

Stark differences in cancer epidemiological data between GLOBOCAN and GBD: Emphasis on oral cancer and wider implications

K.M. Fan,^{a,1} J. Rimal,^{b,1} P. Zhang,^{b,1} and N.W. Johnson^{a,b,c,1*}

^aSchool of Medicine & Dentistry, Griffith University Queensland, Australia

^bMenzies Health Institute Queensland, Griffith University Queensland, Australia

^cFaculty of Dentistry Oral and Craniofacial Sciences, King's College London, England

Summary

Background GLOBOCAN 2020 and Global Burden of Disease (GBD) 2019 are the two most established global online cancer databases. It is important to examine the differences between the two platforms, to attempt to explain these differences, and to appraise the quality of the data. There are stark differences for lip and oral cancers (LOC) and we attempt to explain these by detailed analysis of ten countries at the extremes of differences.

Methods Age-standardised incidence rates (ASIR) of LOC were obtained from GLOBOCAN 2020 and GBD 2019. Five countries with the greatest and smallest fold differences were selected. A systematic search of PubMed and Embase electronic databases was then performed to identify publications reporting the incidence of LOC in the selected countries between 2015 and 2022. Specifically, data sources of the articles were examined and evaluated.

Findings For LOC, greatest differences were found in Papua New Guinea, Vietnam, China, Pakistan, and Indonesia (group A). In contrast, the United States of America (USA), Brazil, France, Germany, and India (group B) had the least differences between the two databases.

Interpretation It is not surprising that when GLOBOCAN and GBD could not obtain high-quality or accessible LOC data from national or local cancer registries, as in group A, discrepancies would be seen between the two online databases. In contrast, where only minor differences were seen between GLOBOCAN and GBD, as in group B, presumably due to those countries having well-established cancer registries and healthcare administrative systems, the literature is more consistent. Moreover, many studies have grouped lip and oral cavity with pharynx and categorised outputs as “oral and oropharyngeal cancer” or “oral cavity and pharynx cancer”. Those categorisations lacked subsite accuracy and failed to realise that oral cancer and oropharyngeal cancer have completely different etiological factors, pathogenesis, prognosis, and treatment outcomes.

Funding This research received no specific grant or funding from any funding agency in the public, commercial, or not-for-profit sectors, and the authors received no financial support for the research, authorship, and/or publication of this article.

Copyright © 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Global Cancer Observatory (GLOBOCAN); Global Burden of Disease (GBD); Lip and oral cavity cancer; Oral cancer; Incidence

Abbreviations: ASR/ASIR, age-standardised incidence rates; CR, cancer registry/registries; GBD, Global Burden of Disease; GHDx, Global Health Data Exchange; GLOBOCAN, Global Cancer Observatory; IARC, International Agency for Research on Cancer; ICD-10, 10th revision of the International Statistical Classification of Diseases and Related Health Problems; IHME, Institute for Health Metrics and Evaluation; LMICs, low- and middle-income countries; LOC, lip and oral cavity cancers; PBCR, population-based cancer registry; PNG, Papua New Guinea; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; USA, United States of America

*Corresponding author.

E-mail address: n.johnson@griffith.edu.au (N.W. Johnson).

¹ Address for all authors: Griffith University, Parklands Drive, Southport, QLD 4222, Australia.

eClinicalMedicine

2022;54: 101673

Published online xxx

[https://doi.org/10.1016/j.](https://doi.org/10.1016/j.eclinm.2022.101673)

[eclinm.2022.101673](https://doi.org/10.1016/j.eclinm.2022.101673)

Research in context

Evidence before this study

A search of PubMed, Embase, and Scopus detected only one study comparing GLOBOCAN and GBD cancer data directly, together with data from a national cancer registry: this concerned annual percentage changes in the incidence of lung cancer. Inclusion criterion used for our search is study/studies which directly compared GLOBOCAN and GBD databases with no time and language limitation. Search terms 'GLOBOCAN' AND 'GBD' were used.

Added value of this study

We have exposed significant discrepancies in age-standardised incidence rates of oral cancer between GLOBOCAN 2020 and GBD 2019, especially in countries without high-quality national or regional population-based cancer registries. Such discrepancies should be explored for all cancer types and sites.

Implications of all the available evidence

We recommend that authors always provide, and journal editors require, statements about the quality of cancer data quoted. This should include statements on timeliness, population coverage, and accuracy of anatomical site and subsite descriptors. We recommend development of a scoring system for data quality which can, with subsequent evaluation studies, be developed for use around the world.

Introduction

Summary and analytical data from the major collations of national data imply that, whilst the incidence of oral cancer remains high in many low- and middle-income countries (LMICs), with wide variation, a declining global trend is underway. Cancer epidemiological studies play a major role in formulating governmental health policies.¹ In the case of oral cancer, as with most malignancies, identification of risk factors and their management at population level, policies on cancer screening, resource allocations for cancer care, and barriers to disease prevention thereto all require accurate data.²

Global Cancer Observatory (GLOBOCAN) 2020 and Global Burden of Disease (GBD) 2019 are the two most established global online cancer databases,^{3,4} provided by the International Agency for Research on Cancer's (IARC) Global Cancer Observatory and the Institute for Health Metrics and Evaluation (IHME), respectively. Both online platforms allow fast and convenient access for the world's researchers and policymakers to interpret the incidence or mortality for most types of cancers in their countries. Estimates of trends over time are also given.

