

## Active case-finding method improves completeness and accuracy of data reported to the rural Eastern Cape Cancer Registry in South Africa

Nontuthuzelo IM Somdyala<sup>1</sup>, Linda Mbuthini<sup>2</sup>, Borna Müller<sup>3</sup>, Nomfuneko Sithole<sup>1</sup>, Akhona Ncinitwa<sup>1</sup> and Debbie Bradshaw<sup>1,4</sup>

<sup>1</sup>South African Medical Research Council, Burden of Disease Research Unit, PO Box 19070, Tygerberg 7505, Cape Town, South Africa

<sup>2</sup>Centre for Lung Infection and Immunity, Department of Medicine, UCT Lung Institute, University of Cape Town, Observatory 7925, Cape Town, South Africa

<sup>3</sup>F. Hoffmann-La Roche Ltd, Pharmaceuticals Division, Global Access, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

<sup>4</sup>Department of Family Medicine and Public Health, University of Cape Town, Observatory 7925, Cape Town, South Africa

### Abstract

The quality and accuracy of the data provided by cancer registries has a significant impact on decision making. Over decades, high-income countries have been successful in monitoring their cancer burden because of well-established data abstraction techniques such as digital systems. Conversely, in low- and middle-income countries, sparsely distributed cancer registries, using alternative less costly, but imprecise methods are struggling to capture all cancer cases. A population-based cancer registry in South Africa covering a resource-limited rural population is faced with challenges in case finding yet the quality and accuracy of the data provided has a significant impact on decision making. The objective of this study was to assess data quality using two data quality attributes 'completeness and accuracy' and also to determine the benefits of using active and passive case-finding methods for cancer registration in this population. Data used were collected between January 2014 and December 2015 from four hospitals to compare the quality of both active and passive case-finding methods. From all four hospitals during the same period, a first set of data obtained through passive reporting was compared with a second set of data obtained through active case finding. Covering multiple facilities during active case finding can significantly improve quality of data, while passive case finding is challenged by data collection being confined to one specific health facility, only. Better investment in active case finding is recommended in settings with resource-distribution disparities.

**Keywords:** *population-based cancer registration, completeness and accuracy, active case finding, resource-distribution disparities, better investment, Africa*

### Introduction

Cancer registries play an important role in monitoring the globally increasing cancer burden. Also, they are of pivotal importance in the planning and evaluation of cancer intervention programmes [1]. Consequently, the quality and accuracy of the data provided by cancer registries has a significant impact on decision making. The International Agency

**Correspondence to:** Nontuthuzelo IM Somdyala  
Email: [nontuthuzelo.somdyala@mrc.ac.za](mailto:nontuthuzelo.somdyala@mrc.ac.za)

ecancer 2021, 15:1251  
<https://doi.org/10.3332/ecancer.2021.1251>

Published: 17/06/2021

Received: 19/01/2021

**Copyright:** © the authors; licensee ecancermedicalsecience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

for Research on Cancer (IARC) of the World Health Organization; the global cancer surveillance controlling body, sets the rules and guidelines for cancer registries. Each registry's data quality is measured against those rules and its value depends on the quality of data generated. In cancer registration, quality of data is described by four attributes: namely, comparability, accuracy/validity, completeness and timeliness [2–6]. However, with the latter, there are no international guidelines at present, although specific standards for abstraction and reporting have been set out by certain organisations [4].

Over decades, high-income countries have been successful in monitoring their cancer burden because of well-established data abstraction techniques such as digital systems [7]. Conversely, in low- and middle-income countries, sparsely distributed cancer registries, using alternative less costly, but imprecise methods are struggling to capture all cancer cases. In Africa, particularly in rural areas, cancer registries are faced with challenges in case finding which include limited resources such as laboratories and oncologist specialists for accurate diagnoses, inaccessibility of hospitals due to poor infrastructure and poor record keeping for tracing and follow-up to patients [8]. This impacts on the generation of reliable, valid and complete data.

The Eastern Cape Cancer Registry (ECCR) is the only rural population-based cancer registry in South Africa covering eight magisterial areas [9–10] with a population of 1.2 million [11]. However, merely collecting cases and establishing a cancer registry is insufficient. Cancer registries have a vested interest in promoting the use of their data and must ensure data generated are of good quality [12]. The ECCR uses both active and passive case-finding methods to mitigate the discrepancies and shortfalls of either method alone [13]. The use of both methods remains financially burdensome to the registry but beneficial in terms of data quality. The active case-finding method involves annual visits to collaborating hospitals by ECCR staff to review and extract all available information from patients' records with a cancer diagnosis [13]. While the passive case-finding method involves data collectors in only four collaborating hospitals who collect and send data to the ECCR on a monthly basis. For uniformity in both methods, a standardised data collection tool is used [14, 15]. The ECCR contributes to cancer incidence (CI) data globally and regionally; Cancer Incidence in Five Continents (CI5); volumes X and XI [16, 17], Cancer in sub-Saharan Africa [18] and survival collaborative studies; CONCORD 2 and 3 [19, 20]. Other than the achievement of these developmental milestones, no internal study has been conducted to evaluate ECCR data quality. The objective of this study was to assess the quality of data generated by ECCR in two data quality attributes 'completeness and accuracy'. Completeness is the extent to which all incident cancers occurring in a population are included in the registry [2], whereas accuracy is defined as the proportion of cases in the data with a given characteristics (e.g. topography and morphology) that truly had the attribute [4]. This study focused on these two quality attributes as an initial attempt to checking internal consistency and uniformity of registry records produced using active and passive data collection methods in a resource-limited setting.

