




# BMJ Open Investigation of an extensive outbreak of HIV infection among children in Sindh, Pakistan: protocol for a matched case-control study

Amna R Siddiqui,<sup>1</sup> Apsara Ali Nathwani ,<sup>2</sup> Syed H Abidi,<sup>3</sup> Syed Faisal Mahmood,<sup>4</sup> Iqbal Azam,<sup>1</sup> Sobiya Sawani,<sup>1</sup> Abdul M Kazi,<sup>2</sup> Aneeta Hotwani,<sup>5</sup> Sikander Ali Memon,<sup>6</sup> Jamila Soomro,<sup>7</sup> Saqib Ali Shaikh,<sup>6</sup> Baseer Achakzai,<sup>8</sup> Quaid Saeed ,<sup>8</sup> Victoria Simms,<sup>9</sup> Palwasha Khan,<sup>10</sup> Rashida Ferrand,<sup>2,10</sup> Fatima Mir 

**To cite:** Siddiqui AR, Ali Nathwani A, Abidi SH, *et al.* Investigation of an extensive outbreak of HIV infection among children in Sindh, Pakistan: protocol for a matched case-control study. *BMJ Open* 2020;**10**:e036723. doi:10.1136/bmjopen-2019-036723

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-036723>).

RF and FM are joint senior authors.

Received 30 December 2019  
Revised 21 January 2020  
Accepted 26 February 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

**Correspondence to**  
Dr Fatima Mir;  
[fatima.mir@aku.edu](mailto:fatima.mir@aku.edu)

## ABSTRACT

**Introduction** In April 2019, 14 children were diagnosed with HIV infection by a private healthcare provider in Larkana district, Sindh province, Pakistan. Over the next 3 months, 930 individuals were diagnosed with HIV, >80% below 16 years, the largest ever outbreak of HIV in children in Pakistan. In this protocol paper, we describe research methods for assessing likely modes of HIV transmission in this outbreak and investigate spatial and molecular epidemiology.

**Methods and analysis** A matched case-control study will be conducted with 406 cases recruited. Cases will be children aged below 16 years registered for care at the HIV treatment centre at Shaikh Zayed Children Hospital in Larkana City. Controls will be children who are HIV-uninfected (confirmed by a rapid HIV test) matched 1:1 by age (within 1 year), sex and neighbourhood. Following written informed consent from the guardian, a structured questionnaire will be administered to collect data on sociodemographic indices and exposure to risk factors for parenteral, vertical and sexual (only among those aged above 10 years) HIV transmission. A blood sample will be collected for hepatitis B and C serology (cases and controls) and HIV lineage studies (cases only). Mothers of participants will be tested for HIV to investigate the possibility of mother-to-child transmission. Conditional logistic regression will be used to investigate the association of a priori defined risk factors with HIV infection. Phylogenetic analyses will be conducted. Global positioning system coordinates of participants' addresses will be collected to investigate concordance between the genetic and spatial epidemiology.

**Ethics and dissemination** Ethical approval was granted by the Ethics Review Committee of the Aga Khan University, Karachi. Study results will be shared with Sindh and National AIDS Control Programs, relevant governmental and non-governmental organisations, presented at national and international research conferences and published in international peer-reviewed scientific journals.

## INTRODUCTION

The HIV epidemic in Pakistan has been predominantly concentrated in adults

## Strengths and limitations of this study

- Our study will provide information on the likely routes of infection and drivers of the HIV outbreak among children in Pakistan, which is critical to inform appropriate interventions and targeted strategies for HIV prevention.
- The study incorporates molecular epidemiology including the prevalence of different HIV subtypes and circulating recombinant forms, phylogenetic clustering, genetic recency and drug resistance, which will provide an in-depth understanding of the dynamics of the HIV epidemic in the country.
- This study also includes spatial data which will allow us to explore concordance between spatial and genetic clustering observed in this HIV outbreak. Additional spatial data on all local government and private sector health facilities including laboratories, blood banks, formal and informal healthcare providers will be collected which will allow investigation of potential relationships between case location and the acquisition of infection.
- The study is prone to selection bias and the controls may not be representative of the population and exposures are ascertained by self-report and therefore subject to recall bias.
- Widespread reporting in the media about the likelihood of the outbreak having occurred as a result of contaminated needles may result in over-reporting of receipt of injections.

among groups with high-risk behaviour such as people who inject drugs (PWID), men who have sex with men, transgenders and sex workers.<sup>1</sup> In 2018, an estimated 160 000 people were living with HIV in Pakistan, with 2.2% being children aged below 15 years.<sup>2</sup>

