R workshop

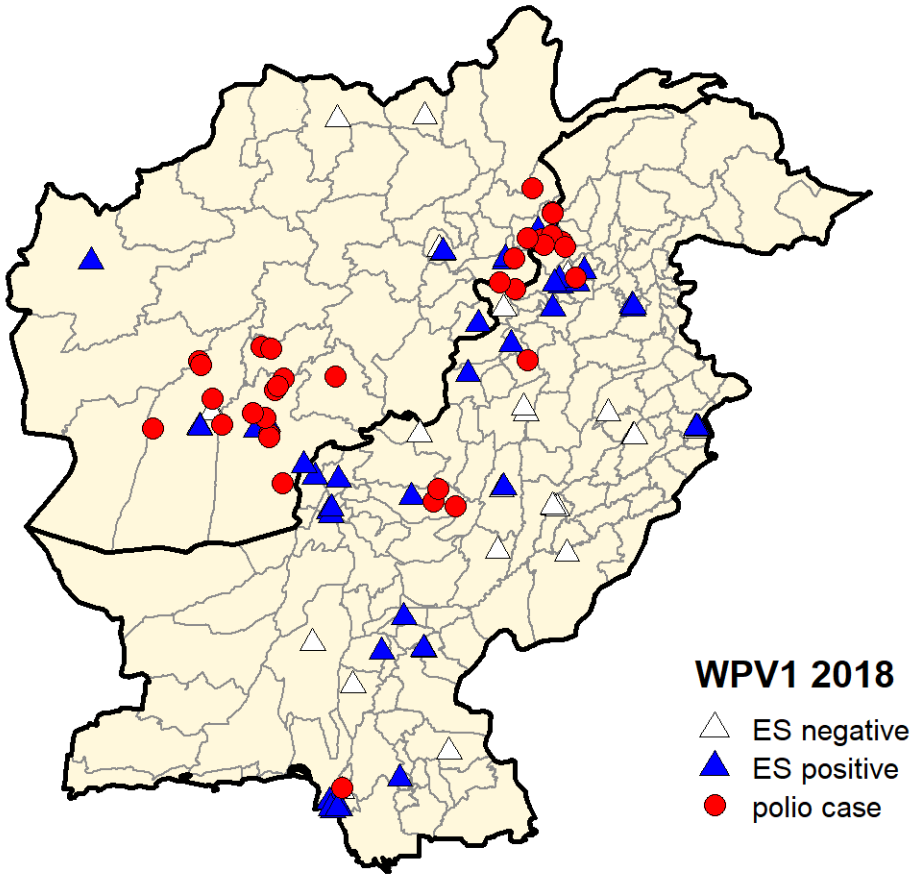
- The next few days will introduce you to using R and Rstudio to interact with and plot data.
- The data we will work with later in the week is related to prediction of the risk of poliovirus in Pakistan.
- The next few slides will introduce how this data is used in risk assessments.

Risk factors and short-term projections for serotype-1 poliomyelitis incidence in Pakistan: A spatiotemporal analysis

Molodecky et al. 2017 Plos. Med.

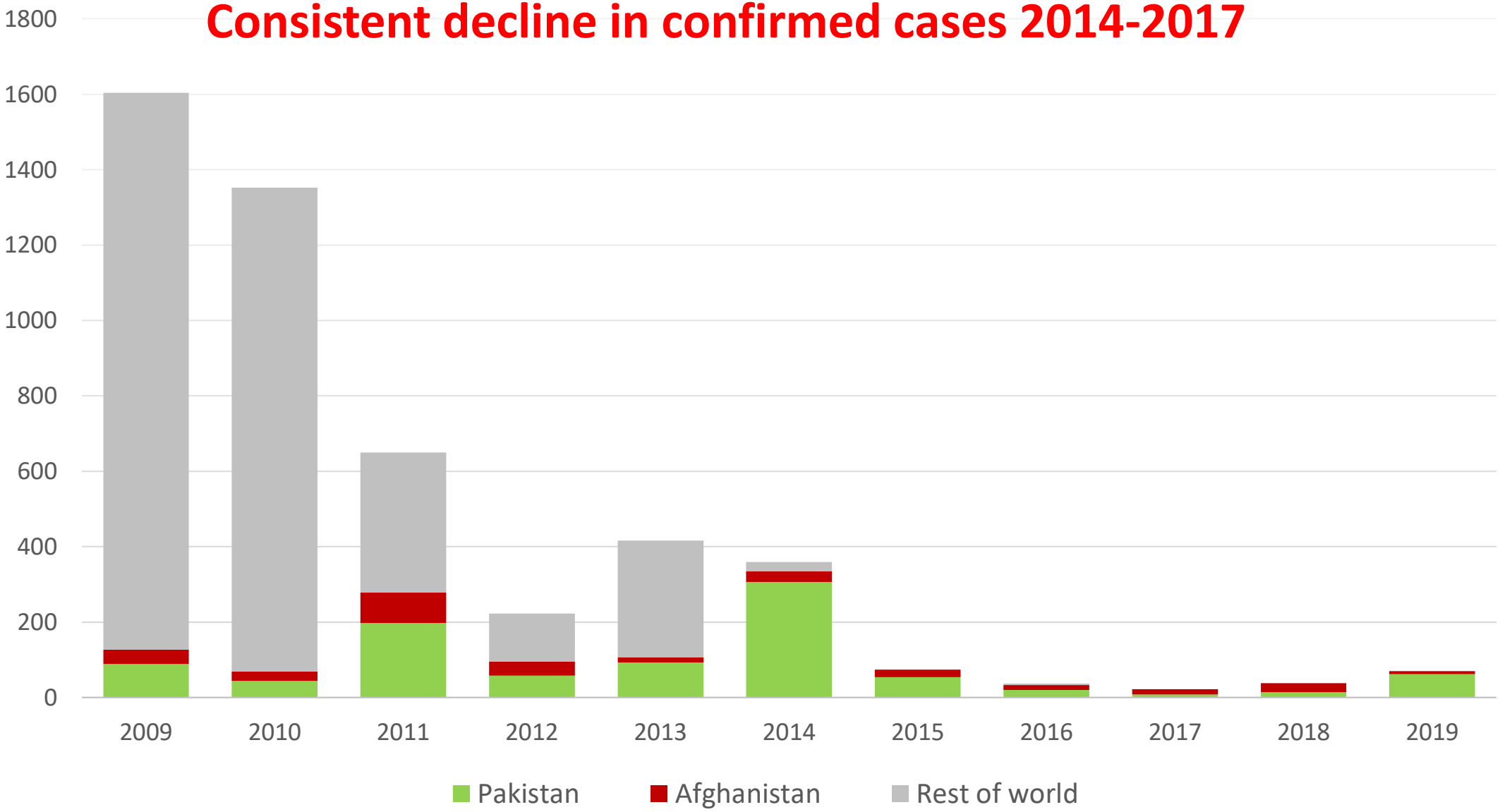
Background

- Pakistan and Afghanistan represent the last remaining cluster of wild poliovirus transmission globally. No cases have been seen in Nigeria since 2016
- RI and SIAs are the key interventions in Pakistan to interrupt transmission
- Accessibility issues are known – estimating risk in inaccessible areas is hard but they also represent risk of transmission to surrounding areas
- OPV immunogenicity poorer in Pakistan – multiple doses needed
- Targeting interventions important to maximise utility of doses



Global WPV1 Cases, 2009–2019*

Consistent decline in confirmed cases 2014-2017



Updated from presentation by Natalia Molodecky, WHO

Model aim

- This work aimed to include movement data in polio risk models to see if polio spread could be better estimated.
- Risk models are used to plan SIAs so accuracy is important to the interruption of WPV transmission.