## Materials and methods

### *Data collection and study sample*

This study was conducted between January 2014 and December 2015. In total, 15 hospitals and their associated pathology laboratories constitute the main data source used by the ECCR. Data from these hospitals generally is obtained by registry staff periodically performing patient chart reviews, also termed active case finding. Four of the 15 hospitals are resourced to proactively perform independent data abstraction and reporting to the cancer registry by hospital staff, also termed passive case finding. Data from four hospitals were used to compare the quality of both, the active and passive case-finding methods. From all four hospitals, in 2014 and 2015, a first set of data was obtained through passive reporting. The second dataset was constituted by cases collected by cancer registry staff from the same hospitals during the same period. The four hospitals included in the study were two peripheral hospitals (Tafalofefe and St Elizabeth) and two referral hospitals (Frere and Umtata General Hospital Complex (UGHC)). All types of malignant cancer cases were included and coded according to the International Coding of Disease for Oncology [20].

### *Data and analysis*

For this study, the main information used included variables for patients' identification (name, surname, address, age, sex and ethnicity), the primary organ invaded by the tumour (topography), histological classification of the cancer tissue (morphology), incidence date (first date the patient was seen by the doctor and cancer diagnosed) and healthcare facility in which the patient was first diagnosed with cancer. Variables

selected constitute the mandatory variables expected for each case to be accepted as valid in the database. The complete standardised data collection tool used during both active and passive case finding is attached ([Appendix 1: Confidential cancer notification form](#)). Before analysis, duplicates were cleaned once the patient's information has been consolidated.

(i) *Dataset to assess completeness*

For each hospital, completeness of active and passive case finding was assessed by calculating the number of patients captured during 2014 and 2015 by either method divided by the total number of patients identified through both methods combined; new cases active =  $N_a$  and new cases passive =  $N_p$

( $C = N_a T_{cases} * 100$  and  $C = N_p T_{cases} * 100$ ). Patients captured by both methods were identical if all identification variables were matching.

(ii) *Dataset to assess accuracy*

To assess accuracy of information on topography and morphology, data were anonymised by assigning each patient a registry number as identification. Only 1,040 cases recorded by both methods (active and passive) were included in the analysis ([Figure 1](#)). Of these, 10% ( $N = 104$ ) were randomly selected to assess and calculate the agreement in information provided with regard to topography and morphology received through both case-finding methods ([Figure 1](#)). Three reviewers independently assessed these records and if results differed re-examination was carried out. Subsequently, analysis using Microsoft Excel 2016 was done. Testing for statistically significant differences between active and passive case-finding methods and between hospitals was done in R using McNemar's and Pearson's Chi-squared tests, respectively ([Table 1a](#) and [b](#)).

**Table 1. (a) How did the performance of active and passive case finding differ between peripheral and referral hospitals?**

By year, *p*-values for pairwise comparison of the sensitivity of active and passive case finding between peripheral and referral hospitals using Pearson's Chi-squared test with Yates' continuity correction

Year	Active	Passive
2014	0.54	<b>0.01</b>
2015	0.37	<b>0.04</b>
Both years	0.25	<b>0.00</b>

*p*-values reaching statistical significance are highlighted in bold

**Table 1. (b) How did the performance of active case finding differ between hospitals?**

*p*-values for pairwise comparison of the sensitivity of active case finding between hospitals (all years combined) using Pearson's Chi-squared test with Yates' continuity correction

	St Elizabeth	Tafalofefe	UGHC	Frere
St Elizabeth		0.92	0.09	0.96
Tafalofefe			0.18	0.95
UGHC				<b>0.03</b>
Frere				

*p*-values reaching statistical significance are highlighted in bold

### How did the performance of passive case finding differ between hospitals?

*p*-values for pairwise comparison of the sensitivity of passive case finding between hospitals (all years combined) using Pearson's Chi-squared test with Yates' continuity correction

	St Elizabeth	Tafalofefe	UGHC	Frere
St Elizabeth		<b>0.05</b>	<b>0.01</b>	<b>0.00</b>
Tafalofefe			0.84	0.28
UGHC				<b>0.01</b>
Frere				

*p*-values reaching statistical significance are highlighted in bold

## Ethics approval

ECCR is an ongoing project of the South African Medical Research Council (SAMRC). The main objective is to generate CI in a defined area. Approval was received from the SAMRC Ethics Committee and permission was sought from the Eastern Cape Department of Health Research Committee. For any additional research which deviates from the original, the principal investigator is expected to review the project proposal. However, for this study, no additional ethics approval was needed as this was a secondary data analysis.

## Results

### Completeness

We compared cancer registry data obtained through both active and passive case-finding methods from four hospitals in the Eastern Cape Province of South Africa. These consisted of two peripheral hospitals (Tafalofefe and St Elizabeth) and two referral hospitals (Frere and UGHC). During the study period (2014–2015), using active and passive case finding combined, a total of 2,961 individual cancer cases were identified at these four hospitals (Figure 1, Table 2). The number of cases jointly identified by both case-finding methods was 2,176 (74% of the total amount of cases diagnosed; Table 2, Figure 2a and b). Neither of the two methods alone identified all diagnoses reported. However, among all cancer cases diagnosed, active case-finding identified a significantly higher proportion than that from passive case finding ( $p < 0.05$ ; Tables 1a, b and 3).

For the peripheral hospitals St Elizabeth and Tafalofefe, respectively, active case-finding identified 94% and 96% of all cases diagnosed (95% for both peripheral hospitals combined; Table 2). In contrast, passive case finding covered significantly fewer cases among all cases identified (66% and 85%, respectively or 71% for both peripheral hospitals combined;  $p < 0.05$ ; Table 2). Combined for the two peripheral hospitals, adding active case finding to the passive case-finding method increased the number of cancer cases identified by 41% (529 cases were reported through passive case finding alone and 215 cases were added by including active case finding; Tables 2 and 3). In contrast, adding passive case finding to the active case-finding method increased case detection by 6%, only (703 cases were reported through active case finding alone and 41 cases were added by including passive case finding).