In April 2019, a private practitioner in taluka (division) Ratodero, District Larkana in the Sindh province of Pakistan diagnosed



14 children with HIV, based on clinical suspicion. As part of the response, the federal and provincial established an HIV screening camp at Tehsil Hospital Ratodero, a public-sector hospital, with voluntary counselling and testing offered to any individual attending the camp regardless of age. Between April and July 2019, 930 individuals tested HIV positive of whom 763 were children aged below 16 years.<sup>3</sup> This constituted a 54% increase in the cumulative number of paediatric HIV diagnoses over the past 13 years. By 4 December 2019, 952 children with HIV had been identified (personal communication: Sindh AIDS Control Program (SACP)). Although there have been several HIV outbreaks over the past two decades, none has ever been so large nor significantly affected children.<sup>4-8</sup>

The scale of the outbreak and the age of the individuals affected by the outbreak raise important questions about the epidemiology of HIV infection and the mode of transmission of HIV infection. The most widely accepted cause of this outbreak is thought to be the reuse of needles by health practitioners and transfusion of contaminated blood products.<sup>9-10</sup> Nosocomial outbreaks of HIV infection have previously been reported in Pakistan and other countries, but have mostly been related to a specific health facility and have not been at this scale.<sup>5-11-12</sup> In a report of the outbreak, we reported that the proportion of infants with an HIV-positive mother was significantly higher than that among older children, implying that vertical transmission may have contributed to transmission.<sup>3</sup> Although there has been an unexpectedly high number of paediatric HIV diagnoses, it is by no means

certain that the acquisition of infection by children was recent or from one source. It is possible that transmission of HIV may have been ongoing in the community for some time and the increased detection may be a consequence of the vigorous screening efforts implemented by the provincial AIDS Control Program in response to the initial diagnoses. Thus, this 'outbreak' may be an indicator of the bridging of the concentrated HIV epidemic in key populations into the general population.<sup>13</sup>

In successive rounds of HIV surveillance surveys among key populations in Pakistan, the HIV prevalence among PWID has remained persistently extremely high in Larkana, and there has been an upward trend in HIV prevalence among both male and female sex workers (table 1).<sup>1-14-17</sup> Importantly, there is evidence for interactions between different groups that could lead to HIV epidemic bridging (table 1).<sup>18</sup> In addition, low condom usage and large population densities of key population groups in Larkana provide considerable potential for the expansion of the HIV epidemic into the general population.<sup>4</sup>

This protocol paper describes a study to investigate the epidemiology, including the spatial and molecular epidemiology of the outbreak in Sindh.

## METHODS

### Study aims and design

The overall aim of the study is to investigate the outbreak of HIV infection among children in Sindh province,

**Table 1** HIV prevalence and interactions between key population groups in Larkana district, Pakistan in surveillance survey rounds

	Round II <sup>14</sup> (2006)	Round III <sup>15</sup> (2008)	UNGASS Report <sup>17</sup> (2010)	Round IV <sup>16</sup> (2013)	Round V <sup>1</sup> (2016–17)
HIV prevalence in PWID	14%	28.5%	28%	18.6%	16.2%
HIV prevalence in TG-SW	–	27.6%	27%	14.9%	18.2%
Clients per month, n		119			106
% Consistent condom use		37.2%			7.9%
HIV prevalence in FSW	0%	–	0.61%	1.9%	4.1%
% Unmarried					11.3%
% Married children					82.9%
HIV prevalence in MSM	–	–	–	–	4.9%
HIV prevalence in MSW	2.5%	0.5%	0.5%	3.1%	5%
% MSM injecting drugs in past 12 months	–	18%	–	–	10.6%
% TG injecting drugs in past 12 months	–	23%	–	–	0.7%
% MSM having sex with PWID in past 12 months	–	6%	–	–	11.2%
% TG-SW having sex with PWID in past 12 months	1.8%	14.1%	–	–	1.1%
% FSW having sex with PWID in past 12 months	0.3%	–	–	–	0.5%
% PWID paid FSW in past 6 months	3.3%	9%	–	–	52.5%
% PWID paid MSW/TG in past 6 months	9%	13%	–	–	68.8%

FSW, female sex worker; IBBS, Integrated Biological & Behavioral Surveillance; MSM, men who have sex with men; MSW, male sex worker; PWID, people who inject drugs; SW, sex worker; TG, transgender; UNGASS, United Nations General Assembly Special Session on HIV/AIDS.