Basic parameters

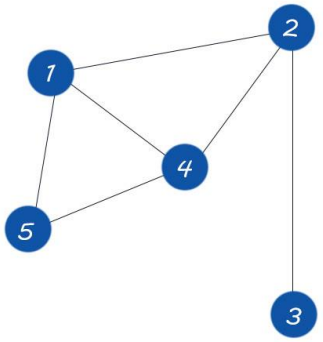
- Vaccination history of AFP cases – reported vaccine doses of non-polio AFP cases divided by number of SIA campaigns in their district between birth and paralysis
- RI coverage - % of children receiving 3 RI doses of OPV

Pop$^{\underline{n}}$ immunity

- Non-polio AFP rate – marker of surveillance intensity
- Cases in the same spatial unit
- Population and density of each spatial unit – number of individuals important to risk of spread
- Population living in poverty
- Precipitation

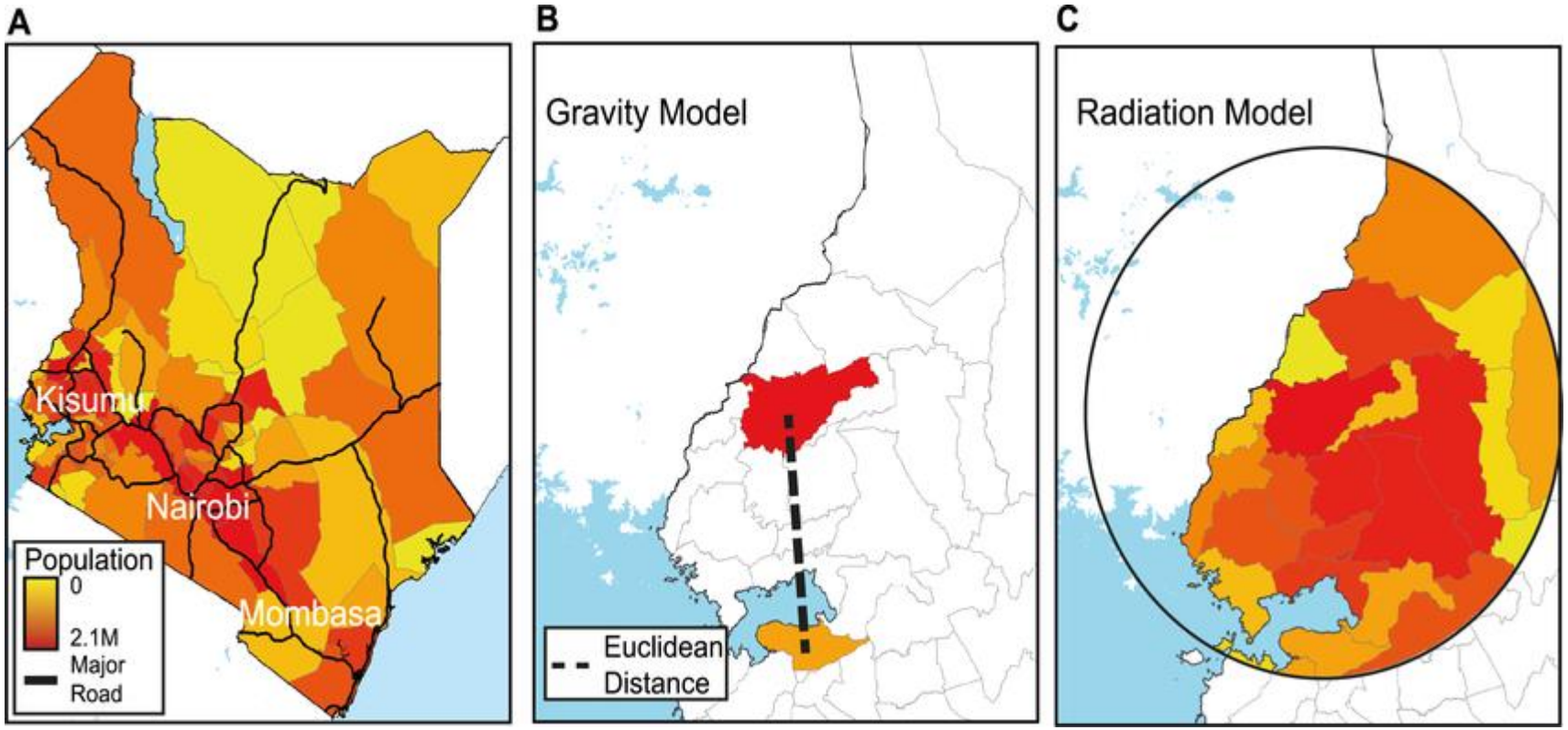
Movement models

First order adjacency



$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Gravity and radiation models



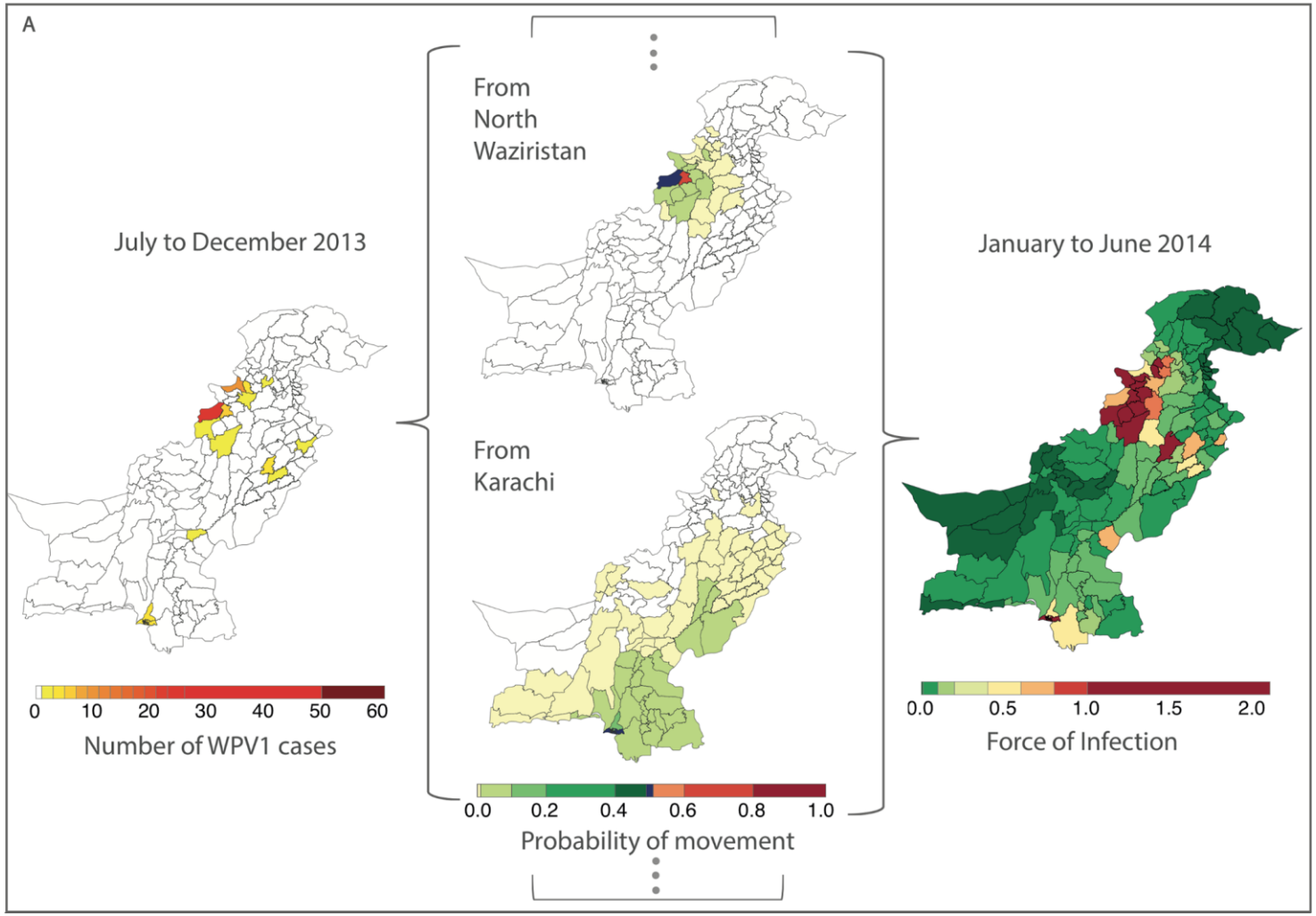
Wesolowski et al. Plos Comp. Biol. 2015

Movement models

Extensions to these movement models incorporated:

- Population density in the radiation model – a small district with high population density is more likely to attract a migrant than a large district with the same population.
- Travel time in the radiation model – rather than direct distance we can use road network maps etc to estimate real world travel time.
- Mobile phone data in the gravity model – mobile phone data gives an indication of where people are travelling in reality to fit the model to.

Force of infection



Force of infection

"The rate at which susceptible individuals acquire disease"

This depends on a number of factors such as likelihood of contacting infected individuals, probability of transmission on contact and the number of infected individuals.

Therefore population, number of local infections, total number of infections, vaccination coverage etc. all contribute to risk of disease.

Results

Table 1. Risk factors associated with the incidence of wild poliovirus type 1 (WPV1) cases based on the best-fitting multivariable mixed-effects lagged regression model for January–June 2010 through July–December 2016. The odds ratio (OR) and the 95% confidence interval (CI) for routine immunization and supplementary immunization activity (SIA) coverage are for an absolute 10% increase in these variables and a 1-unit increase for all other variables. Non-polio acute flaccid paralysis (AFP) rate is per 100,000 persons aged <15 years.

Variable (fixed effects)	**OR (95% CI)**	***P* value**
Routine immunization coverage (previous 6 months)	0.75 (0.67–0.84)	<0.001
SIA coverage (previous 6 months)	0.75 (0.66–0.85)	<0.001
Non-polio AFP rate (previous 6 months)	1.13 (1.02–1.26)	0.025
Log (population size)	2.62 (1.94–3.55)	<0.001
Cases in the same district (previous 6 months)	1.16 (1.04–1.28)	0.006
Cases in all other districts, weighted by probability of movement (previous 6 months; radiation model)	1.14 (1.02–1.27)	0.021
Variable (random intercepts)	**Variance**	**Standard Deviation**
Province	0.393	0.627
Year (6-month interval)	0.838	0.915

Interpreting these odds ratios

Measure of association between an exposure and an outcome.