The significantly higher performance of active over passive case finding for the peripheral hospitals was not observed for the two referral hospitals Frere and UGHC. For these hospitals combined, among all cancer cases diagnosed, the proportion of cases captured by active or passive case finding alone was very similar (88%; Table 2). Also, adding active case finding to the passive case-finding method still increased the number of cancer cases identified by 13% (Tables 2 and 3).

### Accuracy

An assessment of the accuracy of information abstracted from the medical records by each method was done with regard to data on tumour topography and morphology. The overall proportion of patient records matching for tumour topography was 89% in 2014 and 90% in 2015 (Table 4). However, agreement for morphology was less with 83% in 2014 and 76% in 2015 (79% for both years combined; Table 5, Figure 3).

Specifically, for data on tumour topography, at both peripheral hospitals, agreement of active and passive case-finding methods was 100%; however, at a very low sample size ( $N = 19$  records in total). For referral hospitals, agreement was 87% at UGHC and 92% at Frere (89% in total;  $N = 85$ ; Table 4).

For data on tumour morphology, at peripheral hospitals, agreement was 82% at St Elizabeth and 75% at Tafalofefe (79% in total;  $N = 19$ ; Table 5). For referral hospitals, agreement of active and passive case finding was 71% at UGHC and 85% at Frere (79% in total;  $N = 85$ ; Table 5).

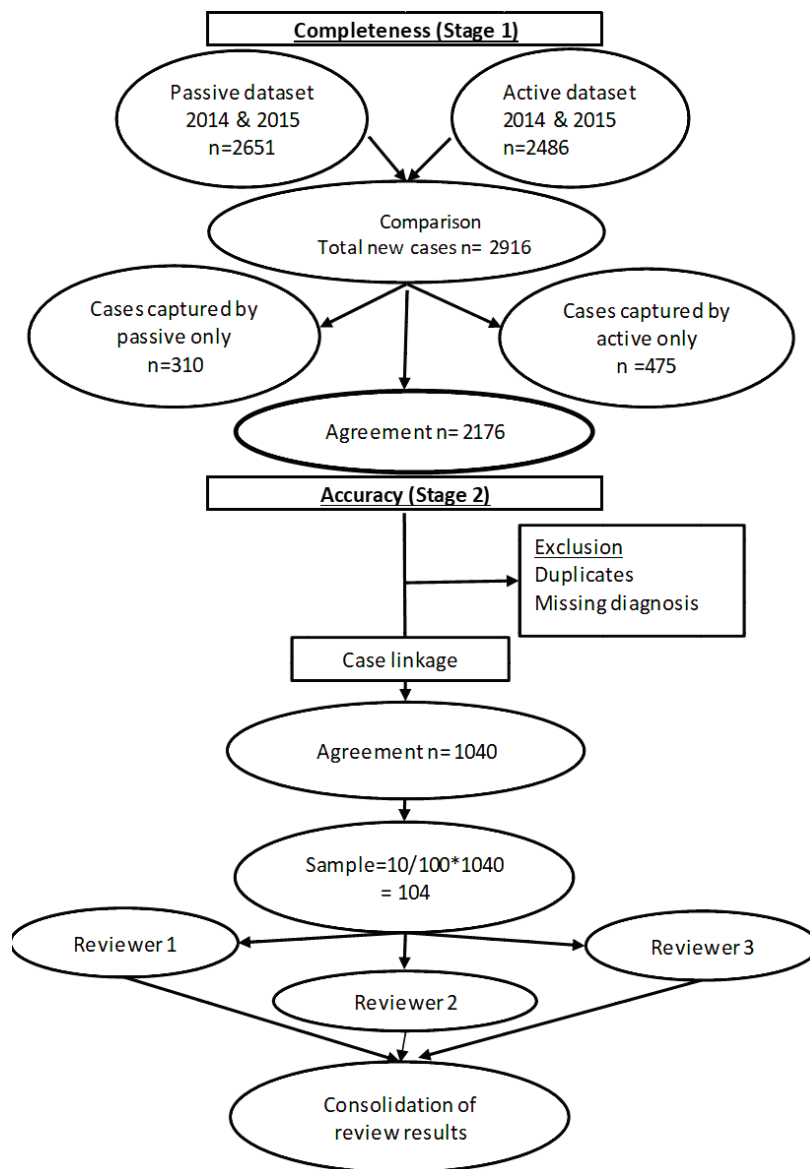


Figure 1. Data evaluation process.

