



 Cancer RADAR

 [cancer.radar@iarc.who.int](mailto:cancer.radar@iarc.who.int)

# Cancer RADAR: assess the current risk and preventable burden of cancer among individuals with a migration background across Europe

version September 2025



## Contents

1 Involved partner .....	2
2 Short summaries .....	3
3.1 Introduction.....	4
3.2 Objectives.....	5
3.3 Work plan and tasks .....	6
3.4 Resources .....	9
3.5 Confidentiality, security and ethical approval .....	9
3.6 Contribution, publication policy and outcome.....	9
4.1 Methods .....	11
4.2 Definitions .....	11
4.3 Data collection process .....	12
4.4 Step 1: Data collection.....	14
4.5 Step 2: Enter data in input file.....	14
4.6 Step 3: Data processing .....	16
4.7 Step 4: Data sharing .....	18
4.8 Data quality .....	18
5 Timeline.....	19
Appendix 1 – Input data .....	20
Appendix 2 – Effect measures .....	22
Literature.....	25

Cancer RADAR: assess the current risk and preventable burden of cancer among individuals with a migration background across Europe

1 Involved partner

*Partners are randomly ordered.*

#### **Advisory board**

**Dr. C.J. Alberts** – Amsterdam UMC, Amsterdam, the Netherlands and Public Health Service Amsterdam, Amsterdam, Netherlands

**D. Georges** – International Agency for Research on Cancer (IARC/WHO), Lyon, France

**Dr. S. Rosso** – University of Turin - Piedmont Cancer Registry

**Dr. I. Man** – International Agency for Research on Cancer (IARC/WHO), Lyon, France

**Dr. F. Bray** – International Agency for Research on Cancer (IARC/WHO), Lyon, France

**Dr. I. Baussano** – International Agency for Research on Cancer (IARC/WHO), Lyon, France

#### **Working group**

**Prof. Dr. L. Richiardi** – University of Turin - Piedmont Cancer Registry

**Dr. E. Migliore** – Piedmont Cancer Registry

**Dr. J. Galceran** – Cancer registry of Tarragona, Cancer Epidemiology and Prevention Service, Hospital University Sant Joan de Reus, IISPV, Reus, Spain

**Dr. M. Carulla** – Cancer registry of Tarragona, Cancer Epidemiology and Prevention Service, Hospital University Sant Joan de Reus, IISPV, Reus, Spain

**Dr. A Bordoni** – Cancer Registry of canton Ticino, Switzerland

**Dr. L. Ortelli** – Cancer Registry of canton Ticino, Switzerland

**Dr. M.D. Chirlaque López** – Cancer registry of Murcia, Spain

**Dr. K. Van Herck** – Belgian Cancer Registry, Brussels, Belgium

**Dr. F. Verdoort** – Belgian Cancer Registry, Brussels, Belgium

**Dr. O. Visser** – Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands

*Experts of specific cancers will be involved.*

Date	Objective of changes
Version September 2025	- CI5-XIII (2018-2022) period has been added as an additional period for which data can be collected - R-package when aggregating takes now into account both numerator and denominator data -'screening-detectable' has been changed into 'amenable to screening detection'
Version October 2024	Initial version of protocol

## 2 Short summaries

Migration to and within Europe has increased and diversified in recent years. Individuals with a migration background often have health needs that differ from the general (host) population of a country. At the same time, they can face important barriers to access the right health care. A unified quantification of current disparities of cancer by migration background in Europe does not exist impeding policy makers to act upon the health needs of migrants. To fill this knowledge gap Cancer RADAR's aim is to develop an infrastructure that allows quantifying the risk of cancer by migration background across Europe. To establish this, Cancer RADAR will first focus on infection-related cancers and cancers amenable to screening, breast, colorectal, cervical, stomach, liver and lung cancer.

### 3.1 Introduction

Migration to and within Europe has increased and diversified in recent years.<sup>1</sup> Individuals with a migration background often have health needs that differ from the general (host) population of a country.<sup>2-4</sup> In addition, they may face important barriers to access the right health care.<sup>5</sup> The available country-level studies on cancer among migrants highlight disparities among individuals with a migration background, yet only offer a partial view as studies typically focus on selected groups and only compare to the host-country population.<sup>6,7</sup> These knowledge gaps are largely due to fragmentation of data and the absence of a framework to systematically integrate and analyse this data. This has hampered the quantification of cancer risk among individuals with a migration background and prevents comparisons of disparities within and between countries in Europe, and over time. To fill this knowledge gap Cancer RADAR aims to develop an infrastructure that enables quantification of the risk of cancer by migration background across Europe. To establish this, Cancer RADAR will first focus on infection-related cancers and cancers that are amenable to detection through screening, namely breast, colorectal, cervical, stomach, liver, and lung cancer. Cancer RADAR will be expanded to other cancer types in the future.

Due to past exposure to infectious agents, healthcare availability, and barriers to accessing healthcare in their birth country and host-country, migrants have a context-specific risk of infection-related cancers.<sup>6,7</sup> For example, hepatitis B virus (HBV) and *Helicobacter pylori*, respectively causing liver and gastric cancer,<sup>8</sup> are typically acquired at young age<sup>9,10</sup> prior to migration. This exposure may result in an increased risk on respectively liver and gastric cancer.<sup>6,11</sup> Cervical cancer is caused by human papillomavirus (HPV). For cervical cancer, the absence of screening in the birth country and lower screening participation in the host-country results in a higher risk to be diagnosed with cervical cancer among individuals with a migration background compared to the general population.<sup>12,13</sup> Individuals with a migration background are often found to have a lower risk to be diagnosed with colorectal and breast cancer compared to general population. This can be partly explained by differences in lifestyle factors. Yet, in some instances migrants have a higher colorectal and breast cancer mortality risk, which might reflect differences in screening attendance which are in some cases observed to be lower among individuals with a migration background.

Developing prevention strategies targeting individuals with a migration background based on their needs and personal risk may therefore result in large health gains. However, quantitative estimates of the risk and (preventable) burden of infection-related and cancers amenable to detection through screening are scarce among migrants in Europe.<sup>3,14</sup> Evidence regarding these cancers is key to inform decision making and design evidence-based prevention programs.<sup>5,15,16</sup>

### 3.2 Objectives

Aim: Cancer RADAR will develop an infrastructure to systematically collect cancer data stratified by migration background from cancer registries. This data will be combined with other data sources to provide a harmonized Europe-wide perspective on the context-specific incidence (and mortality) risk and future (preventable) burden of breast, colorectal, cervical, stomach, liver, and lung cancer among migrants in Europe.