Pakistan, to inform HIV control strategies. Specific objectives are to:

1. Investigate the likely risk factors for and routes of HIV infection in children aged below 16 years.
2. Explore the molecular epidemiology, specifically the range of HIV subtypes recombinant forms, phylogenetic clustering, genetic recency of infections and prevalence of drug resistance mutations in children with HIV identified in the outbreak.
3. Examine spatial clustering of cases using several techniques, including geographic profiling, and assess the concordance between spatial and genetic clusters.

A matched case–control study will be conducted among children aged below 16 years. Cases will be children with HIV and controls will be HIV-uninfected children matched 1:1 to cases by age, sex and neighbourhood. Data will be collected on likely exposures using a structured questionnaire and global positioning system (GPS) coordinates of the address of cases and controls will be recorded. A blood sample will be collected for hepatitis B and C serology and HIV viral lineage studies (cases only) and mothers of all participants will be tested for HIV.

### Study setting

There has been an adult HIV treatment centre at the Civil Hospital in Larkana city since 2007. Given the large numbers of paediatric HIV diagnoses, the SACP established a paediatric HIV treatment clinic at the neighbouring Shaikh Zayed Children's Hospital in April 2019, where >90% of children diagnosed with HIV (potential cases) through the screening camps have been registered. Following a reactive HIV test (Alere HIV Combo) at the screening camp, confirmatory testing is carried out using two different rapid tests (SD Bioline and Unigold).

The majority of children with HIV identified in the outbreak are from Larkana district but any children living in the adjoining districts, Shikarpur and Qambar Shahdadkot, will also be eligible to participate in the study (figure 1). These three districts are situated in Sindh, the second-most populous province of Pakistan. Larkana is divided into four talukas (Larkana city, Ratodero, Dokri and Bakrani) and has a population of 1 524 391 (2015), comprising 13.5% of the province's population, with 40% aged under 15 years and 16.4% aged below 5 years.<sup>19</sup> The district has among the worst health indicators in the province: the infant mortality rate is 109 per 1000 live births, the under-5 mortality rate 142 per 1000 live births, a third of births have low birth weight (<2.5 kg) and 20% of under 5 years of age have severe acute malnutrition. A quarter of households are in the poorest wealth quintiles.<sup>20</sup>

### Participant recruitment

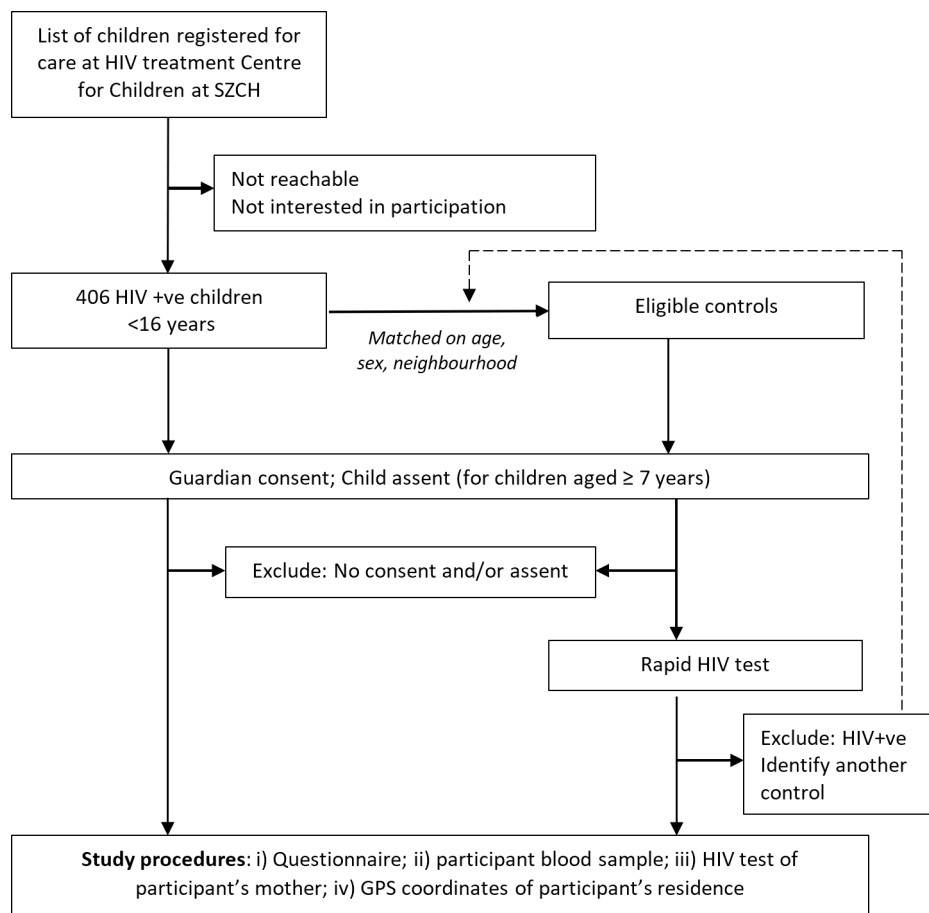
A case will be defined as a child aged below 16 years registered for HIV care at the paediatric HIV treatment centre in Larkana city, regardless of whether she/he is taking antiretroviral therapy (ART). A control matched for sex, age ( $\pm 1$  year) and neighbourhood will be defined as an individual who tests HIV negative. Exclusion criteria will



