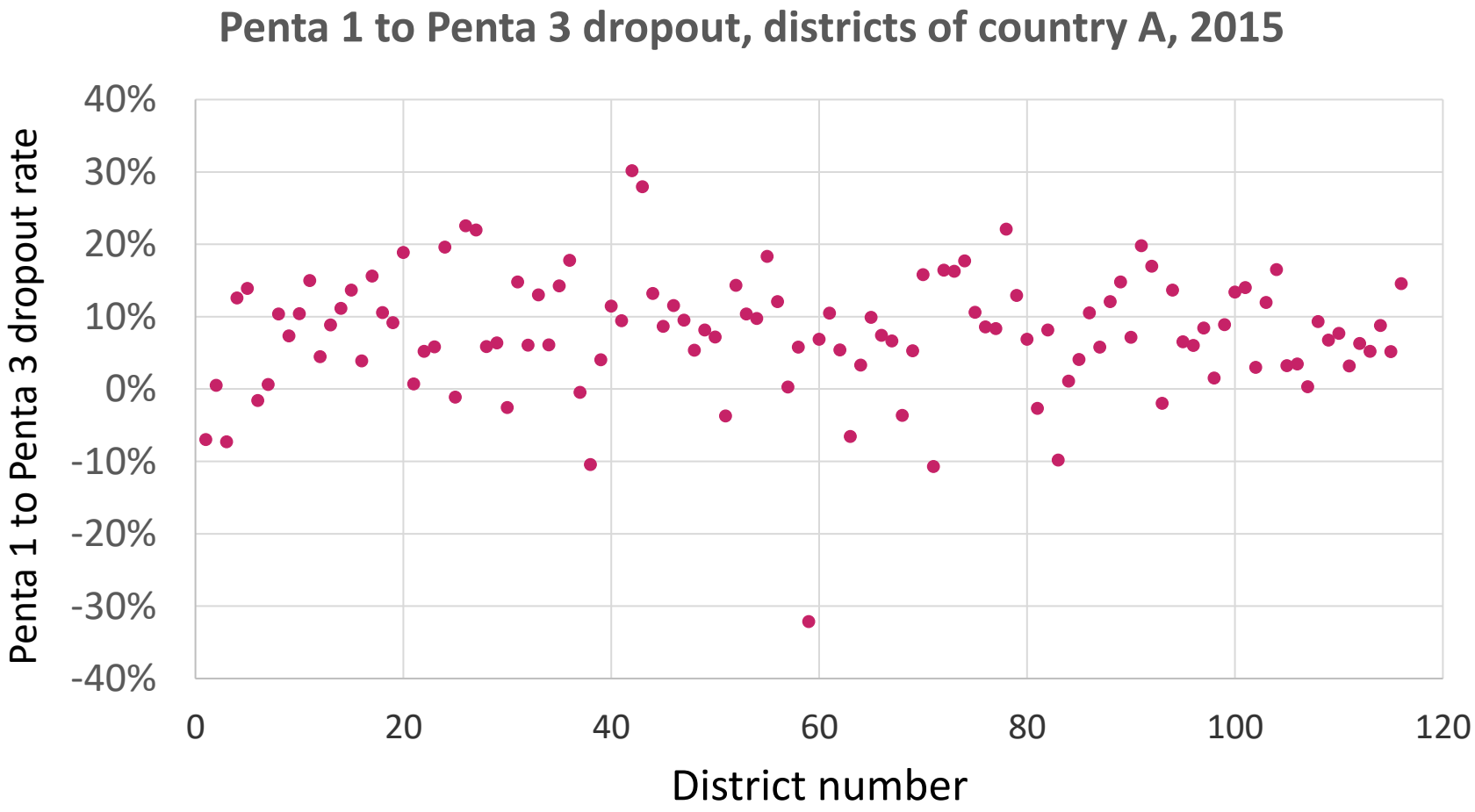
Domain 2: Internal Consistency

Domain 2b: regional data may not show inconsistency between related indicators



Consistency between related indicators

Domain 2c: inconsistency of related indicators is usually more apparent with district-level data