Table 2. Cancer cases registered by detection method, hospital and year

Case detection method	St Elizabeth*		Tafalofefe*		UGHC**		Frere**		Peripheral		Referral		Grand Total	
	Freq	Se	Freq	Se	Freq	Se	Freq	Se	Freq	Se	Freq	Se	Freq	Se
<b>2014</b>														
Active	249	96.1%	107	93.0%	327	83.4%	471	94.8%	356	95.2%	798	89.8%	1154	91.4%
Passive	168	64.9%	94	81.7%	331	84.4%	473	95.2%	262	70.1%	804	90.4%	1066	84.4%
Active not passive	91	35.1%	21	18.3%	61	15.6%	24	4.8%	112	29.9%	85	9.6%	197	15.6%
Passive not active	10	3.9%	8	7.0%	65	16.6%	26	5.2%	18	4.8%	91	10.2%	109	8.6%
Active and passive	158	61.0%	86	74.8%	266	67.9%	447	89.9%	244	65.2%	713	80.2%	957	75.8%
Active or passive (all cases)	259	100.0%	115	100.0%	392	100.0%	497	100.0%	374	100.0%	889	100.0%	1263	100.0%
<b>2015</b>														
Active	250	91.9%	97	99.0%	691	82.1%	459	94.4%	347	93.8%	1150	86.6%	1497	88.2%
Passive	181	66.5%	86	87.8%	683	81.1%	470	96.7%	267	72.2%	1153	86.8%	1420	83.6%
Active not passive	91	33.5%	12	12.2%	159	18.9%	16	3.3%	103	27.8%	175	13.2%	278	16.4%
Passive not active	22	8.1%	1	1.0%	151	17.9%	27	5.6%	23	6.2%	178	13.4%	201	11.8%
Active and passive	159	58.5%	85	86.7%	532	63.2%	443	91.2%	244	65.9%	975	73.4%	1219	71.8%
Active or passive (all cases)	272	100.0%	98	100.0%	842	100.0%	486	100.0%	370	100.0%	1328	100.0%	1698	100.0%
<b>Both years combined</b>														
Active	499	94.0%	204	95.8%	1018	82.5%	930	94.6%	703	94.5%	1948	87.9%	2651	89.5%
Passive	349	65.7%	180	84.5%	1014	82.2%	943	95.9%	529	71.1%	1957	88.3%	2486	84.0%
Active not passive	182	34.3%	33	15.5%	220	17.8%	40	4.1%	215	28.9%	260	11.7%	475	16.0%
Passive not active	32	6.0%	9	4.2%	216	17.5%	53	5.4%	41	5.5%	269	12.1%	310	10.5%
Active and passive	317	59.7%	171	80.3%	798	64.7%	890	90.5%	488	65.6%	1688	76.1%	2176	73.5%
Active or passive (all cases)	531	100.0%	213	100.0%	1234	100.0%	983	100.0%	744	100.0%	2217	100.0%	2961	100.0%

\*Peripheral hospitals

\*\*Referral hospitals

Freq: Absolute number of cases detected

Se: Sensitivity of case detection method for hospital and period indicated

Active: Cases registered by applying active case finding methods

Passive: Cases registered by applying passive case finding methods

## Discussion

The main objective of this study was to assess the completeness and accuracy of cancer patients' data reported to a rural population-based cancer registry in South Africa. Data on the burden of cancer in Sub-Saharan Africa are very scarce and the limited information available is mostly biased towards urban centres with better infrastructure and possibly distinct disease epidemiology. Our study contributes significantly to the understanding of the importance of CI even in rural populations of Africa with resource limitations.

The overall results of the study indicated less variations with regard to completeness compared to accuracy; with a range of 83%–96% in referral hospitals and 83% in peripheral hospitals (Figure 4). However, differences were noted in individual hospitals with a range of 65%–

97%. Neither of the two methods alone identified all patients reported. Since registry staff reviewed the same records used by data collectors during data abstraction, human error in recording resulted in missing cases in either case-finding method. However, in the entire study, active case-finding contributed more cases than passive ( $p < 0.05$ ). We also observed that for the resource-constrained peripheral hospitals, active case-finding captured a significantly larger portion of the total number of patients recorded than passive case finding (95% versus 71%). Hence, active case finding significantly improved case detection; increasing the number of cases identified by as much as 41%. At better equipped referral hospitals, passive and active case finding identified a similar proportion of patients among all cases recorded (88%). Nonetheless, also for referral hospitals, active case finding significantly increased the number of patients captured by 13%.

A higher accuracy agreement for data on topography was observed with disagreement in only 17%–21% of records which is not a nugatory result especially in under-resourced settings where there is manual data collection. Passive case-finding led to poor information recorded on cancer morphology compared to active. Disagreement was frequently based on different or incorrect information recorded regarding diagnoses. A higher record of accuracy was observed in referral compared to peripheral hospitals. Frere Hospital, which is a referral, showed the highest accuracy in recording. This hospital unlike the others has an in-house pathology laboratory linked data to patients' records. The hospital is also fully equipped with oncology and radiation units and several practicing oncologists and registrars. Similar performance of passive reporting (Appendix 2: Resource differences between hospitals observed during the study, possibly impacting quality of passive reporting) was reported at UGHC Hospital which is a referral hospital and at Tafalofefe Hospital which although is a peripheral hospital has a similar infrastructure which is supportive and relevant to cancer registration. In contrast, St Elizabeth Hospital had the lowest completeness for passive case-finding (Table 2) due to frequently interrupted data submissions, incorrect record of information regarding diagnosis which raised some concern. Information regarding high morphology disagreement reflects the absence of oncology trained specialists at both peripheral hospitals and lack of follow-up of referred cases so that information about diagnosis is completed and improved.

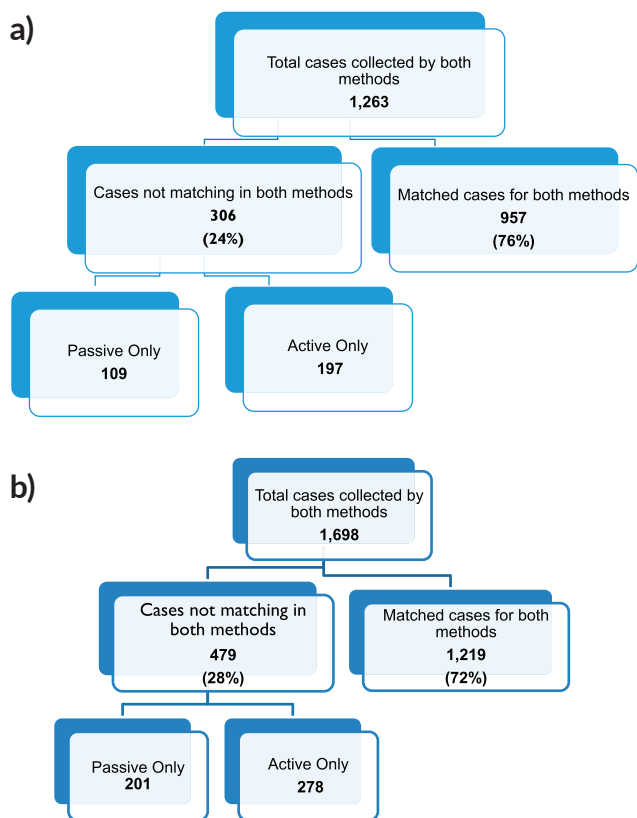


Figure 2. Sorting process for completeness check of case finding in (a): year 2014, (b): year 2015.