These represent the change in the odds of a polio case depending on a 10% change in RI or SIA coverage or a 1 unit change in the other parameters.

Odds below 1 indicate a negative association where an increase in exposure leads to a decrease in the effect of interest.

Table 1. Risk factors associated with the incidence of wild poliovirus type 1 (WPV1) cases based on the best-fitting multivariable mixed-effects lagged regression model for January–June 2010 through July–December 2016. The odds ratio (OR) and the 95% confidence interval (CI) for routine immunization and supplementary immunization activity (SIA) coverage are for an absolute 10% increase in these variables and a 1-unit increase for all other variables. Non-polio acute flaccid paralysis (AFP) rate is per 100,000 persons aged <15 years.

Variable (fixed effects)	**OR (95% CI)**	***P* value**
Routine immunization coverage (previous 6 months)	0.75 (0.67–0.84)	<0.001
SIA coverage (previous 6 months)	0.75 (0.66–0.85)	<0.001
Non-polio AFP rate (previous 6 months)	1.13 (1.02–1.26)	0.025
Log (population size)	2.62 (1.94–3.55)	<0.001
Cases in the same district (previous 6 months)	1.16 (1.04–1.28)	0.006
Cases in all other districts, weighted by probability of movement (previous 6 months; radiation model)	1.14 (1.02–1.27)	0.021
Variable (random intercepts)	**Variance**	**Standard Deviation**
Province	0.393	0.627
Year (6-month interval)	0.838	0.915

Why were these parameters important?

Routine immunisation coverage

Routine immunisation coverage

- Negative association
- This represents both a portion of population immunity to poliovirus alongside SIA coverage and a measure of the efficiency of the health systems in the area to provide 3 doses of vaccine to children.
- Together with SIAs, low coverage can be an indicator of vaccine refusal

SIA coverage

SIA coverage

- Negative association
- This association suggests that greater SIA coverage leads to a decrease in risk of poliovirus in the population.
- This suggests that SIAs carried out in response to higher risk are effective in reducing this risk.

Non-polio AFP rate

Non-polio AFP rate

NPAFP rate is a marker of:

- Sampling rate for poliovirus cases – there is increased chance of reporting a case where more surveillance is being carried out.
- NPAFP reporting also tends to be higher in areas where recent campaigns have been carried out due to increased surveillance by teams in the region.

Log population size

Log population size

- Log is taken of population size to decrease the magnitude of change between urban and rural areas, this allows odds ratios to be calculated more easily.
- Association of population size with polio risk is expected as there is more likelihood of being in contact with an infected person or of an infected person coming to your region if it is a major population centre.
- Sewage systems and sanitation in large urban areas often also serve a larger population leading to increased risk of contact with contaminated water if the sewage network is not contained

Cases in the same district

Cases in the same district

- Cases in the same district represent local risk of disease, you are more likely to come in to contact with an individual or water contaminated by an individual if they are in close proximity to you.

Cases in other districts

Cases in other districts

- As the whole country is considered one population, there is an increased risk for everyone in the population when cases are reported.
- As there is movement between populations it is possible for long distance transmissions to occur.
- It is also possible that cases in bordering districts represent short distance transmissions.

Updated analysis on 2017 data

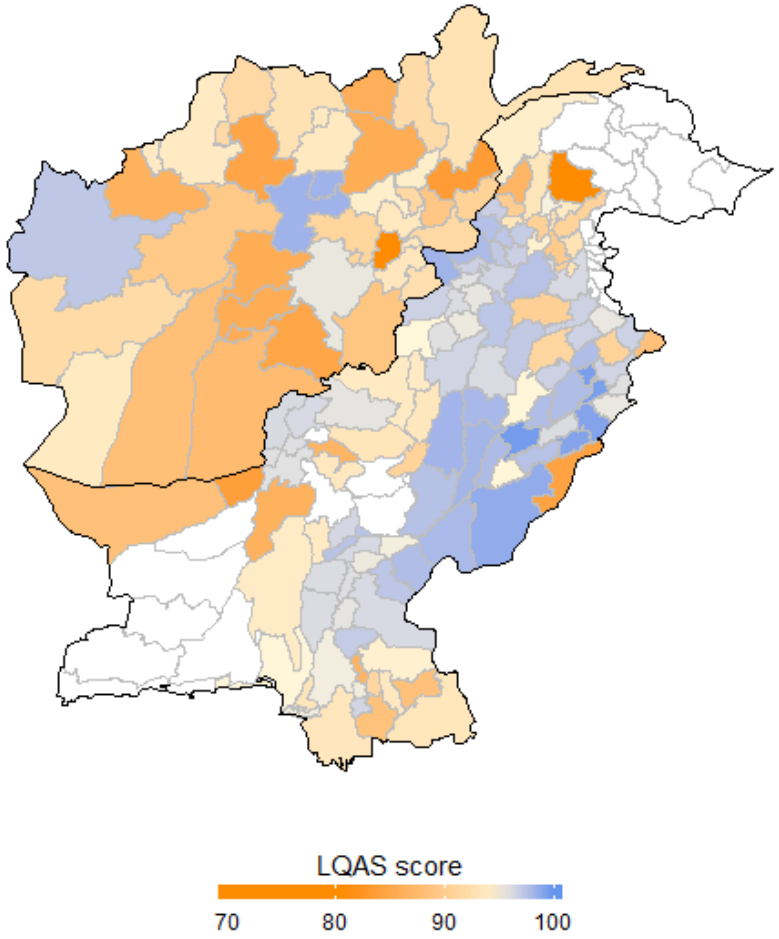
David Jorgensen

2017 analysis

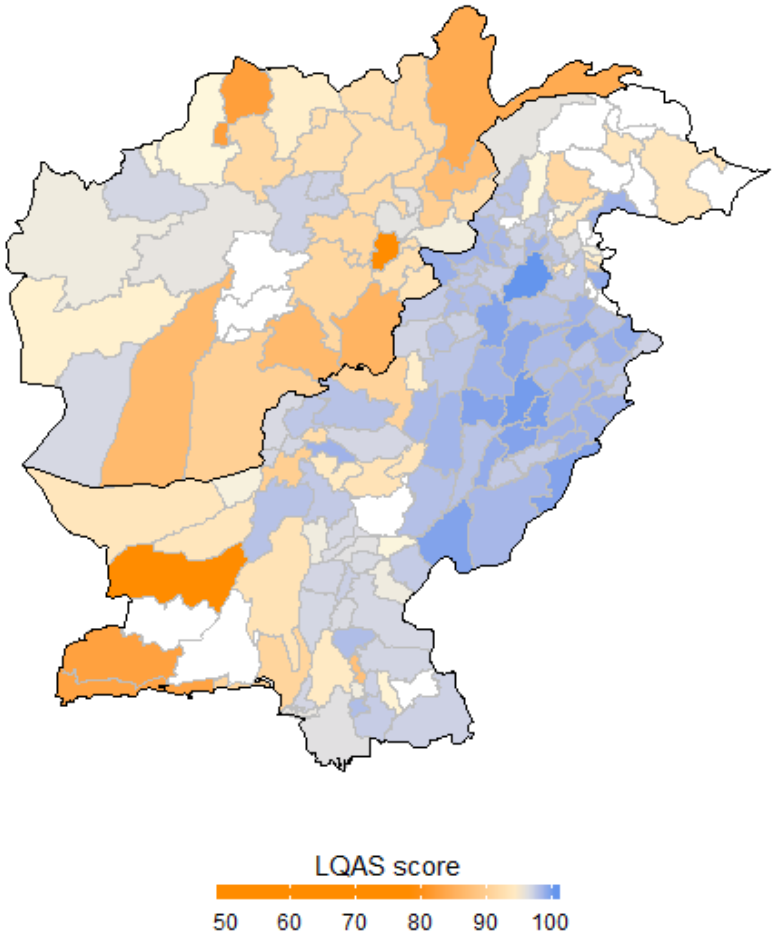
- Included Afghanistan as key risk predictions for border regions are likely to be incorrect when not accounting for cross-border transmission of virus.
- Included LQAS data alongside NPAFP estimates of vaccination coverage.