Figure 1 Map of the study sites.

be the lack of consent from a guardian or assent from the child to participate in the study.

A list of children registered at the Paediatric Treatment Center at Shaikh Zayed Children's Hospital (potential cases) will be obtained after the permission from the SACP. The list will contain the name, age and contact details of children with HIV registered for care. The study team will sign an agreement to keep this list confidential and within office premises. The research team will contact the caregiver of each child by phone from the office and provide information about the study and invite them to participate. If they agree, a home visit will be arranged. If an individual is not contactable, then the team will contact the next individual on the list. The list will be kept under lock and key at the study office. For each case, a control will be enrolled from the same neighbourhood ('muhallah'). Once a case is recruited, the study team will approach the neighbouring dwelling and enquire if there is a child with matching age and sex residing in the dwelling. If there is, information about the study will be provided and the child invited to participate as a control. If there is not, then the study team will approach the next dwelling along on the same street (covering both sides of the street) until an appropriate eligible control



**Figure 2** Study flow. GPS, global positioning system; SZCH, Shaikh Zayed Children's Hospital, Larkana City.

is identified. If there is no control identified in the same street, then the research team will move to the adjacent street. All eligible controls will be screened for HIV prior to recruitment (see the 'Study procedures' section).

### Study procedures

Study staff will be trained on consent procedures, data collection including GPS data, HIV counselling and testing and collection of blood specimens and all study tools will be piloted.

After consent and assent are obtained, eligible controls will undergo rapid HIV testing using a fourth-generation rapid blood-based test (Alere HIV Combo) and recruited into the study if they test HIV negative. For those aged below 18 months, molecular HIV testing will be carried out if the antibody test is positive and the mother's test is HIV positive. In the event that an eligible control's HIV test is reactive, an alternative eligible control will be identified. A structured questionnaire will be administered to the participants' parents/caregivers or the participants themselves if they are aged above 10 years. Sociodemographic data including age, sex, religion, ethnicity, schooling, marital status and household socio-economic status of the participant, and the vital status, HIV status and HIV testing history of the participant's biological parents and natural siblings (defined as those with the same mother) will be collected. For cases only,

information about the HIV testing and treatment history including date of HIV diagnosis and ART initiation (if applicable) will be recorded.

The questionnaire will contain questions about exposures for HIV infection including frequency of contact with different types of healthcare services (private hospitals or clinics, government hospitals, basic health units or rural health centre) and receipt of injections, intravenous fluid ('drips'), surgery and blood transfusions. If participants have received a blood transfusion, then the frequency and reason for transfusion, the location of the transfusion and the source of the blood product (blood bank, donation by a family member or friend) will be recorded. Where possible, responses will be cross-checked against participant-held medical records. Cultural practices such as circumcision (male individuals only), head shaving, tattooing, nose or ear piercing and engaging in 'zanjeer zani' (self-flagellation using knives for mourning the martyrdom of Prophet Muhammad's grandson in the first month of the Islamic calendar) will be recorded. Participants who are aged above 10 years will also be asked about the history of consensual or non-consensual sexual intercourse and intravenous drug use including needle sharing. Questions on the existence of brothels, drug dens and hospital waste dumps in the neighbourhood will be asked. The questionnaires are available on request.