**Table 3. Improvement of case registration by active case finding method**

Year	St Elizabeth*	Tafalofefe*	UGHC**	Frere**	Peripheral	Referral	All
2014	54.2%	22.3%	18.4%	5.1%	42.7%	10.6%	18.5%
2015	50.3%	14.0%	23.3%	3.4%	38.6%	15.2%	19.6%
Both years combined	52.1%	18.3%	21.7%	4.2%	40.6%	13.3%	19.1%

The proportions indicate the cases added by active case finding relative to the number reported by passive case finding

\*Peripheral hospitals

\*\*Referral hospitals

**Table 4. Agreement of topographic data between active and passive case finding.**

Case detection method	St Elizabeth*		Tafalofefe*		UGHC**		Frere**		Peripheral		Referral		Grand Total	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
<b>2014</b>														
Agreement	5	100.0%	4	100.0%	10	83.3%	23	92.0%	9	100.0%	33	89.2%	42	91.3%
Disagreement	0	0.0%	0	0.0%	2	16.7%	2	8.0%	0	0.0%	4	10.8%	4	8.7%
<b>Total</b>	<b>5</b>	<b>0.0%</b>	<b>4</b>	<b>0.0%</b>	<b>12</b>	<b>16.7%</b>	<b>25</b>	<b>8.0%</b>	<b>9</b>	<b>0.0%</b>	<b>37</b>	<b>10.8%</b>	<b>46</b>	<b>8.7%</b>
<b>2015</b>														
Agreement	6	100.0%	4	100.0%	23	88.5%	20	90.9%	10	100.0%	43	89.6%	53	91.4%
Disagreement	0	0.0%	0	0.0%	3	11.5%	2	9.1%	0	0.0%	5	10.4%	5	8.6%
<b>Total</b>	<b>6</b>	<b>0.0%</b>	<b>4</b>	<b>0.0%</b>	<b>26</b>	<b>11.5%</b>	<b>22</b>	<b>9.1%</b>	<b>10</b>	<b>0.0%</b>	<b>48</b>	<b>10.4%</b>	<b>58</b>	<b>8.6%</b>
<b>Both years combined</b>														
Agreement	11	100.0%	8	100.0%	33	86.8%	43	91.5%	19	100.0%	76	89.4%	95	91.3%
Disagreement	0	0.0%	0	0.0%	5	13.2%	4	8.5%	0	0.0%	9	10.6%	9	8.7%
<b>Total</b>	<b>11</b>	<b>100.0%</b>	<b>8</b>	<b>100.0%</b>	<b>38</b>	<b>100.0%</b>	<b>47</b>	<b>100.0%</b>	<b>19</b>	<b>100.0%</b>	<b>85</b>	<b>100.0%</b>	<b>104</b>	<b>100.0%</b>

\*Peripheral hospitals

\*\*Referral hospitals

Freq: Absolute number of cases

%: Percentage relative to total for the year and hospital indicated

active/passive: Information registered by applying active/passive case finding methods

This is the first study of its kind in South Africa and one of very few in Africa due to the general scarcity of cancer registries in this continent. The results highlight the challenges experienced in achieving cancer registration with stable good quality data. These include lack of funding to support these registries, political will and commitment to support data generation activities. Consequently, there is no infrastructure (geographical remoteness, no transport, etc.), human (oncologists) and material resources (laboratories, proper documentation, etc.) and inaccessibility of health medical care centres.

Similar studies include one by Al-Haddad *et. al.* [8] whose findings in Nigeria are similar to our study, because both found evidence of incompleteness in case finding. This was attributed to infrastructure challenges including human and material resources for proper diagnosis and data recording. There is urban bias observed in another study conducted in Tanzania; of the Kilimanjaro Cancer Registry [22]. High performance was evident with 98% completeness and 94% accuracy and can be linked with the number of specialists in the hospital which tends to improve the quality of data collection [4]. The same was observed in two recent studies from the high-income countries: Singapore (Fung *et. al.* [23]) and Switzerland (Wanner *et. al.* [24]). In both studies, infrastructure is not an issue, but these registries showcase the importance of high-quality data which is important for confidence in published results and receiving trust of those who use data including public health planners.



Table 5. Agreement of morphological data between active and passive case finding.

Case detection method	St Elizabeth*		Tafalofefe*		UGHC**		Frere**		Peripheral		Referral		Grand Total	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
<b>2014</b>														
Agreement	4	80.0%	2	50.0%	10	83.3%	22	88.0%	6	66.7%	32	86.5%	38	82.6%
Disagreement	1	20.0%	2	50.0%	2	16.7%	3	12.0%	3	33.3%	5	13.5%	8	17.4%
<b>Total</b>	<b>5</b>	<b>20.0%</b>	<b>4</b>	<b>50.0%</b>	<b>12</b>	<b>16.7%</b>	<b>25</b>	<b>12.0%</b>	<b>9</b>	<b>33.3%</b>	<b>37</b>	<b>13.5%</b>	<b>46</b>	<b>17.4%</b>
<b>2015</b>														
Agreement	5	83.3%	4	100.0%	17	65.4%	18	81.8%	9	90.0%	35	72.9%	44	75.9%
Disagreement	1	16.7%	0	0.0%	9	34.6%	4	18.2%	1	10.0%	13	27.1%	14	24.1%
<b>Total</b>	<b>6</b>	<b>16.7%</b>	<b>4</b>	<b>0.0%</b>	<b>26</b>	<b>34.6%</b>	<b>22</b>	<b>18.2%</b>	<b>10</b>	<b>10.0%</b>	<b>48</b>	<b>27.1%</b>	<b>58</b>	<b>24.1%</b>
<b>Both years combined</b>														
Agreement	9	81.8%	6	75.0%	27	71.1%	40	85.1%	15	78.9%	67	78.8%	82	78.8%
Disagreement	2	18.2%	2	25.0%	11	28.9%	7	14.9%	4	21.1%	18	21.2%	22	21.2%
<b>Total</b>	<b>11</b>	<b>100.0%</b>	<b>8</b>	<b>100.0%</b>	<b>38</b>	<b>100.0%</b>	<b>47</b>	<b>100.0%</b>	<b>19</b>	<b>100.0%</b>	<b>85</b>	<b>100.0%</b>	<b>104</b>	<b>100.0%</b>