LQAS Data

Aggregated LQAS scores for 01-07-2016 to 31-12-2016



Aggregated LQAS scores for 01-07-2017 to 31-12-2017



Data

The models were fitted on binary incidence over time 01-01-2013 to 30-12-2017

- SIA coverage data:
  - LQAS – discretised using the median (Previous 6 months)
  - AFP – Split in to quartiles (Forward 6 months)
- Other parameters from previous work
  - NPAFP rate (Previous 6 months)
  - RI coverage (Previous 6 months)
  - Cases in the same district (Previous 6 months)
  - Population size (worldpop 2015 estimate)
  - Between district force of infection (Previous 6 months)
- BYM model with neighbour joining matrix
- Random walk over time (province, region and global)

Results: 01-01-2017 to 30-06-2017

		Univariable model		Multivariable model	
Parameter	**Group**	**OR**	**95% CrI**	**OR**	**95% CrI**
SIA coverage (Banded LQAS) previous 6mo	1 (0-94%)	3.42	2.16-5.53	2.20	1.27-3.86
	2 (95-100%)	Ref	-	Ref	-
SIA coverage (Banded AFP) (+6mo)	1 (0-56%)	9.97	4.53-24.46	4.07	1.64-11.08
	2 (57-68%)	2.32	0.98-5.99	1.82	0.70-5.12
	3 (69-80%)	0.95	0.36-2.59	0.85	0.31-2.45
	4 (81-100%)	Ref	-	Ref	-
RI coverage previous 6mo	1 (0-50%)	7.99	4.07-16.93	6.16	2.55-15.78
	2 (51-75%)	3.70	1.79-8.13	3.10	1.38-7.38
	3 (76-80%)	2.41	0.85-6.52	1.72	0.57-5.02
	4 (81-100%)	Ref	-	Ref	-
npAFP rate previous 6mo*	-	1.15	1.07-1.23	1.07	0.99-1.15
Cases in the same district previous 6mo*	-	5.70	3.39-5.70	1.94	1.05-3.86
Log(population size)*	-	2.48	1.60-4.04	2.01	1.46-2.81
Force of infection previous 6mo*	-	1.35	1.14-1.60		

Issues with these risk analyses

- Models are always a simplification of reality, a number of other factors for which data is not available likely also contribute to risk of polio cases.

- Poliomyelitis cases, reported in these analyses, represent only a small fraction of those infected with poliovirus, due to the likelihood of contracting paralytic disease polio cases are usually clustered in the regions of high transmission but transmission is able to persist in other regions through asymptomatic individuals and the environment.

All of these risk analyses were carried out in R!

- The next few days will teach you the basics of R up to plotting specific polio data from Pakistan. These risk analyses show how the data we will be plotting contribute to the risk of polio infection.

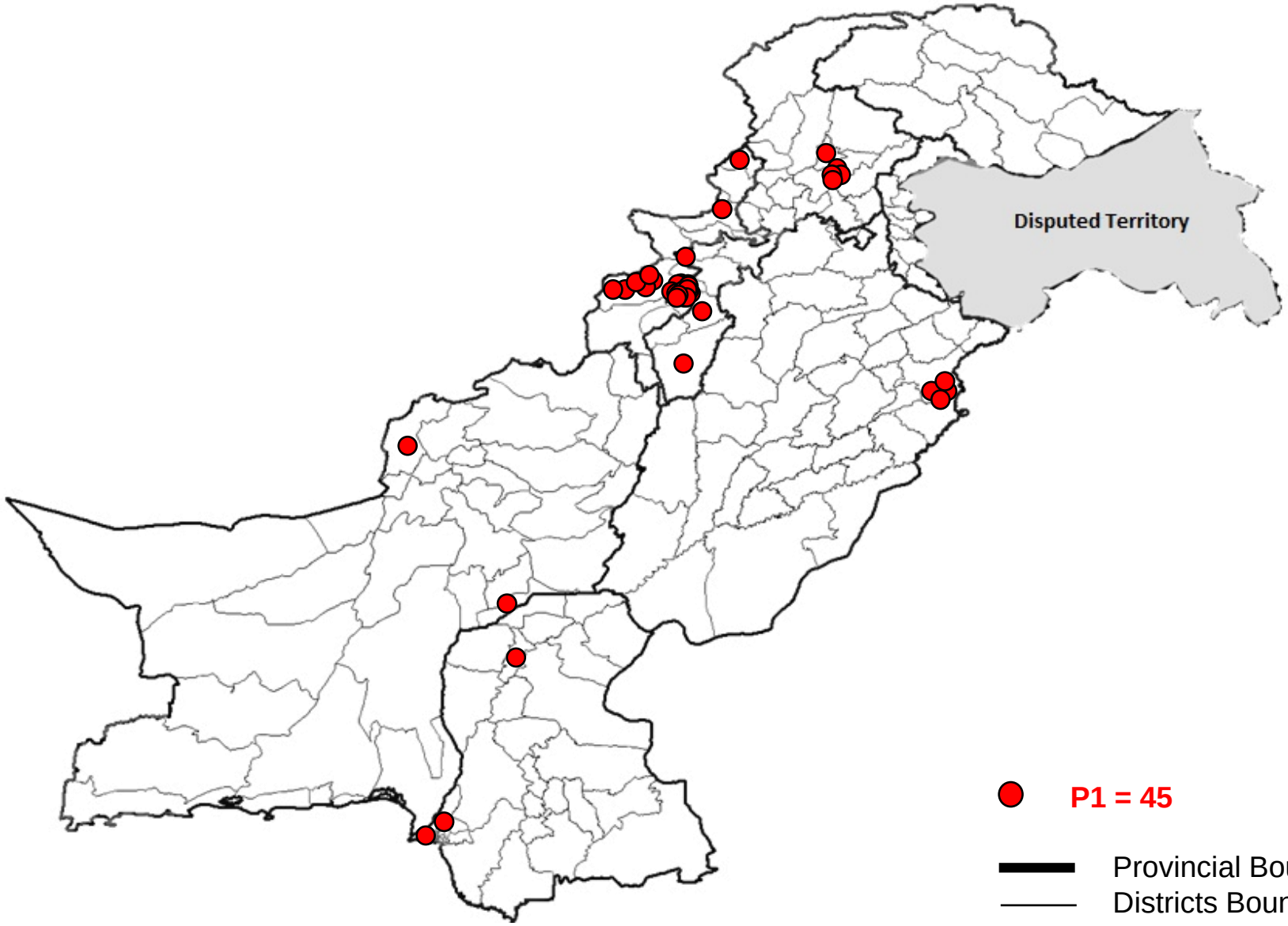
- Being able to clean and plot the input data for these type of models is half the work. Matching your data columns, aggregating data and visualising it give a good indication of specific risks without running these complex, national level models.

Where we are going with this course

- The majority of the course will focus on basic plotting and data handling with R, towards the 3rd day you will start to see these skills come together to recreate plots with important predictors of polio risk.