**Table 2** Sample size estimates

Probability of exposure among controls	Correlation of exposure	Power, %	Minimum OR detected
0.1	0.2	80	1.94
0.1	0.2	90	2.12
0.1	0.4	80	2.18
0.1	0.4	90	2.43
0.5	0.2	80	1.56
0.5	0.2	90	1.67
0.5	0.4	80	1.67
0.5	0.4	90	1.81
0.8	0.2	80	1.81
0.8	0.2	90	2.00
0.8	0.4	80	1.98
0.8	0.4	90	2.21
0.95	0.2	80	3.56
0.95	0.2	90	4.45
0.95	0.4	80	4.23
0.95	0.4	90	5.39

A blood sample will be collected from participants for hepatitis B and C serological testing and viral lineage studies (for cases only). Mothers of all participants will undergo HIV testing using the same rapid HIV test used to test eligible controls. Pretest and post-test counselling will be provided before any HIV testing is performed in the study. Results will be given on the same day. Anyone with a reactive HIV test will be referred to the paediatric or adult HIV clinic in Larkana for confirmatory testing and linkage to care. The study flow is summarised in [figure 2](#).

### Specimen collection and transportation

Blood will be collected from participants in EDTA and gel top tubes (total volume 3–5 mL) labelled with the study ID number and packed in a Biohazard Zip-lock Bag. Blood samples will be maintained in an upright position at 2°C–4°C with frozen ice packs and transported within 6–8 hours to the Aga Khan University (AKU) Infectious Disease Research Laboratory (IDRL) in Matiari (an AKU satellite research laboratory in Sindh) where the sample condition (temperature and evidence of leakage) will be checked and the sample details will be confirmed with those on the accompanying laboratory form. Samples will subsequently be dispatched daily to AKU in Karachi. Hepatitis B and C serology and viral lineage studies will be performed at the IDRL and the Department of Basic Biological Sciences at AKU respectively.

### Viral lineage and drug resistance studies

Viral nucleic acid will be extracted from the blood sample of cases using commercially available kits (Qiagen) following the manufacturer's instructions, and the

Gag-Pol region of the HIV genome will be amplified by PCR using previously optimised protocols. The amplicons will be purified and undergo Sanger sequencing (Macrogen, Korea). The sequences will be edited, the HIV subtype determined and the phylogenetic relationship between different HIV subtypes will be analysed.

### Geospatial data

GPS coordinates of the households of participants will be obtained and recorded using smartphones. Location data of all known local government and private health facilities including hospitals, primary healthcare clinics, local laboratories, local blood banks, informal healthcare providers/pharmacists, local barbers and body piercing outlets in the town and surrounding area will be collected

### Data management

Data will be collected on paper forms and each completed data form will be checked by the field supervisor for missing or incorrect data prior to data entry. Data will be double entered into a password-protected database (Visual FoxPro V.9.0). Data cleaning will include range checks and checks for missing values and the final cleaned dataset will be deidentified for analysis. Where possible, the transmission of data through email will be avoided and a secure file transfer server will be used to transmit the data. Data files will be deidentified with unique study identifiers (study IDs) incorporated. All electronic files will be password protected. Data will be backed up on a server (with weekly offsite backup) and paper forms will be stored in locked cabinets for 3 years after study completion. Passwords will be assigned to all machines storing data and only accessible to authorised personnel. GPS data will be collected using mobile phone devices that will be stored securely during the study. All GPS data will be deleted from devices after entry on to data forms.

### Sample size estimates

Assuming a probability of exposure to injections among controls of 0.8 and a correlation coefficient for exposure between matched case and control of 0.4, a sample size of 406 HIV cases and 406 controls matched on age, sex and area of residence will provide at least 80% power to detect an OR of at least 1.98 at a two-sided significance level of 0.05. [Table 2](#) shows that there will be at least 80% to detect an OR of  $\geq 2.0$  for the probability of exposures ranging from 0.1 to 0.85 and a correlation coefficient for exposure between matched HIV case and control subjects of 0.20.

### Data analyses

#### Case-control study

Statistical analysis will be done using STATA (V.14 or later). Continuous variables will be presented as means (SD) if normally distributed or median (IQR) if skewed. Categorical data will be presented as proportions. Our dataset will include cases individually pair-matched to controls of the same age, sex and neighbourhood (usually living on the same street). The bivariate analysis will be



used to determine crude ORs and 95% CIs for the association of HIV with key exposures related to parenteral (eg, receipt of injections, history of injection drug use, history of blood transfusion, history of surgery, hepatitis B or C antibody positivity and so on), vertical (maternal and/or sibling HIV status) or sexual transmission. Multi-variable conditional logistic regression will be conducted to control for confounding and adjusted ORs for exposures in cases compared with controls with 95% CIs will be computed. Conditional logistic regression is required because each case will be individually pair-matched to a control of the same age, sex and neighbourhood. Adjusting for age and gender will still be required as the matching itself can introduce a bias.<sup>21</sup>

