Progress towards obtaining valid vaccination coverage data in South Africa

In March 2019, the South African National Department of Health (SANDoH) launched South Africa’s first national household vaccination coverage survey to be held since 1994. The need for this survey is driven by several imperatives. First, the World Health Organization (WHO) Global Vaccine Action Plan (GVAP) 2011–2020, had set a 2020 global target for countries to reach 90% national coverage of all their primary series vaccines.1 The administrative fully immunised under one year old coverage (FIC) of the Expanded Programme on Immunisation of South Africa (EPI-SA), reported through the District Health Information System (DHIS), shows that EPI-SA is not on track for reaching GVAP’s target.2 EPI-SA’s FIC steadily increased from 83.6% in 2012/2013 to 89.8% in 2014/2015, before dropping to 82.3% in 2016/2017.2

Second, for the past decade, EPI-SA has reported much higher administrative vaccination coverage than the WHO and United Nations Children’s Fund Estimates (UNICEF) of National Immunisation Coverage (WUENIC).3 South Africa is one of many low- and middle-income countries (LMICs) with annual administrative coverage rates that differ substantially from WUENIC.4 WUENIC are based on both administrative coverage rates and population-based surveys, including household surveys using the WHO EPI vaccination coverage cluster survey method5, and Demographic and Health Surveys (DHS).6

Third, the SANDoH’s DHS (SADHS) conducted in 2016 reported a national FIC for 12–23-month-old children of only 52.7%, which was 36.5% lower than 89.2%, the DHIS FIC for 2015/2016.7 While the sample size of 677 12–23-month-olds was very small for a national survey (ranging from 12 children in the Northern Cape to 180 in Gauteng), the FIC result was nevertheless concerning.

Fourth, reports from published South African population-based studies8–11 and sporadic measles outbreaks12–14 provided the SANDoH with evidence that FIC targets are not being reached. Thus the SANDoH recognised that a large-scale national vaccination coverage survey was needed to provide them with a clearer picture of true FIC, and identify pockets of low coverage too small to be noticed within aggregated data reported through the DHIS.15 Planning for this survey began in 2012; however, funding was a major obstacle, as these large surveys are very costly.16

Immediately following the release of the SADHS 2016, the SANDoH convened a workshop for all stakeholders on the ‘SADHS Findings and EPI-SA Turn-around Strategy’. A key issue identified during this workshop was that the planned large national household vaccination coverage survey was long overdue, and funding for the survey was subsequently secured from the vaccine industry. The Wits Health Consortium is conducting this survey, with technical support from WHO and UNICEF. The survey was launched on 8 March 2019 by the Minister of Health, after which data collection commenced, being conducted by 1600 field workers who were trained by Statistics South Africa. It has been estimated that 1.1 million houses will be visited in all 52 districts of South Africa within 90 days, in order to reach a targeted sample size of 55 120 children.

Because of the high costs associated with large national surveys,2 it is difficult for LMICs such as South Africa to conduct these surveys regularly. Yet it is these resource-constrained countries which are most in need of up-to-date valid survey data for identifying current gaps at district and facility level. To overcome these cost constraints, which may prevent the SANDoH from conducting another large national survey within the next decade, we suggest an affordable, practical alternative with high validity. The suggested methodology is based on the experiences of the South African Vaccination and Immunisation Centre, based at the Sefako Makgatho Health Sciences University, in conducting four small-scale South African vaccination coverage surveys.17,18 All of these studies had the common objectives of determining the (1) FIC of children aged 12–23 months, and (2) reasons for some children not being fully vaccinated. These studies were all funded as part of postgraduate research dissertations, with some students paying for the costs themselves, while others were funded through research grants. All studies received ethics approval from institutional review boards – the first three from the Medunsa Campus Research Ethics Committee, and the fourth from the Sefako Makgatho Health Sciences University Research Ethics Committee. In the following discussion, particular attention is paid to issues impacting on affordability and validity, with the view of suggesting a method which could form the basis for countrywide application.