Surveillance

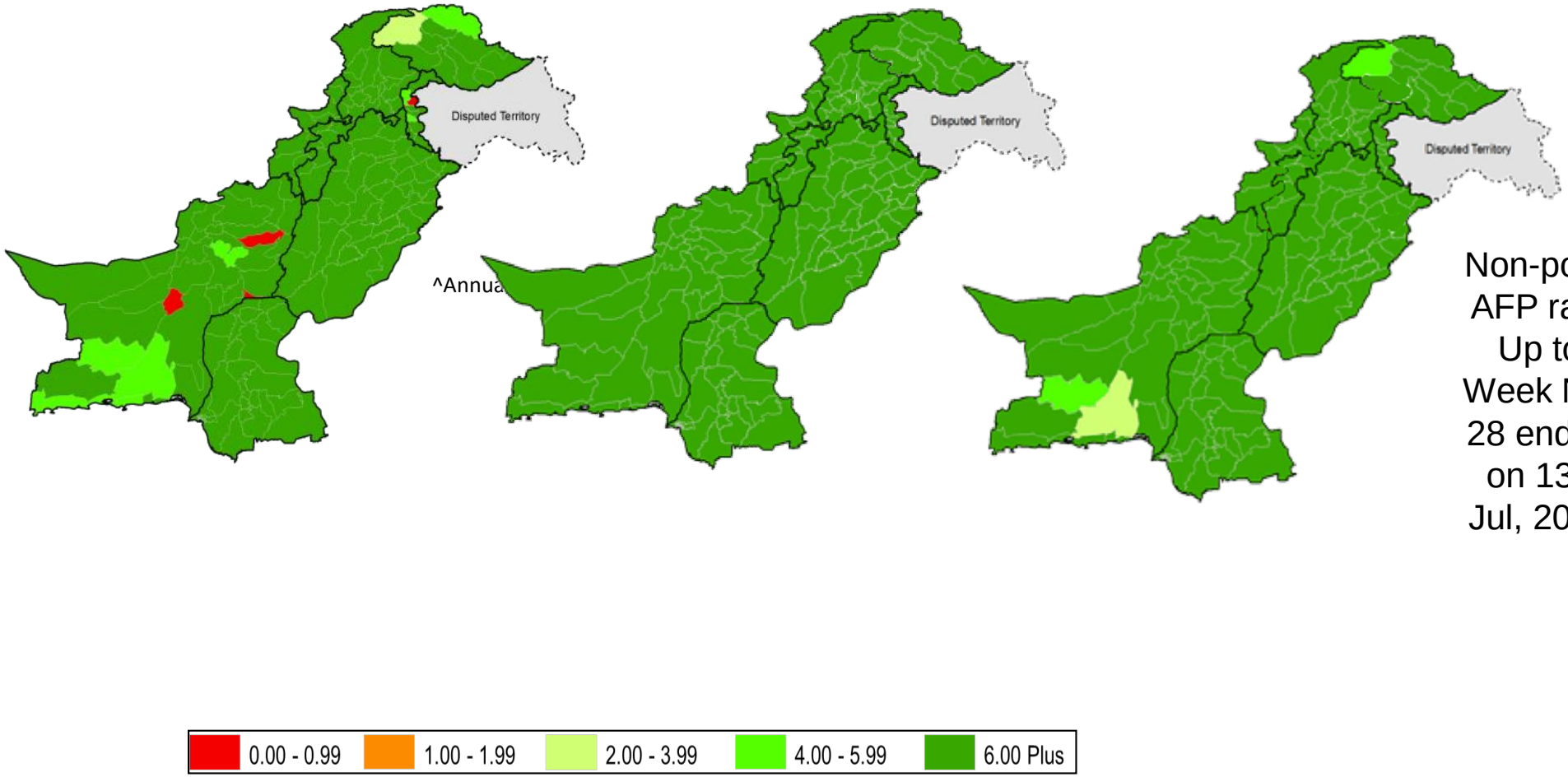
Polio Cases 2019*



* Afp.rec Data as of 15-07-2019

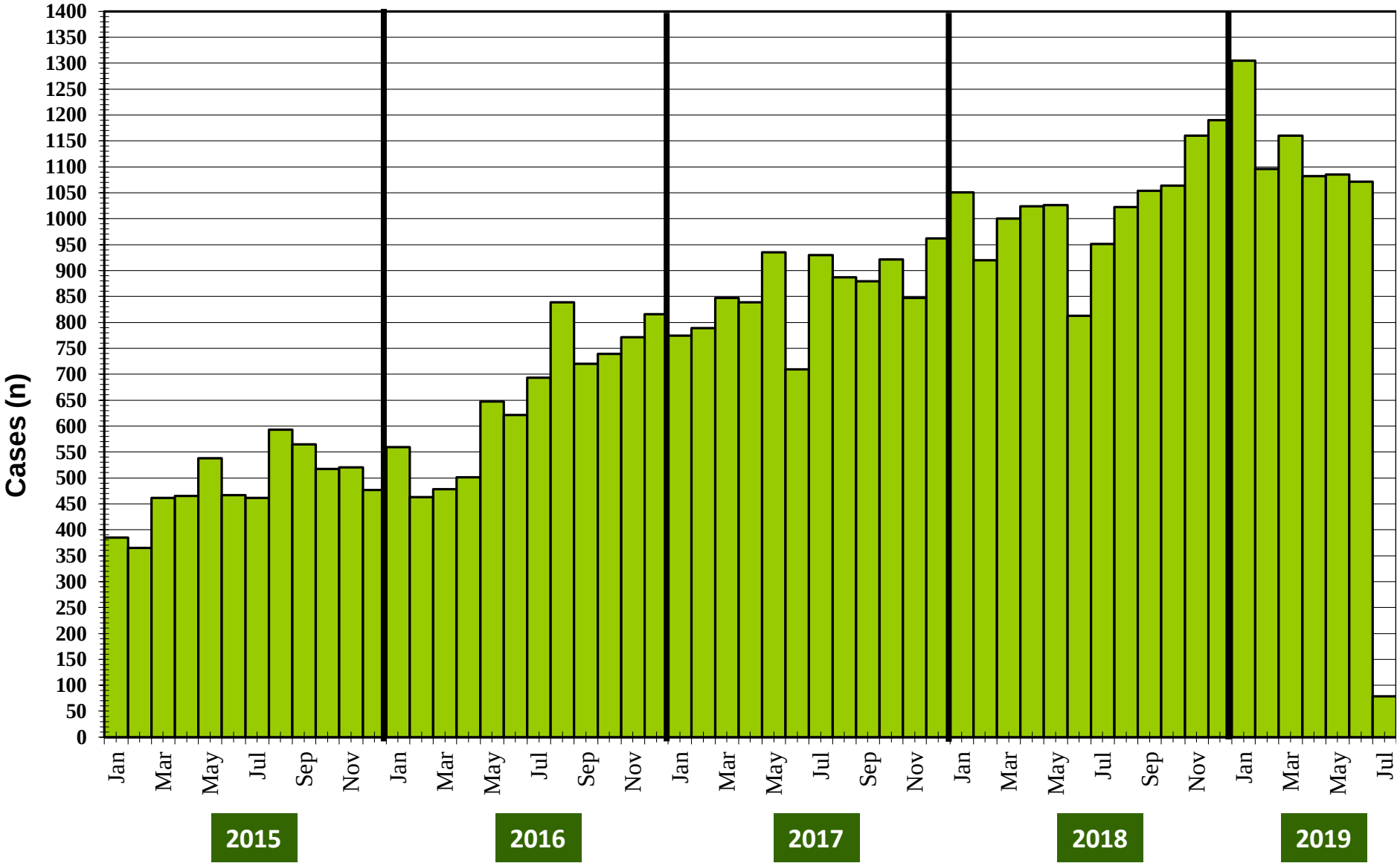
Non-Polio ^ AFP Rate by District

2017 **2018** **2019**



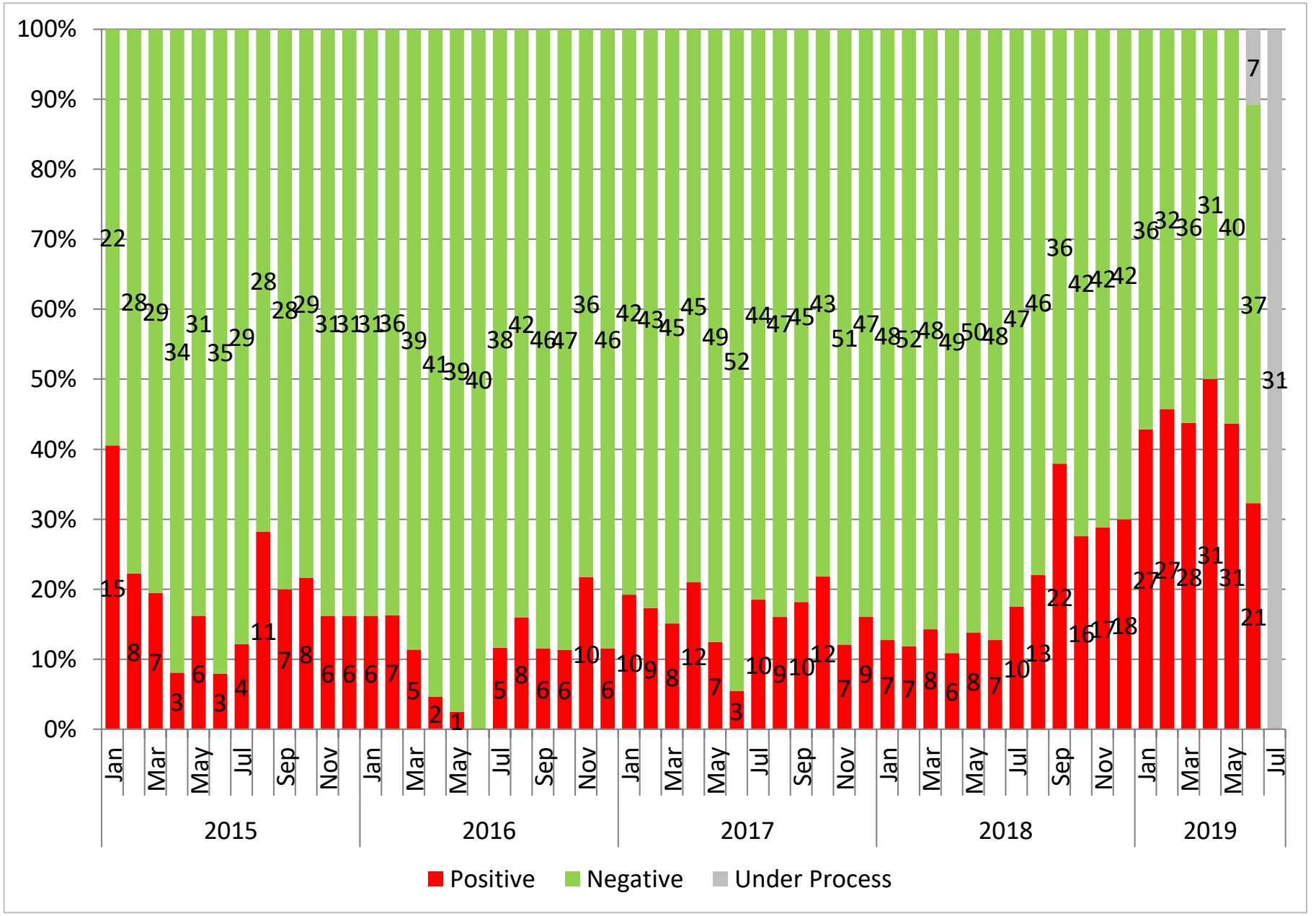
Non-polio AFP rate Up to Week No. 28 ended on 13th Jul, 2019