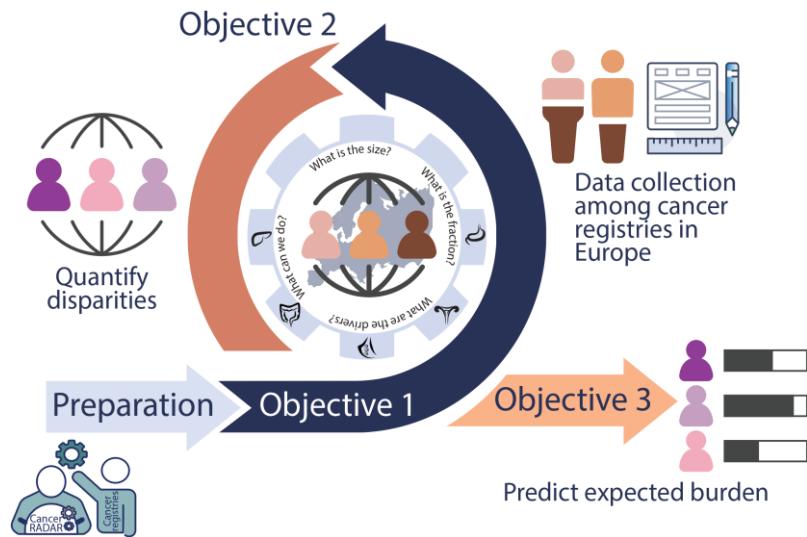


Figure 1: Schematic representation of the planned objectives within Cancer RADAR.

Objective 1 – Collect and synthesize: collect data on cancer by migration background from cancer registries across Europe by developing a flexible data collection tool to integrate various aggregated data sources and context-specific co-variables. *The methods described in this protocol focus on the data collection part of Objective 1.*

Objective 2 - Quantify and contextualize: quantify cancer risks and disparities among migrants by host-country/cancer registry region and identify drivers of disparities. Using analytical approaches to contextualize, interpret, and frame the collected data.

Objective 3 – Predict and act: estimate the future expected and preventable burden of cancer cases among migrants to identify whether the largest health gains can be made through e.g. general or migrant-targeted prevention programs.

Impact: Cancer RADAR's long-term goal is to improve the health outcome of cancer among individuals with a migration background in Europe. With this multidisciplinary approach we plan to establish an infrastructure that will provide policy makers with key analytical insights to develop evidence-based interventions, enabling cost-effectiveness assessment of tailored and targeted prevention strategies to improve health equity.

### 3.3 Work plan and tasks

#### Objective 1 – Collect and synthesize data (Figure 2):

To obtain a Europe-wide perspective on cancer among individuals with a migration background, data will be initially collected on the incidence (with mortality data and screening uptake to be collected at a later stage of Cancer RADAR) of breast, colorectal, stomach, cervical, liver and lung cancer stratified by migration background from cancer registries in Europe. **To adhere to GDPR-regulation only aggregated data and/or indicators (effect measures) will be requested and shared.** The operational definition of a migration background is based on the birth country and thus refers to first generation migrants. In a later stage Cancer RADAR plans to explore data availability on second-generation migrants. In 2023, Cancer RADAR's team developed and adapted a tool to collect data stratified by migration background. April 2024, a questionnaire has been sent to all cancer registries across Europe to assess data availability. September 2024, data collection, validation and processing will start together with pilot cancer registries (**Task-1.1**). Data will be collected through IARC-WHO and IACR. In the case a cancer registry does not have data by birth country linkage to population statistics or other data-sources will be explored (as well as corresponding funding acquisition will be supported). Finally, a map will be created on where in Europe this data is available and the process of gathering cancer data stratified by migration will be documented and evaluated to identify barriers of collection and (re)use of this data (**Task-1.2**).

To place the data obtained in the proper context, we will **integrate the collected data** with: (1) context-specific epidemiological data (prevalence and determinants of the pathogens in birth- and host-country)<sup>8,17-21</sup>; (2) health care availability and utilization in birth country and host-country<sup>3,22-25</sup>; and (3) data on the size of the migrant population per country<sup>26</sup> (**Task-1.3**). This will create an infrastructure to study drivers of disparities specific to host-countries in Europe and birth country.