### Genomic analyses

Cluster analysis will be performed to identify statistically significant nodes on the phylogenetic tree and the supported clusters. The node age, node substitution and branch length of each sequence will be used to identify features such as the similarity between sequences or clusters, ancestral sequences, monophyletic clusters, founder effect and recombination. A Bayesian framework will be used to determine the effective population size (correlated to the effective number of HIV transmission) and the time of the most recent common ancestor in Larkana taxa in relation to recognised HIV sequences from Pakistan. The Bayesian substitution tree will identify the rate of substitution for each cluster, which will be combined with pair-wise genetic distances to calculate the genetic recency of infection with the assumption that substitutions/genetic distance tends to increase with age or duration of infection. A Bayesian birth–death model will estimate the reproductive number for significant taxa to estimate the rate of the epidemic/outbreak growth.

The amino acid sequences of the HIV Pol gene (which codes for reverse transcriptase, HIV-1 protease and integrase) will be aligned with the reference sequences to identify Pol mutations from both the groups. The association of Pol mutations with resistance to ART drugs will be determined using the Stanford Drug Resistance Database and counter-confirmed using the International AIDS Society report on drug resistance. The proportion of ART-naïve and ART-experienced cases with drug resistance mutations will provide an understanding of the probability of transmitted drug resistance and suboptimal treatment adherence.

### Geospatial analyses

Geographical profiling (GP), is a novel geospatial tool originally developed in criminology to prioritise large lists of suspects in cases of serial crime, using spatial locations of linked crimes to identify areas likely to be associated with the criminal. GP has the potential for use in the field of pathogen transmission research, as its aim is to identify source locations, rather than simply identifying spatial clustering of infection.

GP will be used to identify areas that are potential sources of nosocomial transmission using the location of cases whose mothers are HIV negative. The spatial relationship between locations of local government and private health facilities, local laboratories, local blood banks, informal healthcare providers, local barbers and body piercing outlets and so on and any GP-identified hotspots will be examined. GP will be carried out using *Rgeoprofile* (<https://github.com/bobverity/Rgeoprofile>).

Traditional geospatial methods including non-parametric distance-based mapping approach,<sup>22</sup> kernel-density estimation<sup>23</sup> and SaTScan<sup>24</sup> will also be implemented to identify spatial clusters. The concordance between genetic and spatial clusters will be measured using permutation testing to address the hypothesis that there will be a statistically significant association between spatial groups identified by GP and genetic groups identified in the genomic cluster analyses detailed above.

### Patient and public involvement

Patients or the public will not be involved in the design or conduct of the study. Results will be disseminated to the community.

### Study status

Ethical approval was obtained in July 2019. Recruitment began in July 2019 and epidemiological, phylogenetic and geospatial analysis protocols are being finalised.

## DISCUSSION

Pakistan is one of only four countries in Asia where the incidence of HIV has continued to increase in the past decade. The number of new HIV infections increased from 14000 in 2010 to 22000 in 2018, with a 369% rise in HIV-related deaths over the same period.<sup>2</sup> Although the HIV epidemic is mostly limited to key populations, outbreaks in the general population have occurred in previous years.<sup>4,8</sup> The outbreak in Larkana, Sindh in 2019 is the largest reported to date and the first time that children have been predominantly affected.<sup>3</sup> Although there has undoubtedly been a vigorous effort by the provincial and federal government and other stakeholders to implement an appropriate response to the outbreak, understanding the drivers and the evolution of the HIV outbreak is critical to inform coherent middle-term and long-term strategies to control the HIV epidemic in Pakistan.

It has been widely assumed that the HIV outbreak is due to parenteral transmission through the reuse of needles in healthcare settings, although vertical transmission may also play a role.<sup>3,9,10</sup> This study is important to identify the primary mode of HIV transmission in this outbreak to tailor context-specific interventions. There is also a need to establish whether the HIV outbreak has occurred as a result of a bridging of the HIV epidemic from high-risk groups into the general population (primarily children) through the nosocomial transmission.<sup>18</sup>