Graph 1:Distribution of AFP Cases by Month, Pakistan 2015-2019*



* Afp.rec Data as of 15-07-2019

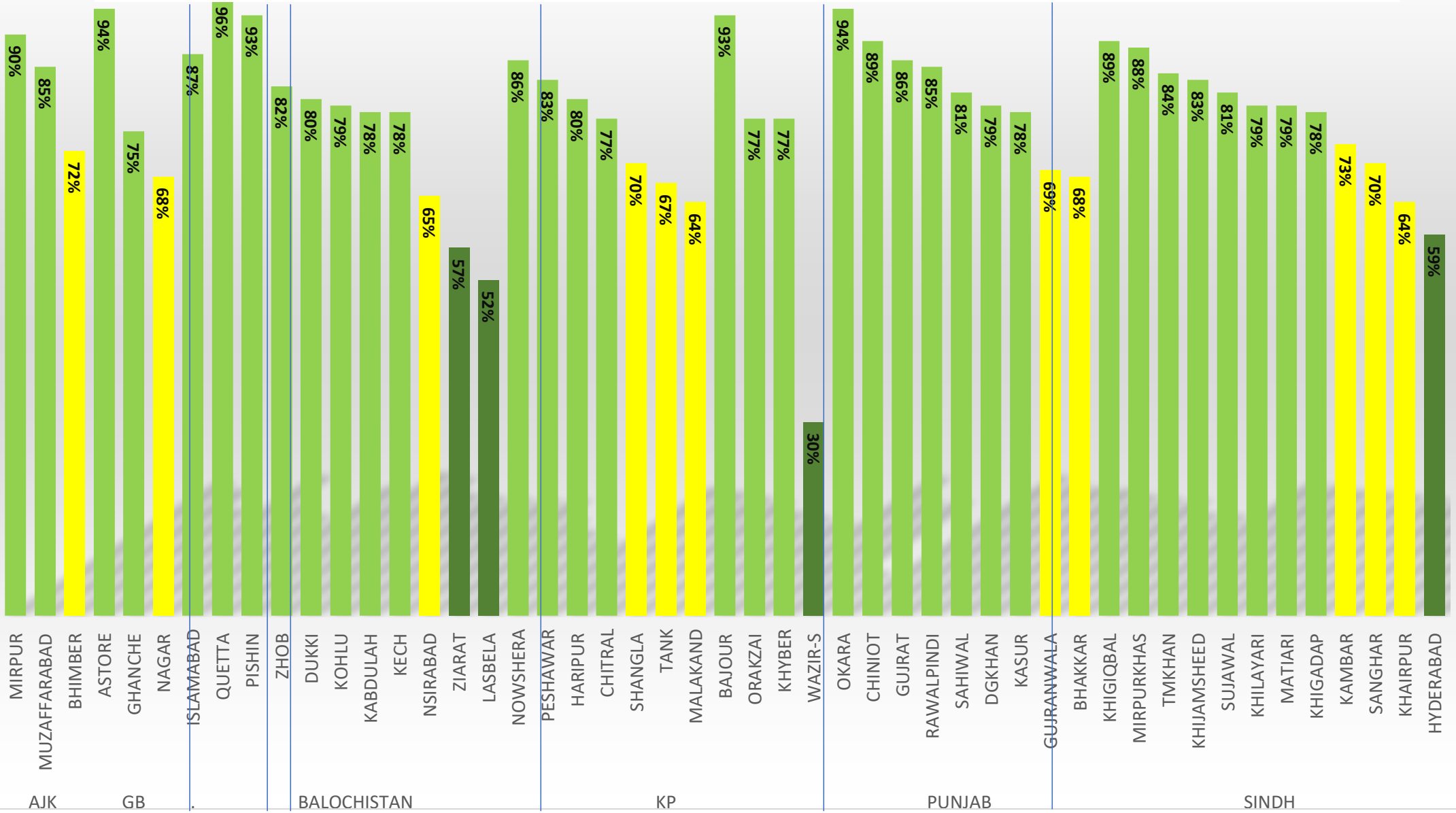
ENVIRONMENTAL SAMPLING RESULTS 2015-19*
PAKISTAN



Micro Plan Quality Assessment

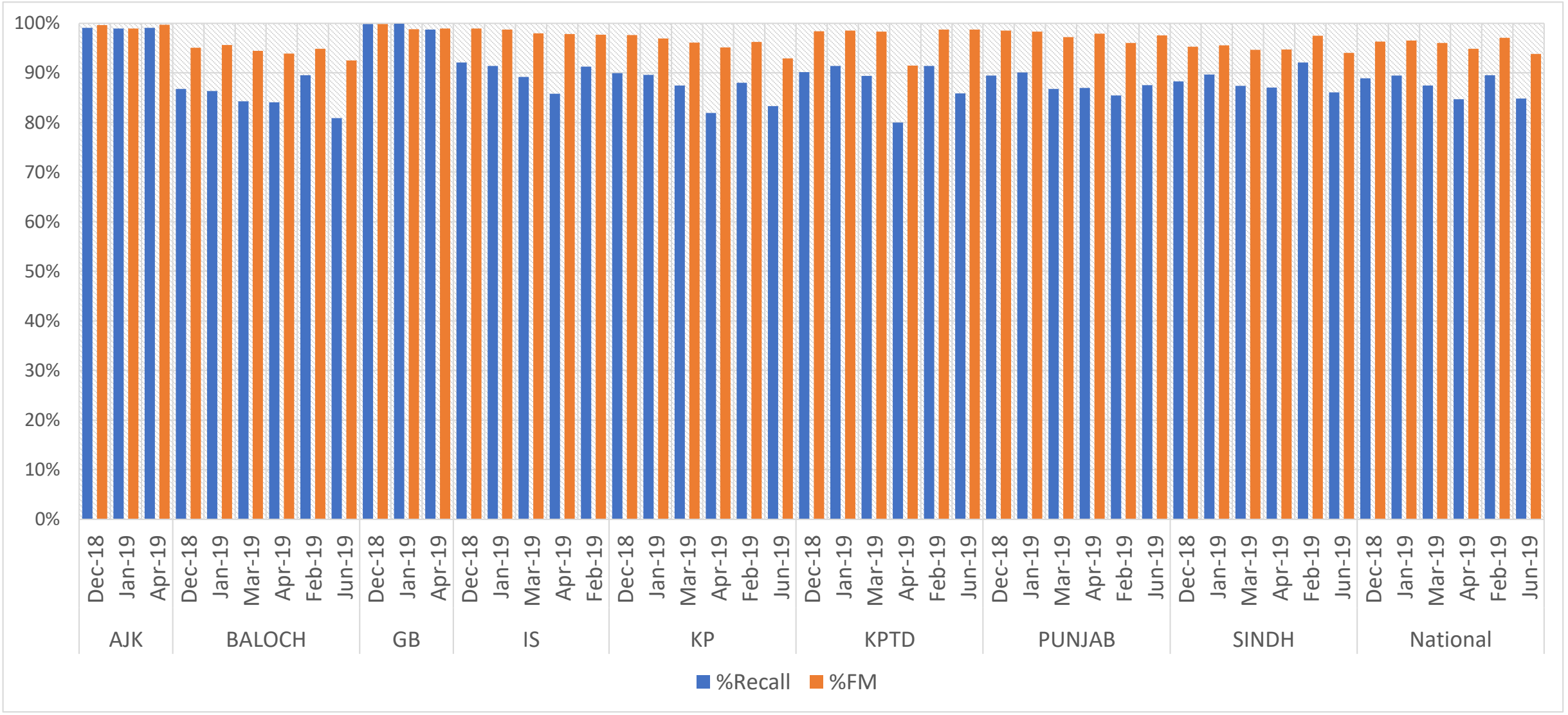
SUMMARY OF MPQA RESULTS

- Micro Plans of Tier-1 Districts from Karachi town, Quetta Block and Peshawar/Khyber are passed
- Substantial gaps identified primarily in Balochistan and pockets of KP and Sindh