\*Peripheral hospitals

\*\*Referral hospitals

Freq: Absolute number of cases

%: Percentage relative to total for the year and hospital indicated

active/passive: Information registered by applying active/passive case finding methods

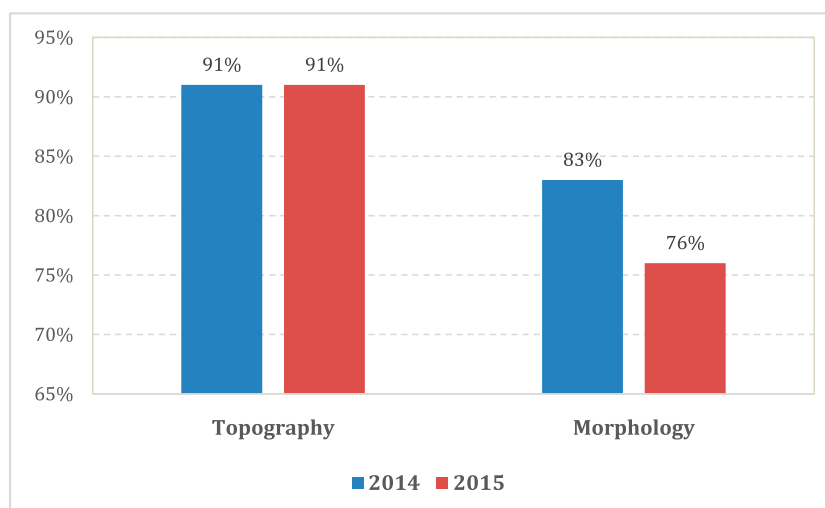


Figure 3. Accuracy; proportion of patient records matched for topography and morphology for both active and passive data case-finding methods by year.

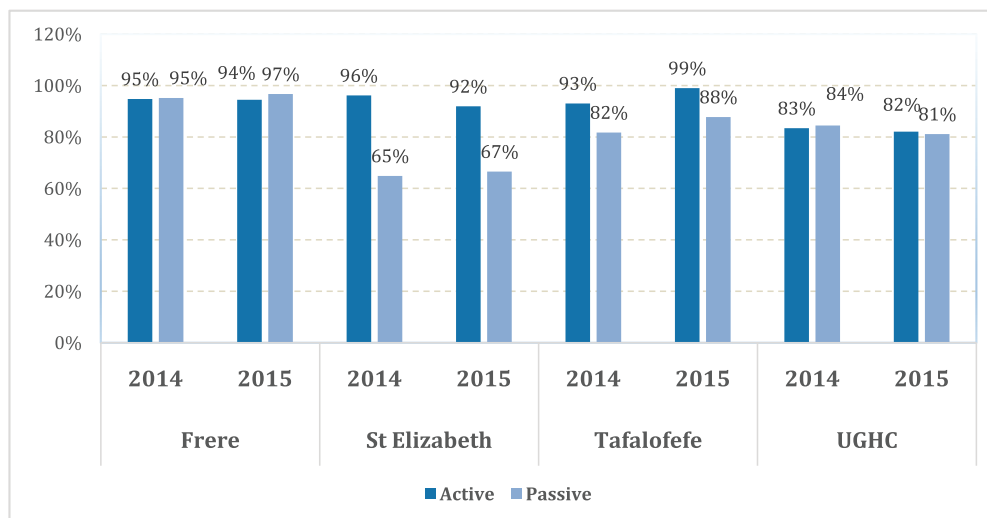


Figure 4. Completeness; proportion of cancer cases reported by both active and passive case-finding methods by facility.

## Limitations

Our study has some limitations. The study population was small; only two peripheral and two referral hospitals were included, which may only be marginally representative of the reporting capabilities of other hospitals in South Africa. Furthermore, their results cannot be compared within the ECCR as they are the only hospitals using the passive case-finding method which constitute only 20% of the total case-finding methods. Patients were double counted especially those collected by active case finding. However, use of multiple sources improves the quality of diagnoses. These limitations did not impact on the general findings of our study but highlighted the challenges of passive reporting especially in resource-limited areas.

## Conclusion

Distinct hospital-specific challenges in cancer registration were observed, particularly in peripheral hospitals. More than a third of the patients diagnosed would not have been reported. Infrequent reporting is associated with staff shortage resulting in delays in reporting and missing information in some cases. This study highlights the importance of regular training for data collectors and retain them in the project for longer time exposure and experience to improve their understanding of concepts such as 'completeness and accuracy' in cancer registration and impact of these in cancer epidemiology. Furthermore, the passive case-finding method alone has negative implications for the quality of data. Two methods used complement each other particularly in settings like South Africa with disparities in resource distribution. It was also evident that utilising and access to multiple sources including pathology laboratories during active case finding improved data quality in the ECCR.