SESSION 4

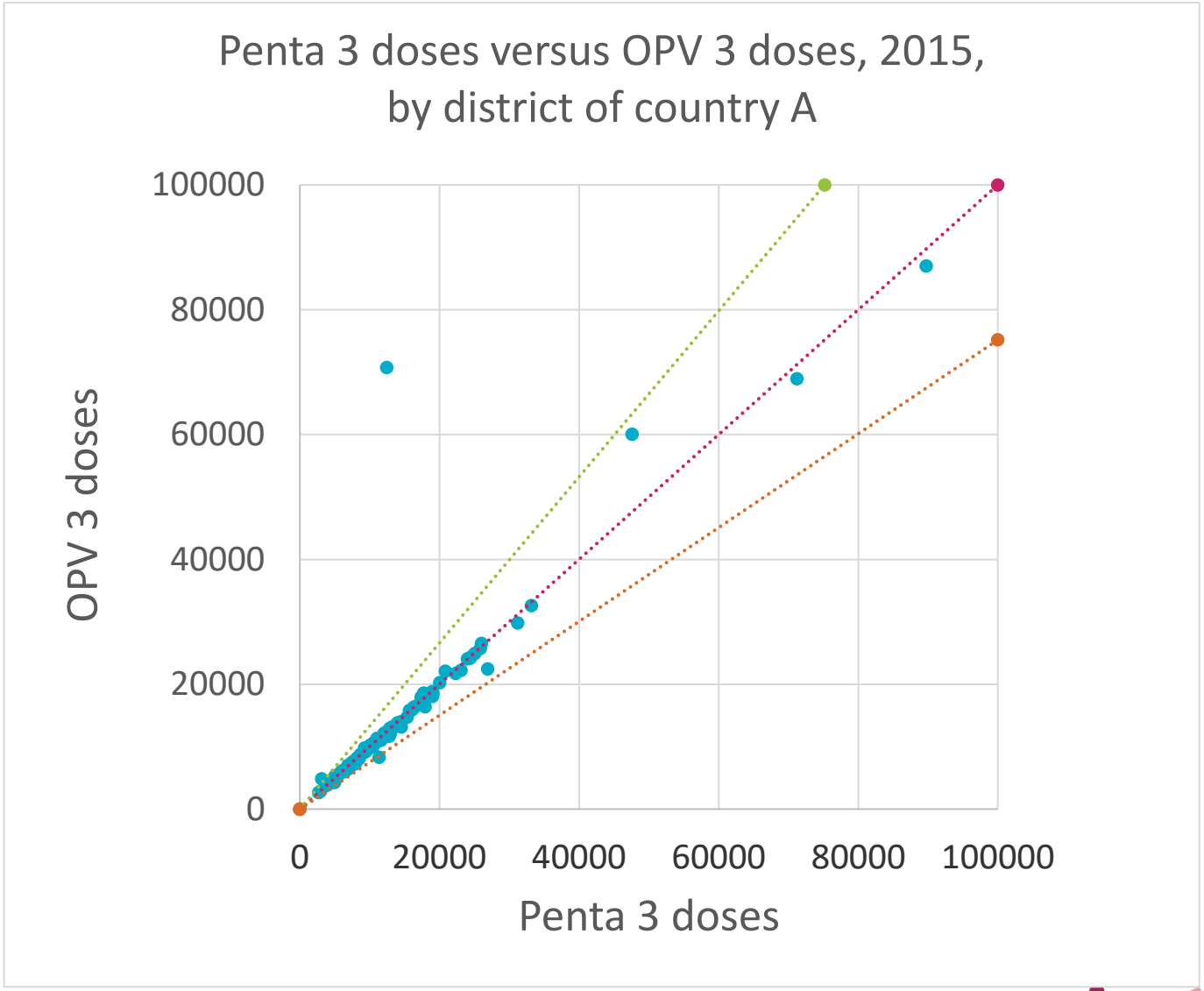
Domain 2: Internal Consistency

Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

Domain 2c: inconsistency between related indicators is usually more apparent with district-level data