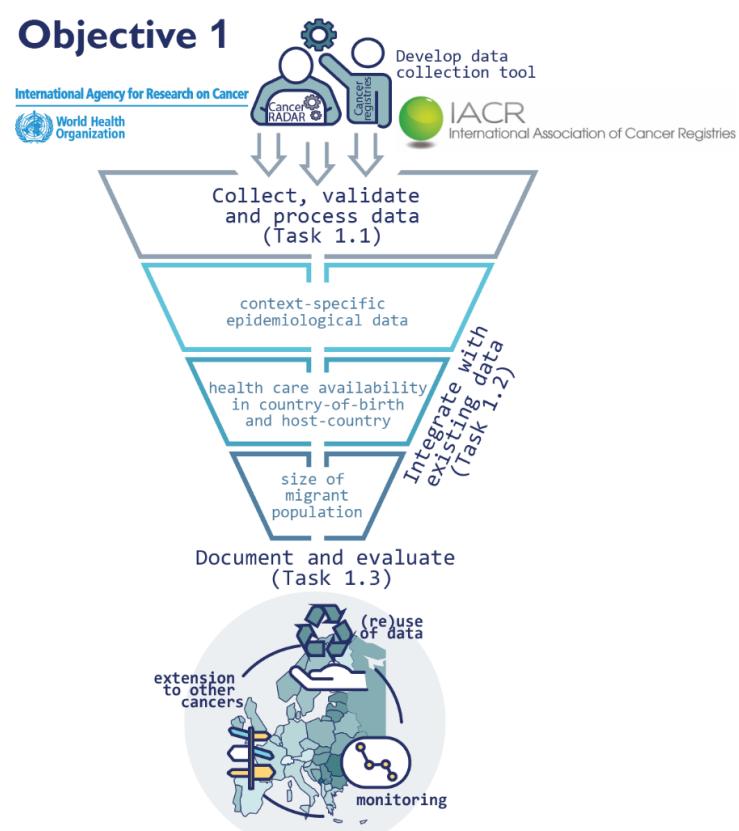


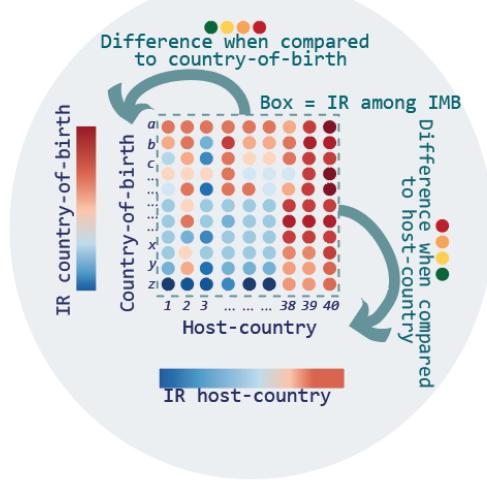
Figure 2: Schematic representation of Objective 1.

## Objective 2 - Quantify and contextualize (Figure 3):

To provide a Europe-wide comparison of cancer among individuals with a migration background, we will first **quantify cancer disparities among migrants by European country**, using the data collected in Objective 1, by estimating the (age-standardized) incidence rate (IR) of migrants, and the relative (Incidence Rate Ratio[IRR]) and absolute (Incidence Rate Difference[IRD]) differences of cancer risk among migrants when compared to the general population of the host-country and birth country (**Task-2.1**). Using the disparity measures obtained (IRD and IRR) as input for a clustering algorithm (e.g. k-means)<sup>27</sup>, we will **identify disparity clusters among migrants (Task-2.2)**. These clusters will subsequently be used to stratify migrants according to their cancer risk. This approach of stratifying migrants' risk relative to the risk in their host-country and birth country across countries in Europe will provide insights into the dynamics of cancer's risk among migrants. In some host-countries, the cancer risk among migrants may resemble the risk of the general population, while in other countries a migrant's risk may be closer to the risk observed in the birth country. Such results will inform if prevention measures targeting migrants are necessary. To identify drivers of disparities, we will determine **predictors of the observed disparity clusters (Task-2.3)**. Depending on data availability, we will use multinomial regression analysis or machine learning techniques (Random Forest) to identify which of the variables collected under **Task-1.3** discriminate most between the clusters. This integration of data sources and data-science approaches will create risk-based strata enabling harmonization of data across countries, providing actionable insights by identifying factors that drive disparities among individuals with a migration background, and hopefully in the future facilitate the creation of personalized cancer risk scores for migrants.

## Objective 2

Quantify IRC disparities among IMB  
(Task 2.1)



Identify disparity clusters (Task 2.2)      Identify disparity drivers (Task 2.3)

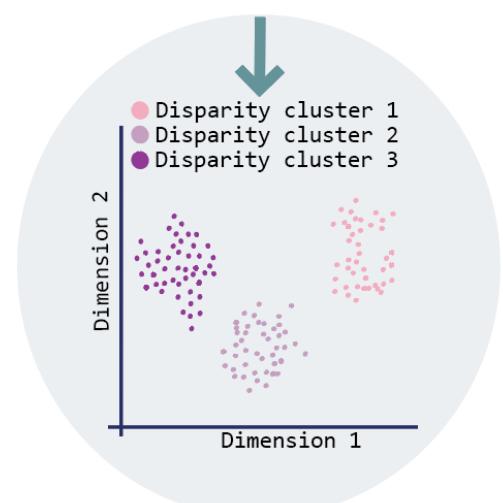
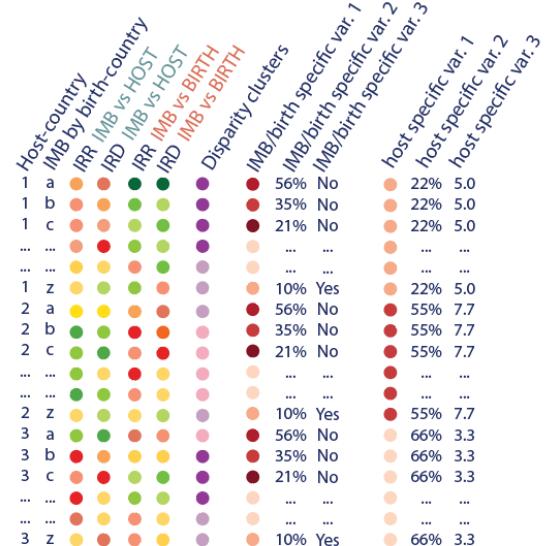
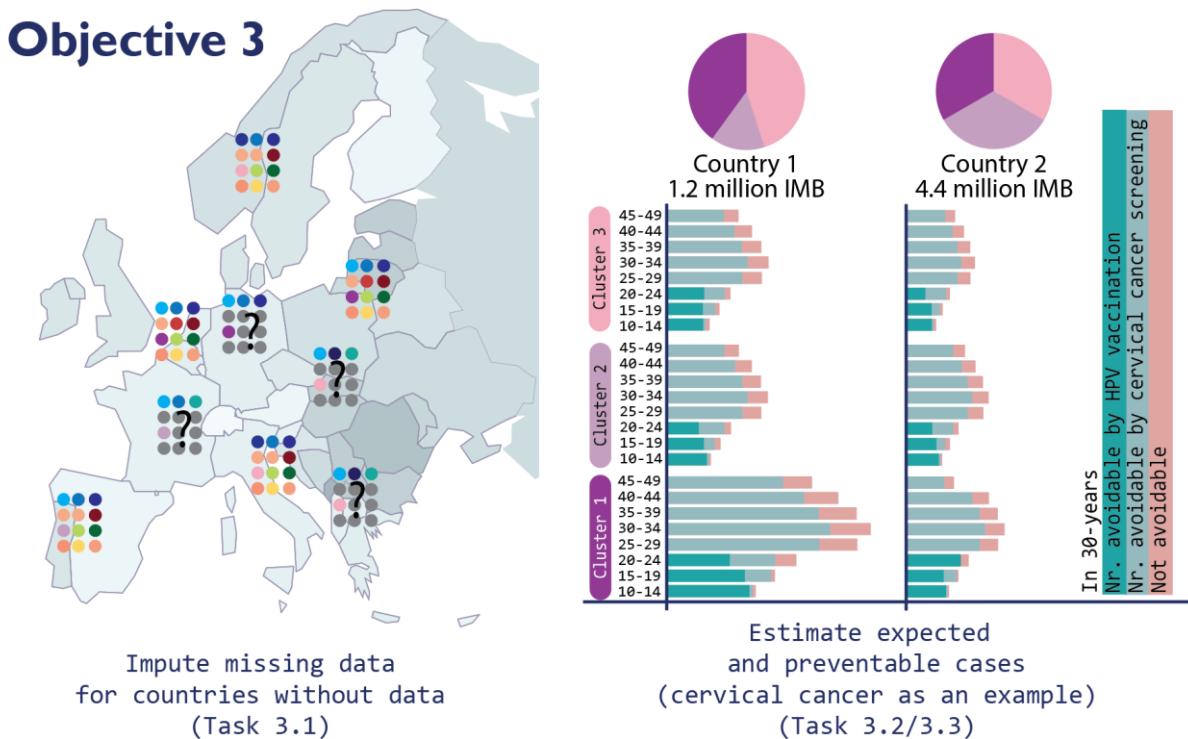