The studies

All four studies adapted the WHO vaccination coverage cluster survey method, with the first three using the 2005 version,19 while the fourth study used the 2016 version (subsequently updated in 2018).20 All studies aimed for a sample size powered at 80–90% at a 95% level of confidence, with a design effect of 2 for 30 clusters.21 For all studies, consenting caregivers of children aged 12–23 months during the study period, who were in possession of their vaccination records, were included. If no eligible child (i.e. a child of the correct age with a vaccination record) was found, or if informed consent could not be obtained, the next closest house was visited. If more than one child in a household was eligible, for the first three studies, data were collected on the youngest eligible child,22 while for the last study, data were collected on all eligible children.23 All studies used researcher-administered questionnaires adapted from the WHO protocol.24

The first study18 was conducted in 2012 in Bela-Bela Township in Bela-Bela Municipality, Limpopo Province. The survey was confined to six wards where low-cost/informal housing predominates. The sample size was increased to 240 (30x8), with 30 clusters (blocks bordered by roads) of 250–300 houses per cluster being based on a map provided by Bela-Bela Municipality. The first house visited in each cluster was randomly selected from the map, and then every nth number of houses per cluster (n = 8) was visited following a pre-determined pattern. The
researcher (R1) was a nurse at a local clinic serving the community being surveyed. There was thus a risk that caregivers of unvaccinated/partially vaccinated children may fear that the vaccination status of their children would have repercussions on future health care from the clinic, thereby introducing nonresponse/volunteer bias. To avoid this, three research assistants (with a minimum qualification of matric) were recruited and trained according to the WHO protocol\(^\text{2}\) for interviewing and collecting data. R1 supervised the research assistants via cellphone calls during data collection, and at the end of each day checked questionnaires for adherence to the inclusion criteria, completeness and consistency. Questionnaires were discarded where inclusion criteria were not adhered to, and research assistants returned to the cluster to continue recruiting eligible participants. Where inconsistencies were found (e.g. subsequent dose received without previous dose), the research assistants returned to re-check the vaccination record the following day. Of all houses visited within 4 weeks, 243 had an age-eligible child with a vaccination record. Of the 243 caregivers, 3 declined to participate, giving a response rate of 98.8% (240/243). Data were captured by R1, and subsequently checked against the questionnaires by R1 to ensure reliability of data capture.

When the results of this study were shared with EPI-SA officials, they pointed out that without copies of vaccination records for data validation, interpretation of these data may not be worthwhile. Furthermore, the data were collected by research assistants who reported back to R1 daily, thus real-time supervision was lacking. To address specific issues potentially impacting negatively on validity, all subsequent studies made use of electronically mailed digital photographs taken with a cellphone camera, ensuring that data collection was conducted under constant supervision.

The second study\(^\text{20}\) was conducted in 2014 in informal settlements on plots in Muldersdrift, a relatively sparsely populated rural part of Mogale City Municipality, Gauteng Province. The Mogale City Municipality provided a list of informal settlements, including plot names, numbers of houses per plot, and geographical position of 39 plots where such settlements were documented. The 39 plots with the largest number of houses (24–1540) were selected as clusters for a sample size of 210 (30x7; 80% power). The first house visited in each plot was that closest to the plot entrance, and then every n\(^\text{th}\) (number of houses per plot ÷ 7) house was systematically visited. The researcher (R2), a community liaison officer employed by the local municipality, with no relationship to the local clinic/health services, collected data on her own. After entering vaccination record data onto the questionnaire, a photograph of the vaccination record was taken using a cellphone camera and sent to the researcher via Facebook Messenger, with the data and time of photographing recorded on the corresponding questionnaire. Of 30 clusters, 23 were successfully surveyed according to protocol. All houses were visited in the other seven clusters, with only three eligible children being found. As the demographics of the seven clusters differed substantially from the successfully surveyed clusters, the three questionnaires were excluded. The previously excluded nine clusters with 10–23 houses per cluster were then surveyed. All houses were visited, but no eligible children were found. The response rate was based on the 23 successfully surveyed clusters, in which 242 houses were visited. An eligible child was identified in 66.5% (161/242) of these houses, with all caregivers consenting to participate. The survey took 16 days, with photographs being sent to the study coordinator in real-time during the first week, and thereafter being sent in batches at the end of each day. After capturing data from the questionnaire, R2 compared these data to, and research assistants returned to the cluster to continue recruiting eligible participants. Where inconsistencies were found (e.g. subsequent dose received without previous dose), the research assistants returned to re-check the vaccination record the following day. Of all houses visited within 4 weeks, 243 had an age-eligible child with a vaccination record. Of the 243 caregivers, 3 declined to participate, giving a response rate of 98.8% (240/243). Data were captured by R1, and subsequently checked against the questionnaires by R1 to ensure reliability of data capture.