A recent review of global trends in the incidence of lip, oral cavity and pharyngeal cancers by Du et al.,⁵ utilising data from GBD, alerted us to significant differences in data from those in GLOBOCAN and prompted this study. The sources of data in GLOBOCAN 2020 are specific to individual countries. IARC describe seven distinct data sources ranging from code 1 "national (or local with coverage greater than 50%) rates projected to 2020" to code 9 "no data: the rates are those of neighbouring countries or registries in the same area". This information is available on the IARC Cancer Today website (<https://gco.iarc.fr/today/data-sources-methods>) with the dataset "Cancer incidence and mortality data: sources and methods by country".

Sources of GBD 2019 data are difficult to divine through their website. However, as of GBD 2017, it is claimed that all data are collected via "more than 90,000 data sources, and each of those sources in turn had anywhere from hundreds to billions of individual data points".⁶ Sources of data on cancer incidence given for individual countries are not described on the website (<https://www.healthdata.org/acting-data/what-data-sources-go-gbd>). "Unusual data sources" (e.g., verbal autopsy) were used in certain situations, "especially in areas of the world that lack consistent and complete censuses and vital registration systems".⁶

Our particular interest are the age-standardised incidence rates (ASIR) of lip and oral cavity cancers (LOC), and their anatomical subsites. Discrepancies between data provided by GLOBOCAN 2020 and GBD 2019 cause confusion and can affect decision making. To our knowledge there is no published examination of these differences. There is also a need to critically interpret the quality of national data, as imported by these global bodies. Therefore, it is important to examine the differences between the two platforms, to attempt to explain these differences, and to appraise the quality of the data. We use LOC as an example of similar differences across other cancer types and anatomical sites.

Our aims were to identify discrepancies in the incidence data for LOC between GLOBOCAN 2020 and GBD 2019, an example of a cancer with substantial discrepancies. Our objectives were to: 1) identify and scrutinise online data from the five countries with the greatest and the five with the least differences in ASIR (per 100,000 population per annum) of LOC between GLOBOCAN 2020 and GBD 2019, 2) compare the data for these countries as presented by GLOBOCAN 2020 with data in the refereed literature which were sufficiently comprehensive as to provide an independent comparator.

Methods

"Cancer Today" was first selected on the website of Global Cancer Observatory (<https://gco.iarc.fr/>). Options were then selected in the order of "compare: populations, indicator: incidence, sex: both, age groups:

0-85+, population: countries worldwide, cancer sites by ICD-10: lip, oral cavity” to obtain the table of “estimated number of new cases in 2020, lip, oral cavity, both sexes, all ages”. The columns of “ASR (World Standard Population)” and “Number” in 2020 were selected.

To obtain the ASIR by using the GBD World Standard Population of LOC in all countries in 2019, GBD results tool (<https://ghdx.healthdata.org/gbd-results-tool>) was first selected on the website, then options were inputted for the Global Health Data Exchange (GHDx) query tool: “base: single, location: select only countries and territories, year: 2019, context: cause, age: age-standardised, metric: rate, measure: incidence, sex: both, cause: B1.1 Lip and oral cavity cancer”. Similarly, number of incident cases were obtained via the same options except “metric: number, age: all ages”.

Two additional inclusion criteria were applied to identify the ten selected countries: 1) number of incident cases were greater than 1000 per year in GLOBOCAN 2020, 2) countries with higher number of incident cases were prioritised over countries with fewer incident cases. Based on the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), GLOBOCAN 2020 defines LOC with the codes of C00-C06, while GBD 2019 includes C00-C08.

By comparing the ASIR (per 100,000 population per annum) of LOC from GLOBOCAN 2020 and GBD 2019, ten countries with the greatest and smallest fold difference (five countries each) were identified. A systematic search of PubMed and Embase electronic databases was then performed to identify publications reporting the incidence of LOC in the selected countries. Specifically, data sources of the articles were examined and evaluated. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagrams as stipulated by the PRISMA 2020 statement were applied, one for each country, to assist the search. Search terms for PubMed were: (mouth cancer[MeSH Terms]) OR (oral cancer[MeSH Terms]) OR (lip cancer [MeSH Terms]) OR (tongue cancer[MeSH Terms]) AND (incidence[MeSH Terms]) AND (country names** [MeSH Terms]), while search terms for Embase were: 'mouth cancer'/exp** OR 'oral cancer'/exp OR 'lip cancer [MeSH Terms]) OR (oral cancer[MeSH Terms]) OR (lip cancer [MeSH Terms]) OR (tongue cancer[MeSH Terms]) OR 'tongue cancer'/exp OR 'incidence'/exp AND 'country names*/exp. Results from PubMed and Embase were screened by two independent authors with an inter-rater reliability of 0.93. Differences were then discussed with the third author and resolved.

** Exp stands for “Emtree term – exploded”.

* Country names included: Papua New Guinea, Vietnam, Pakistan, China, Indonesia, United States of America, Brazil, France, Germany, India.

The inclusion criteria for the publications selected were:

- 1) Studies published in or after 2015
- 2) ASIR or other forms of standardised incidence rates were discussed
- 3) LOC were the focus of the study
- 4) Original studies or articles.