Figure 3: Schematic representation of Objective 2.

Objective 3 – Predict and act (Figure 4):



**Figure 4:** Schematic representation of Objective 3.

A major fraction of cancers can be prevented through vaccination and/or screening programs. To quantify the potential impact of prevention programs, we will estimate the number of expected and preventable cancer cases among individuals with a migration background. We will explore the possibility to **estimate migrant-specific incidence rates for countries without data** using multiple imputation techniques (**Task-3.1**). Next, we will expand an existing statistical model created for the general population<sup>28</sup> to **estimate the future expected number of cancer cases among migrants** for the European countries with data (**Task-3.2**). Next, we will estimate the maximum fraction of cancer that can be prevented (by estimating the population attributable fraction in case of infection-related cancers) and quantify the expected impact of preventative measures targeting migrants (vaccination and/or screening). This will provide a perspective on the prevention efforts necessary to eliminate any existing cancer risk disparities (**Task-3.3**).

### 3.4 Resources

Resources (personnel, computers and facilities) have been allocated by International Agency for Research on Cancer of the World Health Organization (IARC-WHO) and the International Association of Cancer Registries (IACR) in Lyon. A dedicated server is available at IARC-WHO for data storage (called OSIRIS). Project coordination, data checking and analyses will be executed (remotely) by Dr. C.J. Alberts located at Amsterdam UMC who is funded for the period Feb. 2024- Feb. 2027 by a VENI grant of the Dutch science foundation (NOW/ZonMw). Currently no funds to organize in-person meetings and travel are available yet resources for specific, additional initiatives will be sought.

### 3.5 Confidentiality, security and ethical approval

The data requested are in the form of tables of aggregated frequencies (observed cases and population at risk). After processing the generated data contains totals and indicators (e.g. Rates, SIRs, PIRs, standard errors, etc) and are thus considered highly aggregated and anonymous. Data will be stored on a dedicated server at IARC-WHO according to the standard requirements for data security. Cancer RADAR was approved by the IARC Ethics Committee (IEC) under project number IEC 23-38.

### 3.6 Contribution, publication policy and outcome

We propose to assemble an 'Advisory Board' and a 'Working Group' for the Cancer RADAR project.

The *Advisory Board* is composed of the initiators of the project (in alphabetical order) Catharina Alberts, Iacopo Baussano, Freddie Bray, Damien Georges, Irene Man, Stefano Rosso will be supplemented by representatives of cancer registries. For example, piloting registries who have contributed to the development of the protocol and data collection process.

The *Working Group* is composed of all cancer registries contributing data to the Cancer RADAR project. Per cancer type, a separate Working Group will be formed and experts per cancer type will be sought. A provisional list is provided under the section 'Involved partners'. The Working Group will be updated on the status and developments of the project at least once a year.

#### *Collaboration and ownership of data:*

- The International Association of Cancer Registries (IACR) in conjunction with the International Agency for Research on Cancer of the World Health Organization (IARC-WHO) in Lyon, and Amsterdam UMC in the Netherlands, are the co-ordinating centres.
- The Cancer RADAR data is stored at IARC-WHO in Lyon, France.
- The Cancer RADAR data remain the property of the contributing registries, whose consent is required before they can be used for purposes other than those originally envisaged in the Cancer RADAR protocols. All members of the Working Group that provide data will be informed of any analysis being carried out, and each registry has the right to oppose the inclusion of its data in analyses with which it does not agree.

*Authorship:*

- All publications using contributed data of a registry should mention the Cancer RADAR Working Group among the authors (or as the author). A suitable authorship list will be made. We suggest: a list of Authors A, B, C, ... and the Cancer RADAR Working Group. The Cancer RADAR Working Group being a list with all members in a footnote or appendix to the article.
- The first author must provide a justification to the Advisory Board for the names of all authors appearing separately in the authorship list.
- In general, the researchers who performed the analyses and wrote the paper will be named authors in the publication.

*As a minimum the following manuscripts are envisioned to be published:*

- **Objective 1: Quantification** of cancer disparities by migration background.
- **Objective 2: Contextualization** of cancer disparities by migration background. Depending on the availability of data this will be done separately for stomach, liver, cervical, breast, colorectal and lung cancer. As each type of cancer has its own relevant context.
- **Objective 3: Prediction** of expected and preventable burden of cancers among migrants. Depending on the availability of data this will be done separately for stomach, liver, cervical, breast, colorectal and lung cancer.

#### 4.1 Methods

*We will next describe the methods, strategy, and timeframe of Objective 1.*

#### 4.2 Definitions

*Individuals with a migration background (IMB)* = an individual living in another country than where he/she is born (= born outside of the host country)

*Host/general country population* = individuals born in the country where the cancer registry is located.

*Country* = data will be collected by birth country using ISO3-code. This code is used to define internationally recognized codes of letters and/or numbers that we can use when we refer to countries and their subdivisions.

Our IMB definition is in line with the UN Migration Agency (IOM) which defines a “migrant as any person who is moving or has moved across an international border or within a state away from his/her habitual place of residence, regardless of (1) the person’s legal status; (2) whether the movement is voluntary or involuntary; (3) what the causes for the movement are; or (4) what the length of the stay is.”