The third study\(^\text{21}\) was conducted in 2015 in Refilwe Township in Cullinan, Tshwane Municipality, Gauteng Province. The Tshwane Municipality provided an aerial photograph for dividing Refilwe’s 6111 houses into 30 clusters (blocks bordered by roads), with ~200 houses per cluster. Systematic interval sampling was planned for 210 houses as described above. The researcher (R3), a scientist employed by the South African Vaccination and Immunisation Centre, participated in and supervised data collection, and was assisted by three research assistants who were trained according to the WHO protocol.\(^\text{2}\) Field work was conducted by two teams, with R3 being responsible for overall supervision and leading one team, while a MSc student led the second team. Data collection was conducted as for the second study, except that on day one of the survey, the original plan of visiting every n\(^\text{th}\) house was abandoned as no-one was found at home in most houses. Instead, all 6111 houses were visited in a period of 6 weeks, with someone found at home in 33.4% (2041/6111). Of these, 8.5% (173/2041) had an eligible child with a consenting caregiver. An attempt to survey the four local cul-de-sacs failed as none of them required proof of vaccination from caregivers, and thus did not have copies of vaccination records. Data were independently captured by the team leaders, resulting in two data sets which were independently analysed and compared for inconsistencies. Data cleaning and validation proceeded as for the second study.

The last study\(^\text{22}\) was conducted in 2017 in Region 5 of Tshwane. Tshwane Municipality provided a map showing all houses in Region 5, which was divided into 30 clusters, with 400–520 houses per cluster for a sample size of 760 (30x26; 90% power). The first household visited in each cluster was randomly selected from the map using Research Randomiser Software\(^\text{2}\) for clusters with house numbers on the map. For informal settlements, the first house from the main road was visited. Thereafter the closest households on the right were visited until the target of 26 was reached or all the households were visited. If there was no-one home, the house was revisited the following day or weekend if the neighbours reported that there was an eligible child in the house. The researcher (R4), a pharmacist with no relationship to the local clinics/health services, participated in data collection and overall supervision of two teams as described above. Photographs of vaccination records were emailed, as recently introduced Facebook privacy policies blocked Global Positioning System (GPS) coordinates from being transmitted. Unfortunately, the number of households provided by Tshwane Municipality included vacant/undeveloped numbered stands, and gated communities/security complexes, which was discovered only during data collection. Hence there were only 24 clusters instead of 30. In addition, houses enclosed by security fencing were inaccessible. Of the houses visited, someone was found at home in 87.2% (7032/8060), with an eligible child being found in 4.7% (327/7032). Of the caregivers of eligible children, 84.4% (276/327) consented to participate in the study. Data capture, cleaning and validation proceeded as for the third study.

Discussion and recommendations

Masters-level research projects for vaccination coverage surveys

Globally, a growing number of universities offer online Master of Public Health (MPH) degrees, allowing health professionals from LMICs to acquire crucial skills and knowledge for improving the health of their populations. These mature students who enroll for the 2-year online MPH at Sefako Makgatho Health Sciences University come from several African countries, and are generally highly motivated early to mid-career health professionals. Many use the opportunity provided by data collection for their MPH projects, to reconnect with their communities by conducting community-based surveys, including vaccination coverage surveys. Similarly, pharmacists undertaking Master of Pharmacy research at Sefako Makgatho Health Sciences University tend to be early to mid-career pharmacists who are passionate about improving public health and serving the community. Other masters-level students who may be interested in conducting vaccination coverage surveys include clinicians undertaking Master of Family Medicine or Community Health degrees; and nurses undertaking Master of Community Nursing research. Because these targeted surveys are relatively small in terms of sample size and geographical area, operational costs are low and many postgraduate students are able to self-finance their data collection, while others benefit from research grants from external funding agencies.

Usefulness to health authorities

While the results of population-based vaccination coverage surveys can be used to identify gaps and strengths in immunisation programmes, these results must be based on studies with high reliability and validity. The feedback on the first study from EPI-SA resulted in vast improvements to the overall study coordinator via Facebook Messenger, with the date and time of photographing recorded on the corresponding questionnaire. Of 30 clusters, 23 were successfully surveyed according to protocol. All houses were visited in the other seven clusters, with only three eligible children being found. As the demographics of the seven clusters differed substantially from the successfully surveyed clusters, the three questionnaires were excluded. The previously excluded nine clusters with 10–23 houses per cluster were then surveyed. All houses were visited, but no eligible children were found. The response rate was based on the 23 successfully surveyed clusters, in which 242 houses were visited. An eligible child was identified in 66.5% (161/242) of these houses, with all caregivers consenting to participate. The survey took 16 days, with photographs being sent to the study coordinator in real-time during the first week, and thereafter being sent in batches at the end of each day. After capturing data from the questionnaire, R2 compared these data to those on the corresponding vaccination record photo. Finally, during data cleaning, the photos of vaccination records with inconsistent data were used for data validation.