The following types of studies were excluded:

- 1) Not published in English
- 2) Unrelated to national LOC incidence in the targeted countries
- 3) Simply using GLOBOCAN or GBD as their source of data
- 4) Source of data from previous papers
- 5) Combining LOC (ICD-10 C00-C06) with other subsites
- 6) Using a single source of data (e.g., only one population-based or hospital-based cancer registry)
- 7) Not related to humans
- 8) Considering only male or female
- 9) Considering specific ethnic groups (e.g., Asian, American Indian, African American)
- 10) Performed outside the targeted country
- 11) Poorly described or unexplained source of data
- 12) Full-text unavailable

Role of the funding source

This research received no specific grant or funding from any funding agency in the public, commercial, or not-for-profit sectors, and the authors received no financial support for the research, authorship, and/or publication of this article.

All authors have access to the dataset and decided to submit for publication.

Results

Figure 1 demonstrates the differences of ASIR (per 100,000 population per annum) in LOC between the GLOBOCAN 2020 and GBD 2019 data in the 38 countries with over 1,000 estimated incident cases. 16 countries had the ASIR (per 100,000 population per annum) of GLOBOCAN 2020 greater than that of GBD 2019, while the remaining countries had the opposite trend except for the Netherlands which showed no difference between the two databases.

Tables 1 and 2 show the top five countries with the greatest and smallest differences between these databases. Greatest differences were found in Papua New Guinea (PNG), Vietnam, China, Pakistan, and

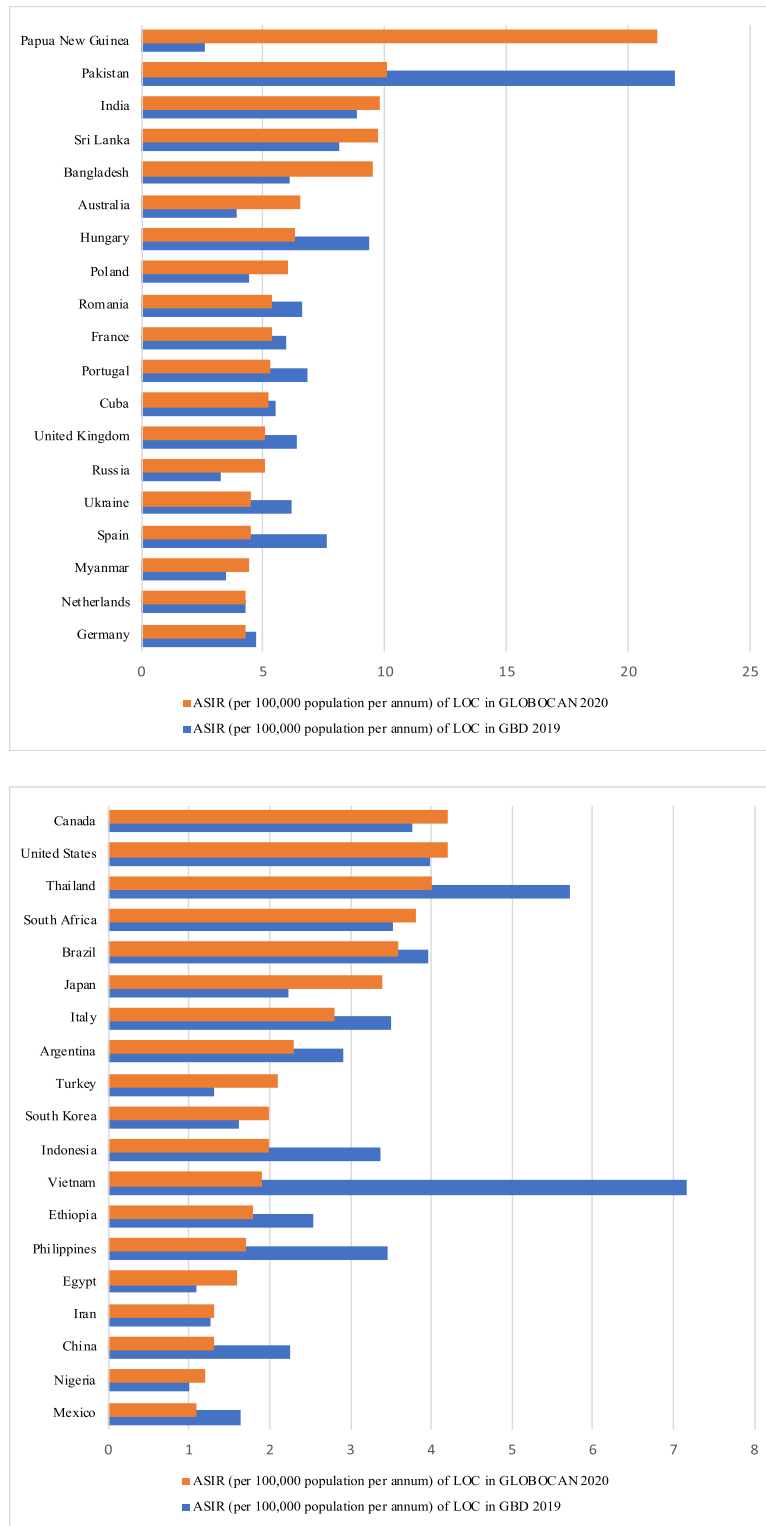


Figure 1. Age-standardised incidence rates (ASIR) per 100,000 population per annum of lip and oral cavity cancers (LOC) from GLOBOCAN 2020 and GBD 2019 in countries with over 1000 estimated incident cases in GLOBOCAN 2020.