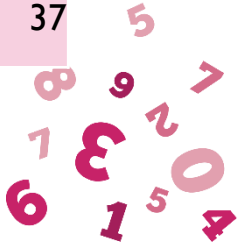
Consistency between related indicators

SESSION 4

Domain 2: Internal Consistency

Domain 2c: "drilling down" shows that the discrepant district values are due to erroneous values from one facility

Doses of vaccine reported from Health Centre 2 of District 5 of Country A, 2015												
Vaccine	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Penta 1	43	44	24	13	25	11	18	37	47	70	29	42
Penta 2	20	27	23	26	18	13	18	16	49	58	26	44
Penta 3	35	25	21	23	39	9	15	15	57	44	24	39
OPV 1	27	55	55	13	34	36	33	35	50	35	24	18
OPV 2	20	49	66	26	61	32	13	34	58	50	34	31
OPV3	38,707	4,023	45	23	31	11	18	45	43	71	14	37



Data verification

SESSION 4

Domain 2: Internal Consistency

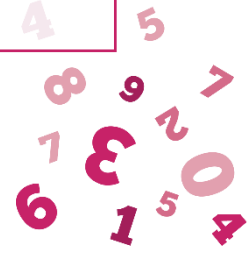
Domain 2: Internal Consistency of Reported Data

Focus

- Plausibility (apparent accuracy) of the data

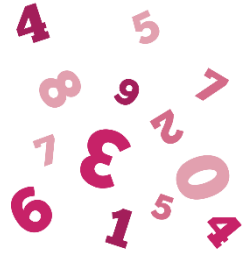
Metrics

- a) Presence of "outliers" (suspicious values)
- b) Consistency from year to year
- c) Consistency of related indicators (e.g. Penta 1 vs Penta 3)
- d) Verification factor (a facility survey) -- consistency between source data (e.g. on clinic registers) and reported data



Internal Consistency of Reported Data—4

Data Quality Metric	Definition	
	National Level	Subnational Level
Verification of reporting consistency through facility survey	% agreement between verified counts for selected indicators in sampled facility records and reported values for the same facilities	**Maternal Health**: ANC 1st visit
		Immunization: Penta/DTP 1-3 in children <1 year
		HIV/AIDS: ART coverage
		TB: Notified cases of all forms of TB
		Malaria: Confirmed malaria cases



Data verification

SESSION 4

Domain 2: Internal Consistency

Example — Verification of Reporting Consistency

Consistency of indicator reporting by district. Indicators with verification factors ≥ ±10% of 1 are flagged in red.

District	Indicator 1			Indicator 2		
	Recounted	Reported	VF	Recounted	Reported	VF
A	1212	1065	1.14	4009	4157	0.96
B	1486	1276	1.16	3518	3686	0.95
C	357	387	0.92	672	779	0.86
D	2987	3849	0.78	1361	1088	1.25
E	4356	4509	0.97	4254	3970	1.07

Discussion

SESSION 4

Domain 2: Internal Consistency

Discussion Questions:

- What constitutes an outlier in routine reporting of public health information systems?
  - What would you do about it if you found one in your dataset?
- What does the trend in the indicator over time tell you about the indicator?
  - What is the expected trend?
  - How can comparing a recent trend to a historical trend inform you about data quality?
- Name some examples of pairs of indicators that have a predictable relationship.
  - How can this relationship be used to judge data quality?
  - If the expected relationship between two indicators is not found what are some plausible reasons to explain the observed relationship?

Exercise

Exercise – Internal Consistency of data

- Navigate to the DQ Demo instance of DHIS2: https://demos.dhis2.org/dq/dhis-web-commons/security/login.action

SESSION 4

Domain 2: Internal Consistency