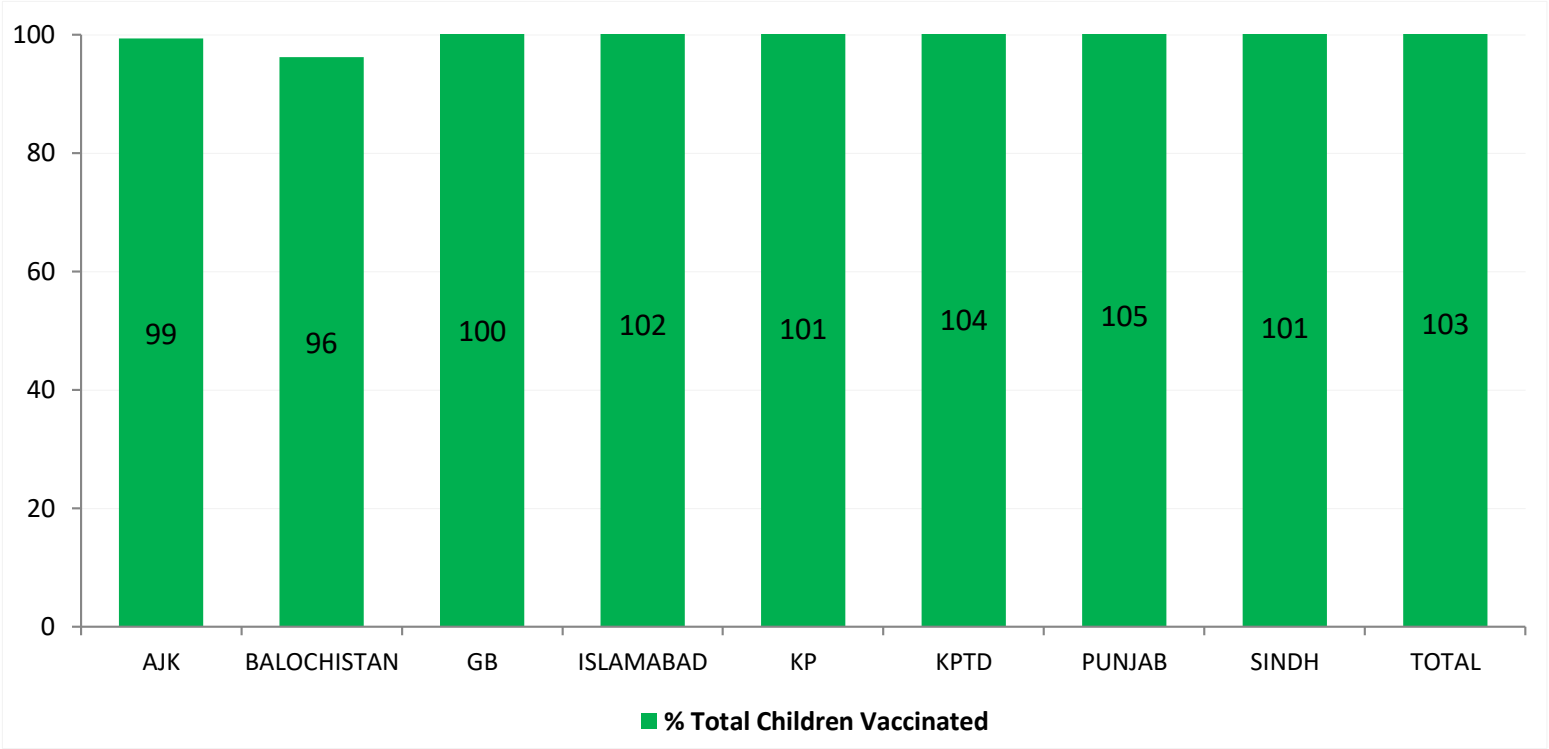
Intra campaign data

Household cluster, Dec18-Jun19



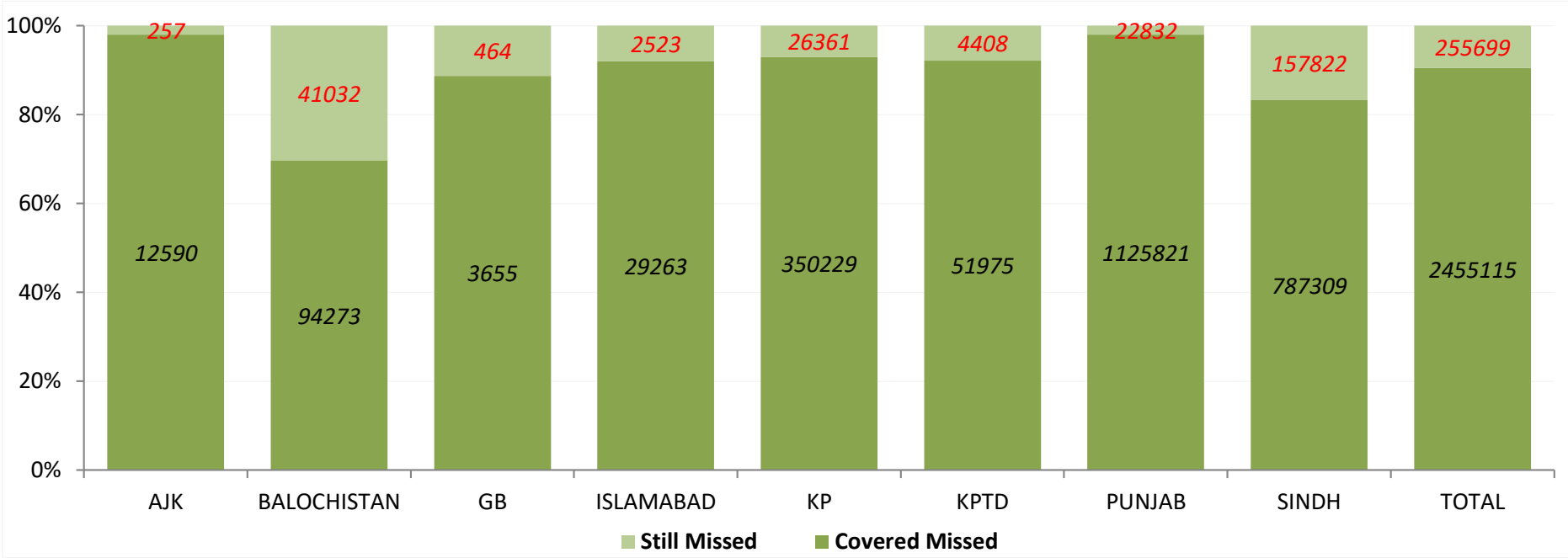
Admin coverage

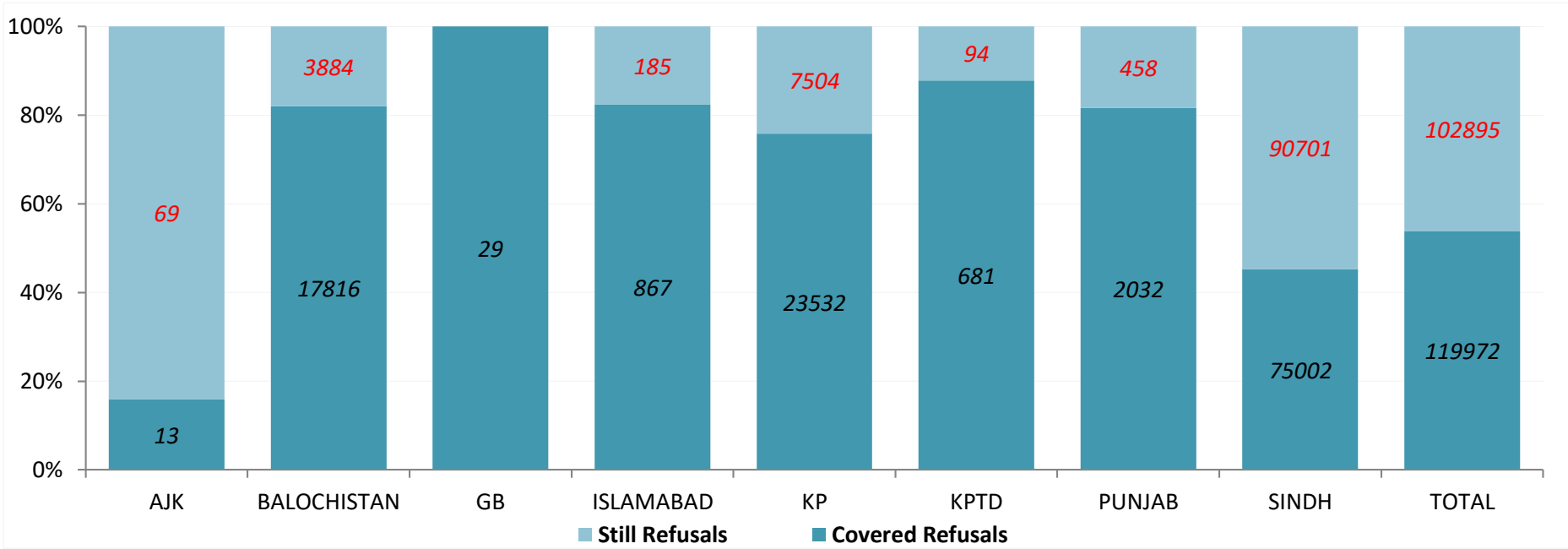
Province	Target Children	% Coverage HH	% Total Children Vaccinated
AJK	703428	86	99
BALOCHISTAN	2541690	86	96
GB	237416	88	100
ISLAMABAD	347932	79	102
KP	5885380	89	101
KPTD	895589	99	104
PUNJAB	19224769	84	105
SINDH	8889020	90	101
TOTAL	**38725224**	**87**	**103**