Countries (Group A)	ASIR per 100,000 population per annum of LOC in GLOBOCAN 2020 (estimated number of incident cases in 2020)	ASIR per 100,000 population per annum of LOC in GBD 2019 (estimated number of incident cases in 2019)	Fold difference (times) ^a
Papua New Guinea	21.2 (1238)	2.62 (142)	+ 8.09
Vietnam	1.9 (2152)	7.16 (7118)	- 3.77
Pakistan	10.1 (16,959)	21.93 (28,579)	- 2.17
China	1.3 (30,117)	2.25 (45,216)	- 1.73
Indonesia	2.0 (5780)	3.38 (7306)	- 1.69

Table 1: Top five countries with the greatest difference in age-standardised incidence rates (ASIR) per 100,000 population per annum of lip and oral cavity cancers (LOC) between GLOBOCAN 2020 and GBD 2019

^a Positive fold difference indicate ASIR of LOC in GLOBOCAN 2020 is greater than that in GBD 2019. Negative fold difference indicate ASIR of LOC in GLOBOCAN 2020 is smaller than that in GBD 2019.

Indonesia. These countries are collectively known as group A in the following discussion. In contrast, the United States of America (USA), Brazil, France, Germany, and India had the smallest differences between the two databases. These countries are described as group B in later sections of this paper. It is noteworthy that seven of the ten countries we have selected on this basis had higher ASIR (per 100,000 population per annum) of LOC in GLOBOCAN 2020 than that in GBD 2019.

Tables 3 and 4 describe the data collection method used by GLOBOCAN 2020 and the methods used in publications which met our inclusion criteria. For group A and India, no study which met our criteria could be identified. PRISMA diagrams for each country are available as supplementary figures.

In GLOBOCAN 2020, if ASIR (per 100,000 population per annum) of LOC could not be obtained directly via national or local data, the Agency uses “estimation from national mortality estimates by modelling using mortality-to-incidence ratios”, “estimation from neighbouring countries partitioned using frequency data”, or “rates from neighbouring countries or registries in the same area”. In contrast, data from the USA, Brazil, France, and Germany had similar, if not totally

comprehensive, means of collecting data from high quality population-based registries.

Discussion

GLOBOCAN 2020 and GBD 2019 allow users to estimate the incidence of LOC in almost all countries in the world easily. With their interactive platforms, researchers can obtain their desired epidemiological data directly from the websites and use these data in their research, policy development and publications. There are many such publications but, unfortunately, the quality and accuracy of the data used are rarely discussed.

In terms of LOC subsites, it is unfortunate that GBD 2019 includes malignant neoplasms of parotid gland (Co7) and of “other and unspecified major salivary glands” (Co8) as LOC because Co7 and Co8, which are outside the mouth, have distinctly different aetiology and behaviours from oral cavity cancers. In numerical terms this would be expected to inflate rates presented by GBD: the opposite, however, is found.

It is not surprising that when GLOBOCAN and GBD could not obtain high-quality or accessible LOC data from national or local cancer registries, as in group A, discrepancies would be seen between the two online

Countries (Group B)	ASIR per 100,000 population per annum of LOC in GLOBOCAN 2020 (estimated number of incident cases in 2020)	ASIR per 100,000 population per annum of LOC in GBD 2019 (estimated number of incident cases in 2019)	Fold difference (times)
The United States of America	4.2 (24,470)	3.99 (30,768)	+ 1.05
Brazil	3.6 (9839)	3.97 (9583)	- 1.10
France	5.4 (6577)	5.95 (6740)	- 1.10
Germany	4.3 (7333)	4.72 (7516)	- 1.10
India	9.8 (135,929)	8.82 (104,838)	+ 1.11

Table 2: Top five countries with the smallest difference in age-standardised incidence rates (ASIR) per 100,000 population per annum of lip and oral cavity cancers (LOC) between GLOBOCAN 2020 and GBD 2019.

*Positive fold difference indicate ASIR of LOC in GLOBOCAN 2020 is greater than that in GBD 2019. Negative fold difference indicate ASIR of LOC in GLOBOCAN 2020 is smaller than that in GBD 2019.

Countries (Group A)	Source category & method of GLOBOCAN 2020	Data source(s) for selected published studies
Papua New Guinea	Category: local & 9 ("rates were estimated as an average of those from selected neighbouring countries") "No data available, thus weighted average (80%, 20% census 2000) of average of Fiji, Vanuatu and New Caledonia rates for the Islands applied to 2020 population"	None identified
Vietnam	Category: local & 2b ("the most recently observed regional incidence rates were applied to the 2020 population") "Weighted average (64%, 36% respectively) of rural rates: using national mortality data (2005–2006) for rural provinces converted to incidence using M:I ratios from Chinese cancer registries; and urban rates obtained from Ho Chi Minh City (2011–2014) and Hanoi City (2007–2009) applied to 2020 national population"	None identified
Pakistan	Category: local & 2b ("the most recently observed regional incidence rates were applied to the 2020 population") "Average of Pakistan, South Karachi (1998-2002, CI5 Vol. IX) and Punjab (2010–2012), and India, Ahmedabad urban (2012–2013) and Patiala district (2012–2014) rates applied to 2020"	None identified
China	Category: local & 2b ("the most recently observed regional incidence rates were applied to the 2020 population") "Rates (2010–2012) from 92 cancer registries in rural and urban settings were applied to 2020 Chinese population. The combined rates were computed as 60% urban rates and 40% rural rates"	None identified
Indonesia	Category: national & 4 ("age- and sex-specific national incidence rates for all cancers combined were obtained by averaging overall rates from neighbouring countries. These rates were then partitioned to obtain the national incidence for specific sites using available cancer-specific relative frequency data") "Average of Malaysia, Penang (2008-2010) and Brunei Darussalam (2010–2012), and 'all sites' rates with frequencies from Indonesia Cancer Registry (2008–2012)"	None identified