Cancer registration in Africa is possible but it has proved to be a difficult task due to many factors, some of which include: difficulty in tracing cases, lack of medical information systems and lack of funding or underfunding. Africa must seriously consider investing in special staff training for efficient digital information exchange between external pathology laboratories and hospitals to address these challenges. Investment in active case finding will be cost effective and is recommended.

## Acknowledgments, funding and conflicts of interest

This study was sponsored entirely by the SAMRC. Co-author Dr Borna Müller is employed by F. Hoffmann-La Roche Ltd, a pharmaceutical producer of cancer drugs and diagnostics. No other co-author nor the project received financial support from Roche.

This study was presented in part as a poster at the African Organization for Research and Training in Cancer (AORTIC), 11th International Conference, Kigali Rwanda in 2017.

The following colleagues are acknowledged for their valuable inputs: Mr Steady Chasimpha of Malawi Cancer Registry for advice and guidance at the early conception of this study, data collectors in the four selected hospitals without whom data would be incomplete, collaborating hospitals in and outside the registration area for co-operation and support during data collection, Mr Thendo Ramaliba for the technical support, Miss Ngcwalisa Jama for editing the earliest version of the draft manuscript and IARC and African Cancer Registry Network for technical training skills and support.

The authors have no conflicts of interest to declare.


## References

1. Bray F, Znaor A, and Cueva P, *et al* (2014) *Planning and Developing Population-Based Cancer Registration in Low-and Middle-Income Settings* (Lyon: IARC)
2. Parkin DM and Bray F (2009) **Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness** *Eur J Cancer* **45**(5) 756–764 <https://doi.org/10.1016/j.ejca.2008.11.033> PMID: [19128954](https://pubmed.ncbi.nlm.nih.gov/19128954/)
3. Zullig LL, Schroeder K, and Nyindo P, *et al* (2016) **Validation and quality assessment of the Kilimanjaro cancer registry** *J Glob Oncol* **2**(6) 381–386 <https://doi.org/10.1200/JGO.2015.002873> PMID: [28717724](https://pubmed.ncbi.nlm.nih.gov/28717724/) PMCID: [5493247](https://pubmed.ncbi.nlm.nih.gov/5493247/)
4. Bray F and Parkin DM (2009) **Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness** *Eur J Cancer* **45**(5) 747–755 <https://doi.org/10.1016/j.ejca.2008.11.032> PMID: [19117750](https://pubmed.ncbi.nlm.nih.gov/19117750/)
5. Shimakawa Y, Bah E, and Wild CP, *et al* (2013) **Evaluation of data quality at the Gambia national cancer registry** *Int J Cancer* **132**(3) 658–665 <https://doi.org/10.1002/ijc.27646>
6. Parkin DM, Chen VW, and Ferlay J, *et al* (1994) *Comparability and Quality Control in Cancer Registration* (Lyon: IARC)
7. Brewster D, Crichton J and Muir C (1994) **How accurate are Scottish cancer registration data?** *Br J Cancer* **70** 954 <https://doi.org/10.1038/bjc.1994.428> PMID: [7947104](https://pubmed.ncbi.nlm.nih.gov/7947104/) PMCID: [2033548](https://pubmed.ncbi.nlm.nih.gov/2033548/)
8. Al-Haddad BJS, Jedy-Agba E, and Oga E, *et al* (2015) **Comparability, diagnostic validity and completeness of Nigerian cancer registries** *J Cancer Epidemiol* **39**(3) 456–464 <https://doi.org/10.1016/j.canep.2015.03.010>
9. Somdya NI, Parkin DM, and Sithole N, *et al* (2015) **Trends in cancer incidence in rural Eastern Cape Province; South Africa, 1998–2012** *Int J Cancer* **136**(5) E470–E474 <https://doi.org/10.1002/ijc.29224>
10. Somdya NI, Bradshaw D, and Gelderblom WCA, *et al* (2010) **Cancer incidence in a rural population of South Africa, 1998–2002** *Int J Cancer* **127**(10) 2420–2429 <https://doi.org/10.1002/ijc.25246> PMID: [20162610](https://pubmed.ncbi.nlm.nih.gov/20162610/)
11. Statistics SA (2013) *Census 2011 Municipal Report Eastern Cape* (Pretoria: Statistics South Africa)
12. Curado PM, Voti L and Sortino-Rachou AM (2009) **Cancer registration data and quality indicators in low- and middle-income countries; their interpretation and potential use for improvement cancer care** *Cancer Causes Control* **20** 751–756 <https://doi.org/10.1007/s10552-008-9288-5>
13. Powel J (1991) **Data sources and reporting** *Cancer Registration, Principles and Methods* eds OM Jensen, DM Parkin, R MacLennan, CS Muir, RG Skeet (Lyon: IARC)