The third study\(^\text{21}\) was conducted in 2015 in Refilwe Township in Cullinan, Tshwane Municipality, Gauteng Province. The Tshwane Municipality provided an aerial photograph for dividing Refilwe’s 6111 houses into 30 clusters (blocks bordered by roads), with ~200 houses per cluster. Systematic interval sampling was planned for 210 houses as described above. The researcher (R3), a scientist employed by the South African Vaccination and Immunisation Centre, participated in and supervised data collection, and was assisted by three research assistants who were trained according to the WHO protocol.\(^\text{2}\) Field work was conducted by
in reliability and validity, with cellphone technology allowing constant, real-time supervision. Importantly, these changes in study design have produced much more credible data from the subsequent studies without increasing the budget appreciably. Furthermore, the independent double-data capture and analysis used in the last two studies ensured reliability of data capture – an aspect of postgraduate research often overlooked by students. Also, health authorities are likely to find targeted surveys such as these very useful for pinpointing underperforming clinics. This may not be possible when conducting large national vaccination coverage surveys, as a major disadvantage of these is that they do not reflect health system performance at the local level.5

**Low-cost, high-quality data**

The high definition of cellphone photographs allows valid data to be captured in the field. As cellphones are ubiquitous in most LMICs including South Africa,6 this method of capturing data does not necessarily add to the budget. While emailing photographs to supervisors adds to the budget, it is relatively inexpensive, with 1 GB of data (far exceeding that needed for 210 vaccination record photographs) currently costing ZAR149.

**Low-cost, continuous supervision**

Using email to send real-time data ensures ongoing supervision throughout data collection, with mistakes being recognised and corrected immediately. In addition, most modern cellphones are equipped with GPS locator technology, allowing the supervisors to track the movements of data collectors, by viewing the GPS coordinates embedded in the vaccination record photographs.

**Representativeness**

Because these surveys were confined to small geographical areas, it was possible to divide the areas into contiguous clusters and sample within all clusters, instead of randomly selecting 30 clusters within a large geographical area, as happens in large national surveys.5,12 While the last three studies had high reliability and validity, none managed to reach their intended sample sizes, despite visiting all the houses in most or all clusters. A major limitation of the Refilwe study was that only 33.4% of houses had someone at home, suggesting that there may be an underrepresentation of employed caregivers in the final sample. This assumption is supported by there being four creches in the township, suggesting both a demand for child care and the means to pay for it. A previous study, conducted in Ga-Rankuwa in northwest Tshwane, found a statistically significant positive association between employment and knowledge and awareness about immunisation.23 This suggests that the FIC in this study may in fact be higher than reported.22 To avoid this selection bias in the last study, data collection was extended into early evenings and was also conducted over weekends. A major limitation of this study was that gated communities/security complexes and houses with security fencing were inaccessible. Caregivers living in these communities may have higher rates of Internet access, which may result in higher rates of vaccination hesitancy and lower FIC than reported.22

**Conclusion**

In 2019, the SANDoH will be producing South Africa’s first national vaccination coverage survey data for all districts since 1994. Thereafter, periodic coverage surveys are essential for tracking progress and identifying gaps in EPA-SA. Masters-level research can contribute greatly to public health in South Africa and other LMICs, by employing an affordable method for obtaining valid vaccination coverage figures. From our experience, the following is essential:

• Recent aerial satellite images (e.g. Google Maps19) of the site to be surveyed should be used instead of maps.
• Cellphones with GPS locator technology must be used for photographing and emailing photographs of vaccination records to supervisors, to allow for real-time remote supervision.
• For urban areas with high employment rates, data collection must not be confined to weekdays during working hours. Personal safety and security are always of concern when conducting household surveys in urban South Africa, thus it may be necessary to conduct the surveys during daytime over weekends. Unfortunately, this means that each survey will take longer than the 30 days recommended by the WHO,13 unless a larger number of research assistants are employed, which will add considerably to the budget. However, these costs could be budgeted for in a grant application, so this option is a viable one that needs to be considered.

• Innovative methods for obtaining access to gated communities/security complexes must be investigated. In the last survey, an unsuccessful attempt was made to gain access by emailing the project details and request for participation to the estate manager, following telephonic contact. Online surveys advertised on Facebook may also be an option, although the response rate to a recent human papillomavirus vaccination online survey was very low.20

**Acknowledgements**

We thank the EPI-SA team for their valuable advice on data collection, and Ms Charlotte van den Broek for assisting in data collection, capture and analysis. We especially thank the caregivers who participated in the studies, and the municipality authorities which provided permission and information. The National Research Foundation of South Africa funded data collection for the Tshwane Region 5 study, and conference attendance for that study and the Refilwe Township study.

**References**


25. Furaha SA. Immunisation coverage of children less than 5 years in Ga-Rankuwa [dissertation]. Pretoria: Tshwane University of Technology; 2009.