Table 3: Comparisons of data collection method of GLOBOCAN 2020 and included literature in Papua New Guinea, Vietnam, Pakistan, China, and Indonesia.

databases. Where only minor differences were seen between GLOBOCAN and GBD, as in group B, presumptively due to those countries having well-established cancer registries and healthcare administrative systems, the literature is more consistent.⁷

Data quality of any truly national cancer registry plays a significant role in the accuracy of what is held in the two online databases. A high-quality population-based cancer registry (PBCR) will collect, validate, and store all cancer data accurately in a timely manner. It will then analyse and interpret those data regularly and report incidence, mortality, disease trends and other epidemiological data, for example, associations with trends in tobacco and alcohol consumption. This facilitates the implementation of health policies for improved cancer control.⁸

There are pertinent examples in the literature for the ten countries which form the basis of the present study. In PNG, cancer registration remains paper-based at Port Moresby General Hospital making retrieval of any meaningful data difficult, and the completeness of national coverage is uncertain.⁹ When about 87% of the population live in rural and remote areas, many cases of

oral cancer would remain undetected, undiagnosed, and hence be unrecorded.

Hanoi and Ho Chi Minh City registries, which covered 20% of the Vietnamese population, were reported as the best cancer registries in the country in terms of their overall practices and data quality by the Global Initiative for Cancer Registry Development Partners Task Force in 2017.¹⁰ In contrast, another study found that data quality was inconsistent among the nine cancer registries in Vietnam, and estimations of cancer incidence by the national cancer registry were regarded as of low quality.¹¹

With the absence of a national cancer registry, Punjab and Karachi South Cancer registries are the only two effective registries in Pakistan. As a country with a large ethnic diversity and heterogeneous population structure, significant regional variations of cancer data are observed.¹² This 2018 publication gives a meta-analysis of all previously published papers describing the prevalence of oral, gastric, prostate, breast, and colorectal cancers in Pakistan. High heterogeneity was observed among the included studies, which was mainly affected by the location from which the data were derived. Due

Countries (Group B)	Source category & method of GLOBOCAN 2020	Data source(s) for selected published studies
The United States of America	Category: national & 1 ("observed national incidence rates were projected to 2020") "National rates with coverage greater than 50% (2003-2012) projected to 2020 and applied to 2020 population"	Cumulative ASR of "cancers in oral cavity region" (ICD-O-3 C000-C009, C020-023, C028-029, C030-C039, C040-049, C050, C058-C059, C060-C069) was 4.3/100,000 from 2000-2010. In other words, the definition of oral cavity region includes "lip, oral tongue, floor of mouth, and gums/hard palate and other sites". Data were obtained via SEER-18 that covered 18 cancer registries from 18 geographical regions namely "San Francisco/Oakland (SF), Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose/Monterey (SJM), Los Angeles (LA), Alaska Natives, Rural Georgia, California excluding SF/SJM/LA, Kentucky, Louisiana, New Jersey, and Greater Georgia", rendering 28% of the US population. Authors also mentioned "SEER registries actively follow-up with and receive cancer-related data from local hospitals, physicians, and laboratories on individuals diagnosed with cancer, who are residents of the geographical area covered by the SEER registry at the time of diagnosis". ²⁵ ASR of oral cavity cancer (ICD-O-3 C020-C023, C028-C029, C030-C031, C039, C040-C041, C048-C049, C050, C058-C059, C060-C061, C062, C068-C069) was approximately 2.1/100,000 in 2014 according to the figure. Data were obtained via SEER 9 database. However, "patients were excluded if they were diagnosed based on their death certificate or autopsy or if there were no data on survival time". ²⁶
Brazil	Category: local & 3a ("rates were estimated from national mortality data by modelling, using mortality-to-incidence ratios derived from cancer registries in that country") "Rates estimated from mortality using M:I ratios from 18 local cancer registries"	This study discussed the ASR of tongue cancer (ICD-10 C01-C02) from 2000 to 2012. "Data were obtained from 23 population-based cancer registries which covered around a quarter of the Brazilian population". According to the figure, ASR of tongue cancer was approximately 1.63/100,000 in 2012. ²⁷
France	Category: local & 3a ("rates were estimated from national mortality data by modelling, using mortality-to-incidence ratios derived from cancer registries in that country") "Rates estimated from mortality using M:I ratios from 15 local cancer registries"	ASR of "oral tongue cancer" (OTC) (ICD-10 C02) and "cancers of the remaining subsites of the oral cavity" (ROC) (ICD-10 C03-C06) from 1975 to 2009 were provided by French Network Cancer Registries, the FRANCIM network. These registries covered 21% of the French population in 19 French districts. ASR of the OTC and ROC were 1.35 and 2.6/100,000 person-years in 2018, respectively. ²⁸
Germany	Category: local & 1 ("observed national incidence rates were projected to 2020") "Rates (2003-2012) from eight cancer registries projected to 2020 and applied to the 2020 national population"	ASR of lip cancer (ICD-10 C00) were 0.35/100,000 in 2012. These data were obtained "from database of nine population-based cancer registries, covering a population of 39 million inhabitants in 14 federal states". ²⁹ ASR by European standard population of oral cancer (ICD-10 C00-C06) were analysed. Data were provided by the Association of Population-based Cancer Registries in Germany, which were "based on data from ten population-based cancer registries, which cover a population of 37.9 million people (46% of the German population)". However, exact numbers of ASR from 2003 to 2011 were not provided in the table but only available in the figure. ³⁰