A disadvantage of the current definition of an IMB is that it only identifies first generation migrants and does not include second generation migrants. Second generation migrants are individuals born in the host country of which one or both parents were born outside of the host country. Because of the scope of this project, we have decided to focus on first generation migrants in this stage of the project. We expect that excluding second generation migrants within Cancer RADAR now may dilute the effects of migration, as second-generation migrants will be included in the reference/general population.

#### 4.3 Data collection process

The data collection process is divided into four steps (see Figure 5 for a schematic representation of the data collection process): [This video](#) provides the process of data collection and Figure 5 provides a schematic representation of the below described steps.

##### **Data is at the cancer registry:**

**Step 1: Data collection:** Cancer registries collect data on number of cancer cases of breast, colorectal, cervical, stomach, liver and lung cancer by birth country. Data on the population-at-risk by country of birth is also collected but optional.

**Step 2: Enter data in input file:** Cancer registries enter the collected data in an input excel-file. The excel-file has a pre-defined format to guide data entry. See Appendix 1 for more details.

**By clicking the following link, an empty input file will open automatically:**

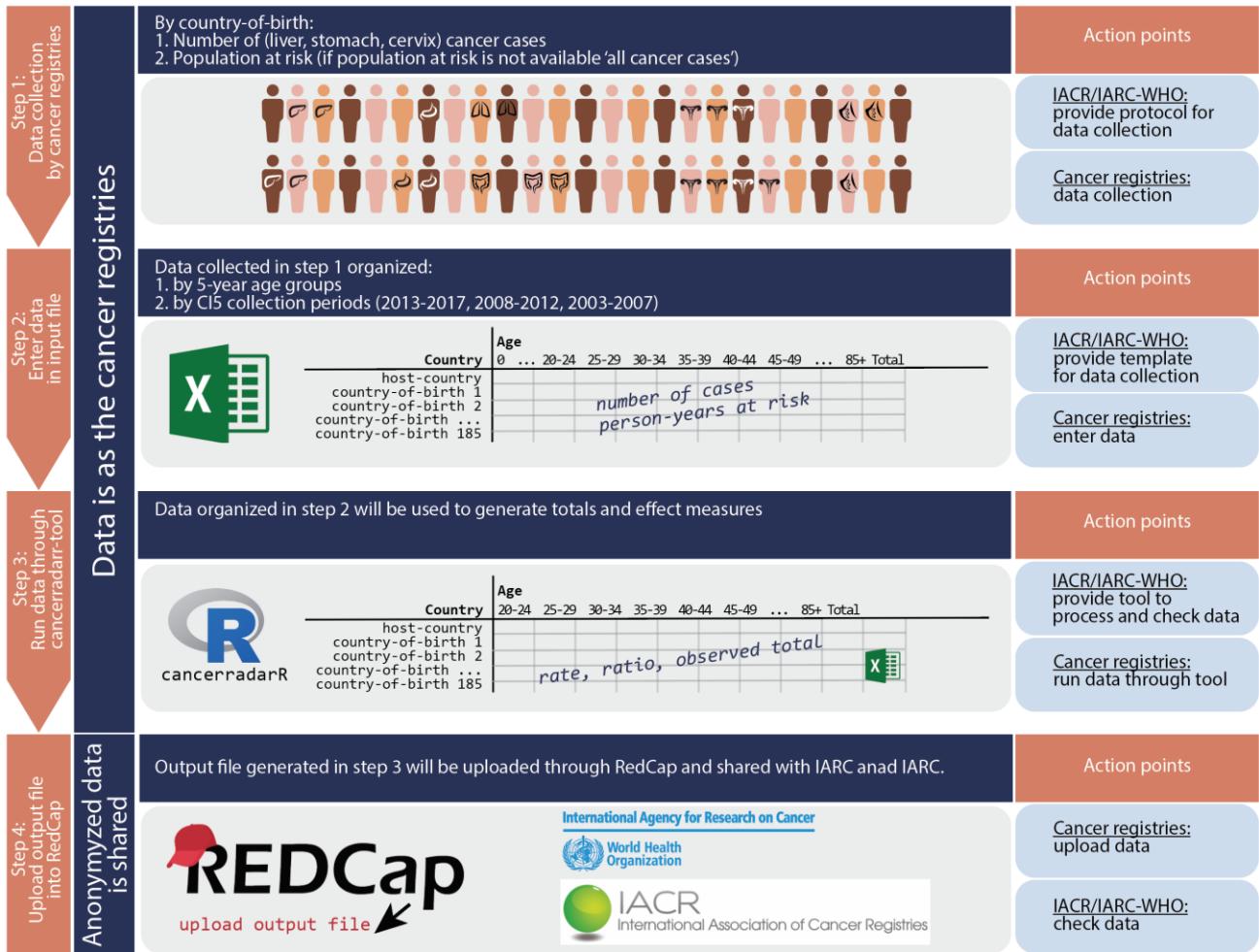
[https://gitlab.com/cancerradar/cancerradarr/-/raw/main/inst/extdata/ex\\_cancerRADAR\\_input.xlsx?ref\\_type=heads&inline=false](https://gitlab.com/cancerradar/cancerradarr/-/raw/main/inst/extdata/ex_cancerRADAR_input.xlsx?ref_type=heads&inline=false)

**Step 3: Data processing:** The completed excel-file is imported in R (an open-source software). Please follow the R commands provided here: <https://gitlab.com/cancerradar/cancerradarr/-/blob/main/README.md>. By following these commands, the Excel file will be processed automatically, generating the effect measures described in Appendix 2. Both the input and output files are stored locally.

If your computing environment does not permit the direct installation of packages, the Cancer RADAR tool for R can be downloaded [here](#).

##### **Highly aggregated data is shared:**

**Step 4: Data sharing:** The output file generated by R can be shared with IARC-WHO/IACR by uploading it to REDCap. Please email [cancer.radar@iarc.who.int](mailto:cancer.radar@iarc.who.int) when your data is ready, and you will receive a REDCap link for uploading. REDCap is a secure data-sharing platform hosted by IARC-WHO. Please avoid sending your data via email.