14. MacLennan R (1991) **Items of patient information which may be collected by registries** *Cancer Registration, Principles and Methods* eds OM Jensen, DM Parkin, R MacLennan, CS Muir, RG Skeet (Lyon: IARC)
15. Finesse AM, Somdyala N, and Chokunonga E, *et al* eds (2015) *Standard Procedure Manual for Population-based Cancer Registries in Sub-Saharan Africa* (Oxford: African Cancer Registry Network) [<http://www.afcrn.org/index.php>]
16. Somdyala NIM, Gelderblom WCA, and Bradshaw D, *et al* (2013) **South Africa, Eastern Cape;2003-2007 Cancer Incidence in Five Continents** vol X, eds B Kohler, C Gombe Mbalawa, and D Forman, *et al* (electronic version) (Lyon, IARC) [<http://Ci5.iarc.fr>]
17. Somdyala NIM, Bradshaw D and Sithole N (2017) **South Africa, Eastern Cape (2008-2012) Cancer Incidence in Five Continents** vol XI, eds F Bray, M Colombet, L Mery, M Piñeros, A Znaor, R Zanetti, J Ferlay (electronic version) (Lyon, IARC) [<http://Ci5.iarc.fr>]
18. Parkin DM, Ferlay J, and Jemal A, *et al* (2018) *Cancer in Sub-Saharan Africa* (Lyon: IARC)
19. Allemani C, Weir HK, and Carreira H, *et al* (2015) **Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2)** *Lancet* 385(9972) 977–1010 [https://doi.org/10.1016/S0140-6736\(14\)62038-9](https://doi.org/10.1016/S0140-6736(14)62038-9) PMID: 4588097
20. Allemani C, Matsuda T, and Di Carlo V, *et al* (2018) **Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries** *Lancet* 391(10125) 1023–1075 [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3) PMID: 29395269 PMID: 5879496
21. Fritz A, Percy C, and Jack A, *et al* (2000) *International Classification of Diseases for Oncology* (Geneva: World Health Organization)
22. Zullig LL, Schroeder K, and Nyindo P. *et al* (2016) **Validation and quality assessment of the Kilimanjaro cancer registry** *Glob J Oncol* 2(6) 381–386 <https://doi.org/10.1200/JGO.2015.002873>
23. Fung JWM, Lim SBL, and Zheng H, *et al* (2016) **Data quality at the Singapore cancer registry; an overview of comparability, completeness, validity and timeliness** *J Cancer Epidemiol* 43 76–86 <https://doi.org/10.1016/j.canep.2016.06.006>
24. Wanner M, Matthes KL, and Korol D, *et al* (2018) **Indicators of data quality at the cancer registry Zurich and Zug in Switzerland** *Biomed Res Int* 2018 7656197 <https://doi.org/10.1155/2018/7656197> PMID: 30009174 PMID: 6020656

## Appendixes

### Appendix 1: Confidential cancer notification form

		<b>Cancer Registry Number:</b> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
<b>EASTERN CAPE CANCER REGISTRY</b> <b>Confidential Cancer Notification Form</b>		
P.O Box 19070 Tygerberg 7505, South Africa    Tel: 021 9380954    Email: Nontuthuzelo.Somdyala@mrc.ac.za		
<b>1. PARTICULARS</b>		
Name of Health Facility/Hospital _____		
Surname _____		
Full names _____		
Date of Birth	<input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Estimated Age <input type="text"/> <input type="text"/> Sex <input type="checkbox"/> M <input type="checkbox"/> F
Folder number	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Pathology Number <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
I.D Number /Passport Number <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
Ethnic Group: Black <input type="checkbox"/> Indian <input type="checkbox"/> White <input type="checkbox"/> Coloured <input type="checkbox"/> Other <input type="checkbox"/>		
<b>2. CURRENT AREA OF RESIDENCE</b>		<b>3. ORIGINAL AREA OF RESIDENCE</b>
City /Town/Village _____		City /Town/Village _____
Postal Code <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		Postal Code <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
How long have you been on this address? <input type="text"/> <input type="text"/>		How long have you been on this address? <input type="text"/> <input type="text"/>
Days/Months/Years		Days/Months/Years
<b>3. TUMOUR INFORMATION</b>		
Date of Incidence	<input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Basis of Diagnosis: <input type="checkbox"/> 0.Deth Certificate Only <input type="checkbox"/> 1.Clinical Only
Primary Site of the Tumour _____		
Morphology _____    M <input type="text"/> <input type="text"/> / <input type="text"/> 1= Pos		
Stage: <input type="checkbox"/> I <input type="checkbox"/> II <input type="checkbox"/> III <input type="checkbox"/> IV <input type="checkbox"/> 9 Unknown    HIV Status <input type="checkbox"/> 2=Neg <input type="checkbox"/> 9=Unknown		
Grade: <input type="checkbox"/> Well differentiated <input type="checkbox"/> Moderately differentiated <input type="checkbox"/> Poorly differentiated <input type="checkbox"/> Unknown		
Laterality <input type="checkbox"/> 1 = Right    2 = left    3 = Bilateral    9 = Unknown    Stage <input type="text"/> <input type="text"/> T: <input type="text"/> <input type="text"/> N: <input type="text"/> <input type="text"/> M: <input type="text"/> <input type="text"/>		
Notes: _____		
<b>4. TREATMENT</b> <input type="checkbox"/> Yes=1    No=2    Unkown=9		
Surgery <input type="checkbox"/> Radiotherapy <input type="checkbox"/> Chemotherapy <input type="checkbox"/> Hormonal Therapy <input type="checkbox"/>		
Immunotherapy <input type="checkbox"/> Palliative <input type="checkbox"/> Other: Specify _____		
<b>5. FOLLOW UP STATUS</b> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		
Date of last contact _____    Status at last contact <input type="checkbox"/> 1=Alive    2=Dead    9= Unknown		
Cause of Death (if deceased only) <input type="checkbox"/> 1=Due to Cancer    2=Other    9= Unknown		
<b>6. INFORMANT PARTICULARS</b>		
Form completed by: _____		Date: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Signature: _____
Data captured by: _____		Date: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Signature _____

Clinical Study

**Appendix 2. Resource differences between hospitals observed during the study, possibly impacting quality of passive reporting.**

Parameter	St Elizabeth*	Tafalofefe*	UGHC**	Frere**
Completeness of passive reporting***	65.7%	84.5%	82.2%	95.9%
Pathology laboratory data linked to hospital patient data	✘	✘	✘	✓
Oncologists present during study period	✘	✘	✘	✓
Oncology-trained nurses present during study period	✓	✓	✓	✓
Uninterrupted reporting of new cancer cases	✘	✓	✓	✓

\*Peripheral hospitals

\*\*Referral hospitals

\*\*\*Proportion of cases reported by passive case finding among all patients captured over both study years combined