Table 4 (Continued)

Countries (Group B)	Source category & method of GLOBOCAN 2020	Data source(s) for selected published studies
India	<p>Category: local & 2b (“the most recently observed regional incidence rates were applied to the 2020 population”)</p> <p>“National incidence was estimated as follows:</p> <ol style="list-style-type: none"> 1. Incidence rates for the North-Eastern region (Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura states) were computed using population weighted average of the rates from nine population-based cancer registries (29% coverage around 2011). These rates were applied to the corresponding 2020 population. 2. Remaining urban population: rates from seven cancer registries (Bangalore, Bhopal, Chennai, Mumbai, Delhi, Nagpur, Pune) (15% coverage of urban population) for the period 2003–2012 were projected to 2020 and applied to the 2020 urban population. These seven registries were supplemented by three cancer registries (Aurangabad, Kollam and Trivandrum) when projection cannot be performed (use of the most recent rates) 3. Remaining rural population: rates from six cancer registries (Ahmedabad, Ambalikkai, Barshi, Mansa, Sangrur and Wardha) (1% coverage of rural population) were applied to the 2020 rural population. <p>The overall incidence estimate for India for 2020 is the sum of the three estimates”</p>	None identified

Table 4: Comparisons of data collection method of GLOBOCAN 2020 and included literature in the United States of America, Brazil, France, Germany, and India.

to numerous political, economic, and military factors, the authors did not anticipate improvement in the quality of Pakistani cancer data in the short term.¹² We expect that the incidence of oral cancer in Pakistan would experience similar regional variations.

By 2019, China had a well-established cancer registration system with 574 PBCRs covering 31.5% of the population.¹³ Although the data quality in the Western and rural areas is generally inferior to that in Eastern and urban areas, with the political support and funding boost by the government and transition to a web-based national cancer registration system, this disparity of cancer data quality is narrowing. The Chinese government has also established “national criteria for Chinese cancer registration data” so only high-quality data would be included in the China Cancer Registry Annual Reports.¹³

Indonesia is another country without a national cancer registry, and yet a “sustainable cancer registry” was reputed to have been in existence in 2007, progressively shifting from hospital-based to population-based.¹⁴ The large population of Indonesia spread across its more than 6000 inhabited islands has over 1300 ethnic groups across its 30 Provinces. In their 2012 publication these authors opined that more collaborative efforts were required to improve coverage and management of the registry.¹⁴ No subsequent studies which evaluate the quality of the registry have been identified to date.

For the USA, the Surveillance, Epidemiology, and End Results (SEER) Program supported by the National Cancer Institute is the main source of incidence and mortality data and covers 48% of the population. The National Program of Cancer Registries of the North American Association of Central Cancer Registries and the SEER program together effectively cover the entire US population.¹⁵ The SEER registries in particular demonstrate at least 95% completeness of their cancer data annually with less than 3% of cases identified through death certificate only.¹⁶ The SEER program is also the “golden standard for data quality amongst cancer registries” globally owing to their continuous training and support to regional registrars, strict data quality control, regular monitoring, and evaluation. This leads to over 40,000 studies citing data from SEER between 1973 and 2015.¹⁵

Data quality of the 20 PBCRs in Brazil is discussed in a very limited number of studies. One such has evaluated improvements in data quality and completeness of the PBCR in São Paulo. This examined several record linkage techniques but did not elaborate on the overall quality of PBCRs in the country.¹⁷ According to the report Cancer in Brazil Volume IV, data quality from Brazilian PBCRs was described as “improving progressively and positively” since 2003. Nonetheless, proportion of cases with microscopic verification, a “golden standard” for cancer diagnosis and an important data

quality indicator, still has notable geographical variation, including the data pertaining to oral cancer.¹⁸

The European Network of Cancer Registries consists of 200 PBCRs, covering about 60% of the European population. Data quality varies across Europe in terms of completeness and validity. Limitation of resources, of funding, and absence of political support are not uncommon in some regional registries resulting in incomplete outputs, sometimes with low accuracy.¹⁹ High-quality national and regional incidence data covering between 10% and 50% of the population are available in Germany and France,²⁰ but comprehensive studies exploring the data quality of all European countries are few.