Potential challenges include difficulties in tracing potential cases and in identifying eligible controls and low participation rates. A limitation of the study is that exposures to possible risks for HIV infection will be ascertained by self-report and therefore there is a risk of recall bias. Another limitation is that we have restricted length of recall of exposures to a period of 6 months for common exposures (use of injections) and 12 months for less common exposures (such as surgery) despite the possibility that children may have been infected earlier. Participants may not accurately recall exposures beyond a certain length of time and increasing length of recall of exposures can increase the risk of recall bias. Where possible, responses will be cross-checked against participant-held records. It is also possible that in children, particularly older children, a particular exposure may have occurred after the acquisition of HIV infection. There has been widespread reporting of the likelihood of the outbreak occurring as a result of the widespread use of contaminated needles by healthcare providers. In particular, a doctor practising in Ratodero who treated many of the children who were subsequently diagnosed with HIV infection was implicated in spreading the infection, a story that received much media coverage in Pakistan.<sup>25</sup> This may affect the accuracy of reporting of contact with healthcare providers by participants. Another limitation is that the study is prone to selection bias. The study design has been chosen to improve efficiency and to enable control in the analysis of unquantifiable factors such as neighbourhood characteristics that might influence the risk of HIV infection (eg, access to healthcare providers).<sup>21 26</sup> However, the controls may not be representative of the level of exposure in the population.

The study will include geospatial and genomic analyses and will investigate the concordance between the genetic and spatial epidemiology to provide an in-depth understanding of the outbreak.

### Ethics and dissemination

Ethical approval was sought and granted by Ethical Review Committee of the AKU, Karachi. Written informed consent in the local language will be obtained from parents or guardians of all participants and verbal assent from children aged 7 to 15 years. Data forms will contain no identifying information other than age, sex and Geographic Information System (GIS) coordinates. Laboratory forms will contain no identifying information, specimens will be identified by a study number only and test results will not be linked to any individual by name. All survey staff will sign a confidentiality agreement to ensure that they do not release participant identities and test or study results to individuals who are not part of the study team.

Study progress and findings will first be reported to the SACP and the National AIDS Control Program (NACP). Subsequently, results will be disseminated to health institutions and other relevant governmental and non-governmental organisations (NGOS) and community

groups, for example, People's Primary Healthcare Initiative (PPHI), Integrated Health Services (IHS), Sindh Healthcare Commission (SHCC), Sindh Blood Transfusion Authority (SBTA), Nai Zindagi Trust (an NGO working with PWIDs) and PRIH (an NGO working with sex workers). The study team is already engaged with relevant stakeholders addressing this outbreak (SACP, NACP, WHO, UNICEF) and will participate in policy and advocacy meetings to facilitate the development and implementation of strategies to strengthen HIV prevention and care in the country. Results will be presented at national and international research meetings and conferences and also prepared for publication in international peer-reviewed scientific journals. Study findings will be disseminated to the study communities.

### Author affiliations

<sup>1</sup>Department of Community Health Sciences, Aga Khan University, Karachi, Sindh, Pakistan

<sup>2</sup>Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan

<sup>3</sup>Department of Biological and Biomedical Sciences, Aga Khan University, Karachi, Sindh, Pakistan

<sup>4</sup>Section of Infectious Disease, Department of Internal Medicine, the Aga Khan University, Karachi, Sindh, Pakistan

<sup>5</sup>Infectious Disease Research Laboratory, Department of Pediatrics and Child Health, the Aga Khan University, Karachi, Sindh, Pakistan

<sup>6</sup>Sindh AIDS Control Program, Ministry of Health, Karachi, Sindh, Pakistan

<sup>7</sup>Public Health Wing, Ministry of Health, Karachi, Sindh, Pakistan

<sup>8</sup>National AIDS Control Program, Islamabad, Pakistan

<sup>9</sup>Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>10</sup>Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK

**Twitter** Fatima Mir @FatimaMirPedsID

**Contributors** FM and RF: conceptualised project. ARS, AAN, SFM, SS, FM and RF: developed study design and questionnaires. AAN, SS and FM: oversight of data collection. SAM, JS, SAS, BA and QS: support of study logistics and access to patient lists. IA and VS: statistical analyses. SHA: phylogenetic analyses. AH: hepatitis serology and sample QC. AMK and PK: geospatial analyses. All authors have read and approved the final manuscript.