**Figure 5:** Schematic representation of the data collection process within Cancer RADAR.

#### 4.4 Step 1: Data collection

##### Diagnostic classification, cancer types, and population data:

Cancer registries can collect, based on availability, data on **number of cancer cases stratified by birth country** for stomach, liver, cervical, breast, colorectal, and lung cancer.

Data on the **population-at-risk stratified by birth country** is also collected but optional. However, if a cancer registry doesn't have data on the population-at-risk, we plan to use '**all cancer**' cases as the denominator, which is an indirect measure of standardization using the Proportional Incidence Ratio (PIR).<sup>29</sup> See Appendix 2 for more details on the PIR.

Cancer RADAR relies on previous infrastructure of CI5 considering data definition, quality and checks. Cancer registries can also participate if they have data for only a single cancer site.

##### For the different cancer types, we will use the following ICD codes:

- C50: Breast
- C18-20: Colorectal cancer
- C53: Cervix uteri
- C16: Stomach
- C22.0: Liver (without intrahepatic bile ducts)
- C33-34: Lung
- C00-97/C44: All cancers excl. non-melanoma skin cancer

##### Population data / Population-at-risk:

Data on the population of the area covered by each registry are optional but preferred if available. When provided, this data should be the average population over 5-year period stratified by birth country. See Appendix 1 for more details on how to calculate this.

##### Time period over which data is requested:

We expect that the fifteen-year period covering the four most recent CI5 publications (2003-2022) can offer an updated and statistically solid estimate and allows to observed changes over time. However, **cancer registries can also participate if they have data for only one time period or just a few years within a given period.**

We therefore propose to collect data stratified for the following periods:

- CI5-XIII 2018-2022
- CI5-XII 2013-2017
- CI5-XI 2008-2012
- CI5-X 2003-2007

#### 4.5 Step 2: Enter data in input file

Upon completion of data collection, cancer registries can input the data into a standardized dataset. This dataset is structured as an Excel file with predefined sheets designed to systematically organize the information based on **birth country** and **5-year age groups**. Each sheet represents one cancer type, one gender, and one time period, an additional sheet is generated for the population-at-risk. This structured approach facilitates comprehensive and uniform data analysis across registries.

By clicking the following link an empty input file will can be downloaded:

[https://gitlab.com/cancerradar/cancerradarr-/blob/main/inst/extdata/ex\\_cancerRADAR\\_input.xlsx?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr-/blob/main/inst/extdata/ex_cancerRADAR_input.xlsx?ref_type=heads)

Sheet Names: Below are the names of each sheet along with their corresponding meanings. If all data is available, for one time period, this could result in the completion of up to 14 distinct data sheets (1 [men] x 5 [cancer categories] + 1 [women] x 7 [cancer categories] + 2 [men/women] x 1 [population at risk]).

	Name sheet	Content sheet
	readme	An explanation of the collected data is provided here.
	data_info	Indicates the country and time period represented by the data in the overall Excel file.
1	py_male	Data on the population at risk (optional but recommended) for <b>men</b> . Refer to Appendix 1 for details on estimation methods.
2	nallC_male	<i>Cumulative number* of all cancer (C00-97/C44)</i> cases per 5-years age group and country of birth in <b>men</b> .
3	nliv_male	<i>Cumulative number* of liver cancer (C22.0)</i> cases per 5-years age group and country of birth in men.
4	nstm_male	<i>Cumulative number* of stomach (C16)</i> cancer cases per 5-years age group and country of birth in men.
5	ncolo_male	<i>Cumulative number* of colorectum (C18-20)</i> cancer cases per 5-years age group and country of birth in men.
6	nlung_male	<i>Cumulative number* of lung (C33-34)</i> cancer cases per 5-years age group and country of birth in men.
7	py_female	Data on the population at risk (optional but recommended) for <b>women</b> . Refer to Appendix 1 for details on estimation methods.
8	nallC_female	<i>Cumulative number* of all cancer (C00-97/C44)</i> cases per 5-years age group and country of birth in women.
9	ncx_female	<i>Cumulative number* of cervical cancer (C53)</i> cases per 5-years age group and country of birth in women.
10	nbrea_female	<i>Cumulative number* of breast cancer (C50)</i> cases per 5-years age group and country of birth in women.
11	nliv_female	<i>Cumulative number* of liver cancer (C22.0)</i> cases per 5-years age group and country of birth in women.
12	nstm_female	<i>Cumulative number* of stomach cancer (C16)</i> cases per 5-years age group and country of birth in women.
13	ncolo_female	<i>Cumulative number* of colorectum cancer (C18-20)</i> cases per 5-years age group and country of birth in women.
14	nlung_female	<i>Cumulative number* of lung cancer (C33-34)</i> cases per 5-years age group and country of birth in women.

\*Depending on the represented period, the cumulative number of cancer cases should reflect the entire duration of that period.

#### Age-categories:

For all input sheets data is requested in 5-year age groups for the full age-range 0 to 85+ years. This data is used to generate the highly aggregated output variables.

#### Birth country:

Data will be collected by birth country using ISO3-code representing 248 countries/territories. This code is used to define internationally recognized codes of letters and/or numbers that we can use when we refer to countries and their subdivisions.

For birth countries without an ISO3 code, for example due to historical reasons, a conversion table will be generated. To make sure you receive the most up to date conversion table please request this table through e-mail [cancer.radar@iarc.who.int](mailto:cancer.radar@iarc.who.int). We will send you the most up to date table.

For undocumented individuals (a person without legal documents to reside in a country) a separate category is made. Please note, that as per ENCR guidelines cancer registries are requested [not to include refugee or undocumented individuals among the cancer incidence cases](#). We have therefore added an additional category in each excel-sheet that is called 'undocumented'. This means that individuals from different birth countries are

combined into one category. A limitation of the undocumented category is that we do not differentiate by birth country. We chose this approach because it is unlikely we will have enough power to analyse this group separately. However, stratifying by birth country already allows us to understand the risk by birth country.

#### Time period and country of the registry:

In the sheet named ‘data\_info’ the registries are asked to enter:

- Country: Indicate the country where the cancer registry is located.
- Cancer Registry Designation: Provide the name of the registry or the national registry network represented by this data.
- Data Collection Period: To differentiate between the time periods, specify here which period the data pertains to (e.g., 2013-2017, 2008-2012, 2003-2007, 2008-2022). Up to three separate input files can be generated for each time period.