36 PBCRs are currently registered with the National Cancer Registry Programme under the National Centre for Disease Informatics and Research of the Indian Council of Medical Research.²¹ Low population coverage, however, remains a challenge. In contradiction to the data collection method of GLOBOCAN 2020, which claimed population coverage of 15–29%, two studies have found that India's PBCRs only cover 10–15% of the population: in particular, rural coverage is merely 0.1% when 68% of the Indian population reside in rural areas.²² ICD-10 coding is not practised in all Indian hospitals so many diagnoses would be subject to the vagaries of hospital records and death certificates. Many medical records from Indian hospitals are of low quality, with substandard quality indicators including lack of microscopic verification. PBCRs are also unevenly distributed across India so fail to capture information on differences related to ethnicity.²² Whilst GLOBOCAN 2020 ranks lip and oral cavity the most common cancer among men in India, and second for both sexes combined, most national reports combine oral cavity and pharynx 'unspecified',²¹ a significant limitation when HPV-driven cancer of the oropharynx and EBV-driven cancer of the nasopharynx are such different diseases to the smokeless tobacco, areca nut, and/or betel quid driven cancers of the oral cavity.²³

The Cancer Incidence in Five Continents (CI5) series are a collaboration between the International Agency for Research on Cancer (IARC) and the International Association of Cancer Registries, published every five years. The latest, Volume XI, was published in 2017. Cancer registries from each country are assessed for their comparability, completeness, and validity, and high-quality incidence data are published from selected PBCRs.²⁴ No data from PNG or from Indonesia are provided in CI5 Volume XI. Several other changes relevant to the present work include: The Hanoi Cancer Registry (CR) is no longer included since CI5 Volume IX. South Karachi CR, the only Pakistani Registry previously included, has not been included since Volume X. 25 new Chinese PBCRs have been added. The USA was one of the countries with a large number of included PBCRs in CI5 and

this number continues to rise. Four new PBCRs in Brazil were included in Volume XI but four PBCRs from Volume X were excluded at the same time. Both Germany and France had a stable number of selected PBCRs with the number rising steadily. Five new Indian PBCRs were added to Volume XI.²⁴ Overall, these changes are encouraging. The Call for data documents for Volume XII closed on 31 May 2022 (http://www.iacr.com.fr/index.php?option=com_content&view=article&id=90&Itemid=566). This updated Volume XII is expected to include more PBCRs from countries, regions, and continents.

Lastly, from our systematic searches, many studies have grouped lip and oral cavity with pharynx and categorised outputs as "oral and oropharyngeal cancer" or "oral cavity and pharynx cancer". These categorisations lacked subsite accuracy and failed to realise that oral cancer and oropharyngeal cancer have completely different etiological factors, pathogeneses, prognosis, and treatment outcomes.²³ Oral cavity cancer remains largely tobacco, areca nut, and alcohol-driven, while oropharyngeal cancer has largely become human papillomavirus-driven in recent years. Lip cancer has an important component of ultraviolet light in its pathogenesis.²³ Future studies must consider LOC and oropharyngeal cancer, and their anatomical subsites, separately. ICD-10 codes remain an important tool and should be used when defining LOC.

The quality and coverage of cancer registration show steady improvements in many well-resourced nations but are deteriorating elsewhere – particularly in LMICs. Armed conflicts, huge internally and externally displaced populations and tensions over water and food security and climate change do not auger well for improvements in the near term. Population growth in high incidence parts of the world, especially Africa and South Asia, will compound efforts at better cancer care.

We applaud and encourage the work of IARC and its Global Initiative for Cancer Registry Development whilst recognising the challenges in obtaining "buy-in" from many LMICs, several of which have exceptionally high rates of LOC. We recommend that a working party of IARC, GBD and of the International Association of Cancer Registries convene and make recommendations as to how this should be structured, approached, and evaluated. We also encourage countries' ownership and participation of international or supranational bodies producing estimates on their behalf. All future presenters and users of data from either GLOBOCAN or GBD are recommended to state a data quality score given by the creators of the database(s) and/or state their own judgement of the reliability and accuracy involved: frequently such quotations need to state that the numbers are estimates only. Further studies are needed to explore discrepancies between the two databases for all cancer types and sites.

A limitation of this study is the one-year difference between presentation by the two platforms: GLOBOCAN 2020 and GBD 2019. However, both institutions collate reports and estimates from a wide range of sources, many of which are several years older. Estimates are updated to take account of population growth and breadth of coverage and the algorithms used are published. Since our focus was mainly on the methods of data collection and interrogation, we submit that this is a minor matter. Restriction of language to English in our systematic search may have introduced selection bias. Nevertheless, both databases are presented in English, so it is publications in the English language which are of greatest value.

Contributors

All authors made substantial contributions to the conception or design of the work and reached agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ka Ming (Henry) Fan: acquisition, analysis, and interpretation of data for the work; Drafting the work and revising it critically for important intellectual content; Final approval of the version to be published

Jyotsna Rimal: analysis and interpretation of data for the work; Revising it critically for important intellectual content; Final approval of the version to be published

Ping Zhang: analysis of data for the work; Revising it critically for important intellectual content; Final approval of the version to be published

Newell Johnson: acquisition, analysis, and interpretation of data for the work; Drafting the work and revising it critically for important intellectual content; Final approval of the version to be published.

Data sharing statement

Data of this study are openly available in the online databases of GLOBOCAN 2020 and GBD 2019 as described in Methods.

Declaration of interests

The authors declare that they had no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research received no specific grant or funding from any funding agency in the public, commercial, or not-for-profit sectors, and the authors received no financial support for the research, authorship, and/or publication of this article.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2022.101673.