Province	Recorded Missed Children covered after revisits (same day)				Still Missed Children				
	Covered Refusals	% Among Recorded	Covered Missed	% Among Recorded	Still Refusals	% Among Targeted	Still Missed	% Among Targeted	% Among Recorded
AJK	13	12.26	12590	43.44	69	0.01	257	0	0.9
BALOCHISTAN	17816	51.67	94273	39.02	3884	0.15	41032	1.6	17
GB	29	100	3655	38.57	0	0	464	0.2	4.9
ISLAMABAD	867	57.72	29263	61.07	185	0.05	2523	0.7	5.3
KP	23532	42.19	350229	48.35	7504	0.13	26361	0.4	3.6
KPTD	681	34.85	51975	45.07	94	0.01	4408	0.5	3.8
PUNJAB	2032	46.7	1125821	48.23	458	0	22832	0.1	1
SINDH	75002	29.47	787309	50.64	90701	1.02	157822	1.8	10.2
TOTAL	**119972**	**34.02**	**2455115**	**48.55**	**102895**	**0.27**	**255699**	**0.66**	**5.06**

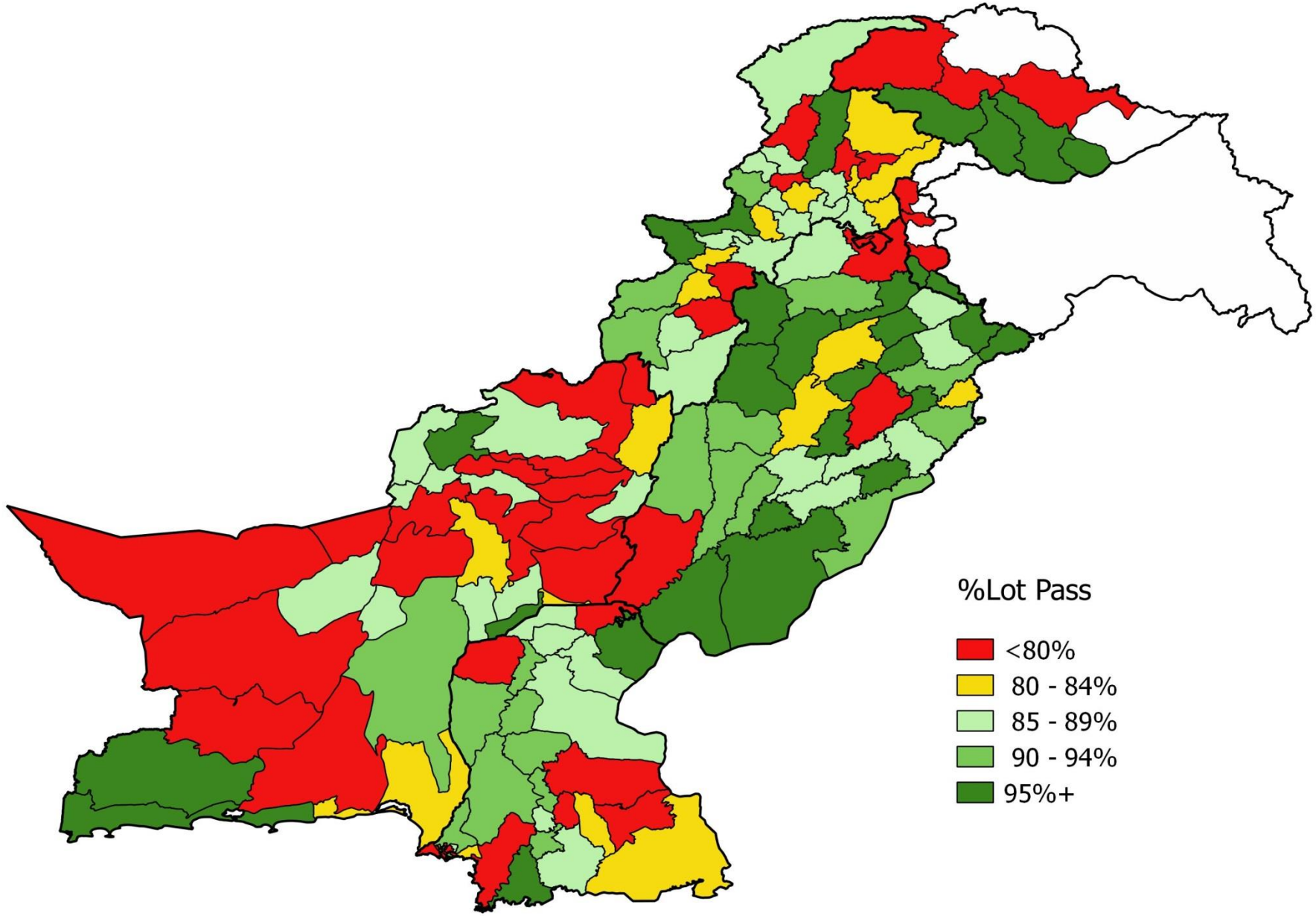
Missed children analysis (Covered and Still Missed Children; Covered and Still Refusals)





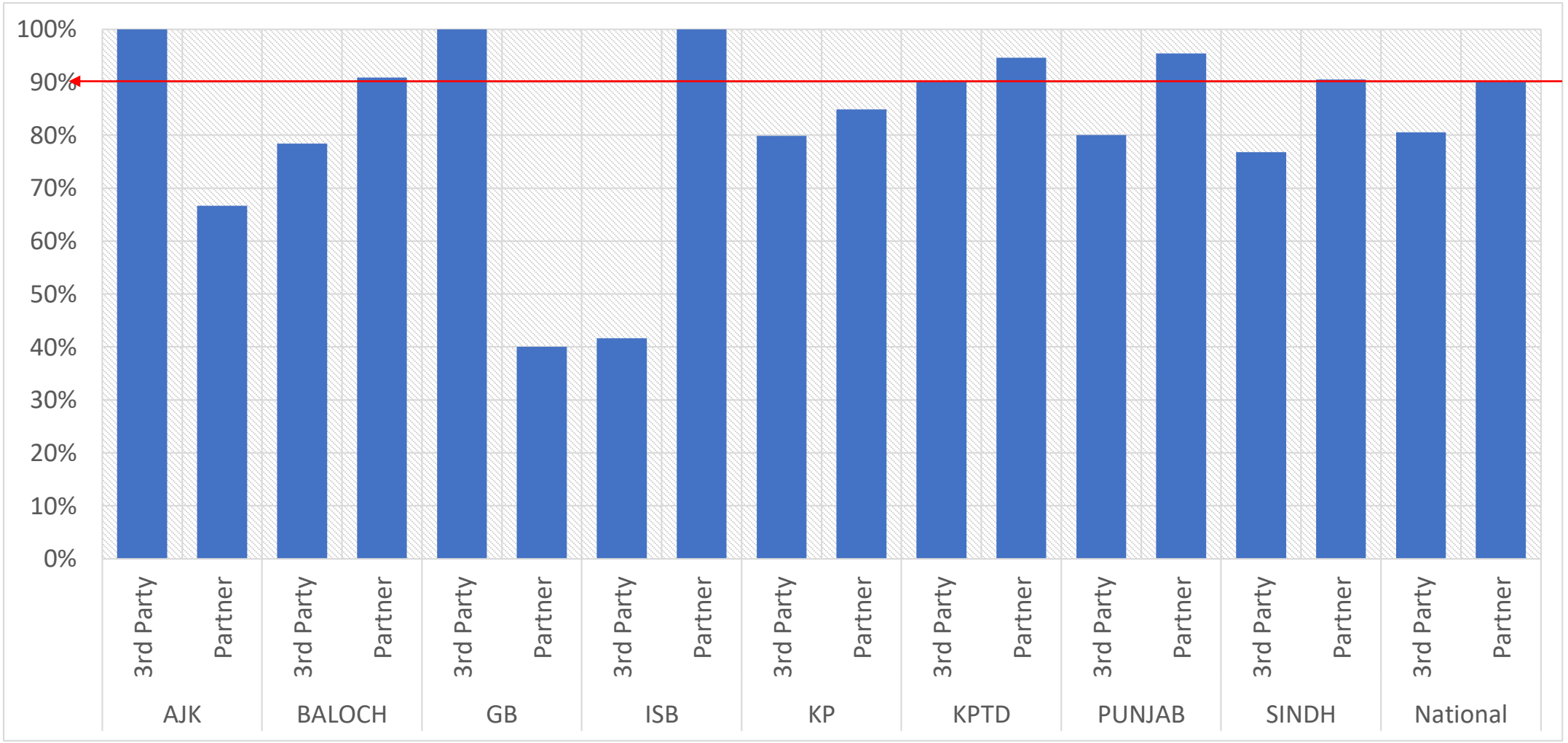
LQAS

%UC s passing LQAS, Sep18-Mar19



LQAS pass rate – Partner vs Third party

- Overall, LQAS pass rate was higher by partner staff compared with third party
- LQAS pass rate higher by partner staff in Balochistan, ISB, KP, KPTD, Punjab and Sindh



Reasons for missed children- Partner vs Third party