#### 4.6 Step 3: Data processing

Once the data has been collected, it must be processed to generate highly aggregated estimates. This is done using R (or RStudio, which provides a more user-friendly interface for interacting with R). The decision to use R was because it is open-source software and can be freely installed by any user. By executing the provided R scripts, the input file is processed, and an output file is produced. Additionally, an online guide has been developed to facilitate step-by-step execution of this process. **It is important to note that R scripts and input/output files are stored locally on the user's system.**

A step-by-step online guide is accessible via the following link:

<https://gitlab.com/cancerradar/cancerradarr/-/blob/main/README.md>

#### Data processing:

To ensure that the data remains fully anonymized and highly aggregated, the input file will be processed to generate an output file with summary effect measures, including totals, incidence rates, rate ratios and rate differences.

#### The effect measures generated are:

- Incidence Rate: crude and world age standardized (using age distribution of EU)
- *Mortality Rate: crude and age-adjusted (will be collected at a later stage of the project)*

Ratios of individuals with a migration background versus **different reference categories**

- Incidence Rate Ratio and corresponding 95% confidence interval
- Incidence Rate Difference and corresponding 95% confidence interval
- Standardised Incidence Ratio (SIR) and corresponding 95% confidence interval
- Proportional Incidence Ratio (PIR) and corresponding 95% confidence interval

Appendix 2 provides the corresponding formula for each effect measure. A data dictionary is generated, describing the Excel sheets and variable names within the output file. The data dictionary can be downloaded here:

[https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/dictionary\\_cancerradarr\\_output\\_file.xlsx?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/dictionary_cancerradarr_output_file.xlsx?ref_type=heads)

### Different reference categories will be used:

Region of the registry itself: The number of cancer cases recorded in the host country category is used to estimate regional incidence, serving as the reference point. **The script identifies the host country using the country information provided in the sheet 'data\_info'.**

National incidence of the respective cancer type in the host country: The reference incidence is derived from estimates of the respective cancer type in the host country, as provided by GLOBOCAN 2022. The script identifies the host country using the country information provided in the sheet 'data\_info'.

EU27 incidence: The reference is based on the incidence of the respective cancer type across the 27 member countries of the European Union.

EuropeUN incidence: The reference incidence is the incidence across all European countries, as defined by the UN.

### These effect measures will be calculated across different stratification levels:

1. By region, namely UN region, US subregion,
2. By Human Development Index (HDI),
3. By cancer-specific risk in the birth country,
4. Migration background (any versus no),
5. **Optional but recommended**, by birth country.

During data processing, users can opt to generate output sheets stratified by birth country. We encourage the cancer registries to first run the script generating the output by birth country. The visual output generated will allow you to identify important characteristics of the population and check for any potential mistakes that might have been generated. If a registry prefers not to share data stratified by birth country, it can **rerun** the R command specifying that stratification by birth country should not be generated. This allows the registry to choose to only share data aggregated by one of the other higher aggregation estimates.

Detailed data by birth country is requested to serve as input for Objectives 2 and 3 of the Cancer RADAR project, as described earlier in the document. The more detailed the data, the more effectively it can be utilized in our models (Objective 2 and 3 of Cancer RADAR).

### **How do we ensure anonymity/highly aggregated output?**

As indicated above, data is generated at different stratification levels. However, it is still possible that, particularly when stratified by birth country, a category contains fewer than five cancer cases. To address this, we have implemented the following precautions:

1. Evaluation of aggregation categories per cell: If an aggregation category contains frequencies below a threshold that poses a risk of personal data disclosure, the data will be aggregated into a larger category to meet the requirement of at least five cancer cases per category. For example, if the frequency in a 5-year age category falls below the defined threshold (e.g., fewer than five cancer cases per cell), the effect measure will be aggregated to a larger category, such as a 10-year (or a larger) age group. Depending on the needs, the threshold of five cancer cases per category can be increased manually. Finally, once the numerator reaches its largest aggregation, a second step is carried out to ensure that a sufficiently large denominator is available.

2. Age thresholds for data sharing: For cervical cancer, data will be provided starting from age  $\geq 20$  years. For liver, stomach, breast, colon, and lung cancers, data will be provided starting from age  $\geq 30$  years.

3. Omission of confidence intervals: If the largest possible category contains fewer than five cancer cases, only the overall incidence rate or rate ratio will be reported, without including a confidence interval. This prevents the possibility of back calculating the original number of cancer cases.

#### 4.7 Step 4: Data sharing

After data processing, three output components are generated using the R script:

Output file: A file containing tables with highly aggregated estimates (in Excel format) as described in the section above. This output file will be shared with IARC-WHO/IACR by uploading it to REDCap. Please email [cancer.radar@iarc.who.int](mailto:cancer.radar@iarc.who.int) when your data is ready, and you will receive a REDCap link for uploading. REDCap is a secure data-sharing platform hosted by IARC-WHO. Please avoid sending your data via email.

Dynamic report: An interactive report that enables users to visually explore the data and generate customized graphs based on specific interests. Cancer registries can use this output for their own purposes. We encourage cancer registries to use this output both internally and externally to describe the situation within their region or country. Additionally, we can assist in summarizing the results and compiling them into a short report.

Static report: A standardized report designed to facilitate the visual assessment of the correctness of the entered data. Each graph is accompanied by a brief text to aid in the interpretation of the graph.

#### 4.8 Data quality

The data utilized is derived from previously cleaned and processed data for the CI5 submissions, thereby ensuring high-quality data collection.

Simple checks will be done regarding the validity of the data by the R script, namely, to check for:

- Positive numbers
- Denominators larger than numerators

Checks that can be done by the cancer registries are:

- Following execution of the R script, a visual report is generated alongside the output file, containing graphical representations and standard statistical summaries. This report enables visual assessment of the dataset to verify the accuracy of the information provided in the input file.

We will also collect data on the following indicators:

- Number of **death certificate only (DCO)** cases, stratified by birth country
- Number of **microscopically verified (MV)** cases, stratified by birth country
- Number of individuals for whom the **birth country is unknown** (i.e., it is known that the individual was born outside the host country, but the specific country is not specified)
- Number of individuals for whom the **birth country is missing** (i.e., no information about the birth country is available)

Please note that we aim to minimize the amount of missing or unknown data on birth country, as such missing information can significantly bias our estimates. The exact threshold at which missing data introduces

substantial bias will need to be further investigated in this study. However, if your registry has the capability to retrieve birth country information for missing or unknown cases, it would greatly benefit the project.