References

- Miranda-Filho A, Bray F. Global patterns and trends in cancers of the lip, tongue and mouth. *Oral Oncol.* 2020;102:104551.
- Toporcov TN, Wunsch Filho V. Epidemiological science and cancer control. *Clinics.* 2018;73(suppl. 1):e627s.
- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249.
- Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* 2020;396(10258):1204–1222.
- Du M, Nair R, Jamieson L, Liu Z, Bi P. Incidence trends of lip, oral cavity, and pharyngeal cancers: global burden of disease 1990–2017. *J Dent Res.* 2020;99(2):143–151.
- Redford SB, Alexander L. What data sources go into the GBD? Institute for Health Metrics and Evaluation: Institute for Health Metrics and Evaluation. Available from: <https://www.healthdata.org/acting-data/what-data-sources-go-gbd>. Accessed 30 May 2022.
- Antoni S, Soerjomataram I, Møller B, Bray F, Ferlay J. An assessment of GLOBOCAN methods for deriving national estimates of cancer incidence. *Bull World Health Organ.* 2016;94(3):174–184.
- Wormald JS, Oberai T, Branford-White H, Johnson LJ. Design and establishment of a cancer registry: a literature review. *ANZ J Surg.* 2020;90(7-8):1277–1282.
- Kelwaip RA, Fose S, Siddiqui MS, et al. Oral cancer in Papua New Guinea: looking back and looking forward. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2020;130(3):292–297.
- Nguyen SM, Deppen S, Nguyen GH, Pham DX, Bui TD, Tran TV. Projecting cancer incidence for 2025 in the 2 largest populated cities in Vietnam. *Cancer control.* 2019;26(1):1073274819865274.
- Pham T, Bui L, Kim G, Hoang D, Tran T, Hoang M. Cancers in Vietnam—Burden and control efforts: a narrative scoping review. *Cancer Control.* 2019;26(1):1073274819863802.
- Idrees R, Fatima S, Abdul-Ghafar J, Raheem A, Ahmad Z. Cancer prevalence in Pakistan: meta-analysis of various published studies to determine variation in cancer figures resulting from marked population heterogeneity in different parts of the country. *World J Surg Oncol.* 2018;16(1):129.
- Wei W, Zeng H, Zheng R, et al. Cancer registration in China and its role in cancer prevention and control. *Lancet Oncol.* 2020;21(7):e342–e349.
- Wahidin M, Noviani R, Hermawan S, Andriani V, Ardian A, Djarir H. Population-based cancer registration in Indonesia. *Asian Pac J Cancer Prev.* 2012;13(4):1709–1710.
- Duggan MA, Anderson WF, Altekruse S, Penberthy L, Sherman ME. The surveillance, epidemiology, and end results (SEER) program and pathology: toward strengthening the critical relationship. *Am J Surg Pathol.* 2016;40(12):e94–e102.
- Brown LM, Check DP, Devesa SS. Oral cavity and pharynx cancer incidence trends by subsite in the United States: changing gender patterns. *J Oncol.* 2012;2012:649498.
- Peres SV, Latorre MD, Tanaka LF, et al. Quality and completeness improvement of the Population-based Cancer Registry of São Paulo: linkage technique use. *Rev Bras Epidemiol.* 2016;19(4):753–765.
- National Cancer Institute José Alencar Gomes da Silva (INCA). *Cancer in Brazil: Data from the Population-Based Registries, volume 4.* Brazilian National Cancer Institute José Alencar Gomes da Silva (INCA); 2013. Accessed 30 May 2022.
- Forsea A-M. Cancer registries in Europe-going forward is the only option. *Ecancermedicalscience.* 2016;10:641.
- Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015;136(5):e359–e386.
- Mathur P, Sathishkumar K, Chaturvedi M, et al. Cancer statistics, 2020: report from National Cancer Registry Programme. *India. JCO Glob Oncol.* 2020;(6):1063–1075.

- 22 Behera P, Patro BK. Population based cancer registry of india – the challenges and opportunities. *Asian Pac J Cancer Prev*. 2018;19(10):2885–2889.
- 23 Shah JP, Johnson NW. *Oral and oropharyngeal cancer*. Second ed Boca Raton, FL: CRC Press/Taylor and Francis; 2019.
- 24 Bray F, Colombet M, Mery L, et al. *Cancer Incidence in Five Continents, Vol. XI (electronic version)*. Lyon: International Agency for Research on Cancer; Lyon: International Agency for Research on Cancer; 2017. Accessed 30 May 2022.
- 25 Weatherspoon DJ, Chattopadhyay A, Boroumand S, Garcia I. Oral cavity and oropharyngeal cancer incidence trends and disparities in the United States: 2000–2010. *Cancer Epidemiol*. 2015;39(4):497–504.
- 26 Kim Y-J, Kim JH. Increasing incidence and improving survival of oral tongue squamous cell carcinoma. *Sci Rep*. 2020;10(1):7877.
- 27 Cohen Goldemberg D, de Araújo LHL, Antunes HS, de Melo AC, Santos Thuler LC. Tongue cancer epidemiology in Brazil: incidence, morbidity and mortality. *Head Neck*. 2018;40(8):1834–1844.
- 28 Deneuve S, Pérol O, Dantony E, et al. Diverging incidence trends of oral tongue cancer compared to other head and neck cancers in young adults in France. *Int J Cancer*. 2022;150(8):1301–1309.
- 29 Singer S, Zeissig SR, Emrich K, Holleczeck B, Kraywinkel K, Ressing M. Incidence of lip malignancies in Germany-data from nine population-based cancer registries. *J Oral Pathol Med*. 2017;46(9):780–785.
- 30 Hertrampf K, Eisemann N, Wiltfang J, Pritzkeleit R, Wenz HJ, Waldmann A. Baseline data of oral and pharyngeal cancer before introducing an oral cancer prevention campaign in Germany. *J Craniomaxillofac Surg*. 2015;43(3):360–366.