**Funding** This study is funded by the Department of Paediatrics and Child Health, The Aga Khan University, Karachi, Pakistan. RAF is funded by the Wellcome Trust through a Senior Fellowship in Clinical Science (206316/Z/17/Z)

**Map disclaimer** The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iDs

Apsara Ali Nathwani <http://orcid.org/0000-0002-2899-1577>

Quaid Saeed <http://orcid.org/0000-0001-6736-4253>

Fatima Mir <http://orcid.org/0000-0001-8602-5353>



## REFERENCES

- 1 National AIDS Control Program Pakistan. Integrated biological and behavioral surveillance in Pakistan 2016-2017; 2018.
- 2 UNAIDS. *Country Factsheets*. Pakistan, Geneva, Switzerland: UNAIDS, 2018.
- 3 Mir F, Mahmood F, Siddiqui AR, *et al*. Hiv infection predominantly affecting children in Sindh, Pakistan, 2019: a cross-sectional study of an outbreak. *Lancet Infect Dis* 2020;20:362–70.
- 4 Wahid B. An update on the severe outbreak of HIV in Kot Imrana, Pakistan. *Lancet Infect Dis* 2019;19:241.
- 5 Altaf A, Pasha S, Vermund SH, *et al*. A second major HIV outbreak in Larkana, Pakistan. *J Pak Med Assoc* 2016;66:1510–1.
- 6 Shah SA, Altaf A, Mujeeb SA, *et al*. An outbreak of HIV infection among injection drug users in a small town in Pakistan: potential for national implications. *Int J STD AIDS* 2004;15:209.
- 7 Ansari JA, Salman M, Safdar RM, *et al*. Hiv/Aids outbreak investigation in Jalalpur Jattan (JPJ), Gujrat, Pakistan. *J Epidemiol Glob Health* 2013;3:261–8.
- 8 Zaid M, Afzal MS. HIV outbreak in Pakistan. *Lancet Infect Dis* 2018;18:601.
- 9 Arif F. HIV crisis in Sindh, Pakistan: the tip of the iceberg. *Lancet Infect Dis* 2019;19:695–6.
- 10 Altaf A, Iqbal S, Shah SA. A third major human immunodeficiency viruses (HIV) outbreak in Larkana, Pakistan: caused by unsafe injection practices. *J Pak Med Assoc* 2019;69:1068–9.
- 11 Yerly S, Quadri R, Negro F, *et al*. Nosocomial outbreak of multiple bloodborne viral infections. *J Infect Dis* 2001;184:369–72.
- 12 Pan X, Jiang J, Ma Q, *et al*. Outbreak of HIV infection linked to nosocomial transmission, China, 2016–2017. *Emerg Infect Dis* 2018;24:2141–9.
- 13 Niccolai LM, Shcherbakova IS, Toussova OV, *et al*. The potential for bridging of HIV transmission in the Russian Federation: sex risk behaviors and HIV prevalence among drug users (DUs) and their non-DU sex partners. *J Urban Health* 2009;86 Suppl 1:131–43.
- 14 National AIDS Control Program. HIV second generation surveillance in Pakistan: national report round II 2006–2007. Islamabad, Pakistan; 2007.
- 15 National AIDS Control Program. HIV second generation surveillance in Pakistan: national report round III, 2008. Islamabad, Pakistan; 2008.
- 16 National AIDS Control Program. HIV second generation surveillance in Pakistan: national report round IV, 2013. Islamabad, Pakistan; 2013.
- 17 National AIDS Control Program. Progress report on the Declaration of commitment on HIV/AIDS for the United nations General assembly special session on HIV/AIDS. Islamabad, Pakistan; 2010.
- 18 Khanani MR, Somani M, Rehmani SS, *et al*. The spread of HIV in Pakistan: bridging of the epidemic between populations. *PLoS One* 2011;6:e22449.
- 19 Sindh Bureau of Statistics and UNICEF. Sindh multiple indicator cluster survey; 2015.
- 20 National Institute of Population Studies (NIPS) Pakistan and ICF. Pakistan demographic and health survey 2017–18. Islamabad, Pakistan and Rockville, Maryland, USA; 2018.
- 21 Pearce N. Analysis of matched case-control studies. *BMJ* 2016;352:i969.
- 22 Zelner JL, Murray MB, Becerra MC, *et al*. Identifying hotspots of multidrug-resistant tuberculosis transmission using spatial and molecular genetic data. *J Infect Dis* 2016;213:287–94.
- 23 Yang C, Lu L, Warren JL, *et al*. Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis. *Lancet Infect Dis* 2018;18:788–95.
- 24 SaTScan. *SaTScan v8.0: Software for the spatial and space-time scan statistics [program]*, 2009.
- 25 Kaleem M. What lies behind the spread of HIV in Larkana. *Herald* 2019.
- 26 Rose S, van der Laan MJ. Why match? investigating matched case-control study designs with causal effect estimation. *Int J Biostat* 2009;5:Article 1.