When data is shared through the REDCap tool, the Cancer RADAR team will review it for trends and identify any inconsistencies, which will then be reported to the respective cancer registry.

## 5 Timeline

In November 2023, the Cancer RADAR project was launched during the joint ENCR/IACR meeting in Granada.

In April 2024, the Cancer RADAR survey was distributed to all cancer registries across Europe, and it will remain open until at least the end of 2024.

In September 2024, the data collection procedure was further detailed and explained during an online meeting with cancer registries. Recordings of the webinar are available [here](#). Data collection started and will continue until the end of December 2025.

In September 2025, analyses on data availability and feasibility of data collection will be finished

## Appendix 1 – Input data

### Entered data - general description of variable:

#### Population at risk (optional):

- py\_male: Person-years for men over a period of 5 calendar years stratified by birth country and 5-years age groups and totals.
- py\_female: Person-years for women over a period of 5 calendar years stratified by birth country and 5-years age groups and totals.

*To calculate the average population for each 5-year period, follow these steps:*

1. First, calculate the average population for each calendar year by taking the January 1<sup>st</sup> population count of that year and the January 1<sup>st</sup> population count of the following year, then divide the sum by two.
2. Once you have the average population for each calendar year, calculate the mean of these values over the 5-year period.
3. Finally, multiply this 5-year mean by 5 to obtain the total population at risk for the entire 5-year period.

#### Stomach cancer (C16):

- nstm\_male: for a period of **5 calendar years** the cumulative number of **stomach cancer cases** among **men** stratified by birth country in 5 year age-groups and totals.
- nstm\_female: for a period of **5 calendar years** the cumulative number of **stomach cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

#### Liver cancer (C22.0):

- nliv\_male: for a period of **5 calendar years** the cumulative number of **liver cancer cases** among **men** stratified by birth country in 5 year age-groups and totals.
- nliv\_female: for a period of **5 calendar years** the cumulative number of **liver cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

#### Cervical cancer (C53):

- ncx\_female: for a period of **5 calendar years** the cumulative number of **cervical cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

#### Breast cancer (C50):

- nbrst\_female: for a period of **5 calendar years** the cumulative number of **breast cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

#### Colorectal cancer (C18+C19+20):

- ncol\_male: for a period of **5 calendar years** the cumulative number of **colorectal cancer cases** among **men** stratified by birth country in 5 year age-groups and totals.
- ncol\_female: for a period of **5 calendar years** the cumulative number of **colorectal cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

### Lung cancer (C33+34):

- lnc\_male: for a period of **5 calendar years** the cumulative number of **lung cancer cases** among **men** stratified by birth country in 5 year age-groups and totals.
- lnc\_female: for a period of **5 calendar years** the cumulative number of **lung cancer cases** among **women** stratified by birth country in 5 year age-groups and totals.

### All cancer cases (C00-97/C44)\*:

This data will be used as denominator for those countries without population data. For countries for which the population at risk is known this data will also be requested. This will allow to estimate both SIR and PIR (see Appendix 2) for countries for which the denominator is known. With this in hand we will be able to establish the relationship between SIR and PIR and infer the SIR for countries for which the population-at-risk is unknown. In previous work the following relation between PIR and SIR has been established:  $\text{PIR} = \text{SIR}/\text{SIR}(\text{all cancers})$ .

- nallC\_male: for a period of **5 calendar years** the cumulative number of **all cancer cases** among **men** stratified by birth country in 5 year age-groups.
- nallC\_female: for a period of **5 calendar years** the cumulative number of **all cancer cases** among **women** stratified by birth country in 5 year age-groups.

An example of an input file can be downloaded here:

[https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/ex\\_cancerRADAR\\_input\\_filled.xlsx?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/ex_cancerRADAR_input_filled.xlsx?ref_type=heads)

An empty input file can be downloaded here:

[https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/ex\\_cancerRADAR\\_input.xlsx?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/ex_cancerRADAR_input.xlsx?ref_type=heads)

A step-by-step online guide to process the input file accessible via the following link:

[https://gitlab.com/cancerradar/cancerradarr/-/blob/main/README.md?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr/-/blob/main/README.md?ref_type=heads)

The data dictionary of the output file can be downloaded here:

[https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/dictionary\\_cancerradarr\\_output\\_file.xlsx?ref\\_type=heads](https://gitlab.com/cancerradar/cancerradarr/-/blob/main/inst/extdata/dictionary_cancerradarr_output_file.xlsx?ref_type=heads)

## Appendix 2 – Effect measures

### Formulas corresponding to the effect measures:

**Incidence Rate:** Incidence Rate expresses the number of new cases of cancer which occur in a defined population (of disease-free individuals), and the incidence rate is the number of such events in a specified period of time.

**Crude Incidence Rate (cir):** The crude, all-age rate per 100,000 will be calculated by dividing the total number of cases (R) by the total number of person-years of observation (N) over a five-year period, then multiplying the result by 100,000.

$$cir = \frac{\sum_{i=1}^A r_i}{\sum_{i=1}^A n_i}$$

Where

A = the age groups for which the number of cases and the corresponding person-years of risk can be assessed (85+ years)

$r_i$ = is the number of cases which have occurred in the  $i^{th}$  age class (0 till 85+)

$n_i$ = is the person-years of observation in the  $i^{th}$  age class during the same period of time as cases were counted (i.e. 5 year)

**Age-standardized incidence rate (asir):** An age-standardized rate represents the rate that would have occurred if the observed age-specific rates had been applied to a reference population, commonly known as the standard world population.

$$asir = \frac{\sum_{i=1}^A \frac{r_i}{n_i} * w_i}{\sum_{i=1}^A w_i}$$

**Incidence rate ratio (irr):** Is the ratio between the ir from a specific birth country and the ir of the host country.

$$irr = \frac{ir_{birth\ country}}{ir_{host\ country}}$$

**Incidence rate difference (ird):** Is the difference between the ir from a specific birth country and the ir of the host country.

$$ird = ir_{birth\ country} - ir_{host\ country}$$

**Standardised Incidence Ratio (SIR):** SIR is the ratio between the observed and expected number of cases. The expected number of cases reflects the number of cases expected in the migrant group under the assumption that the incidence rate in the group is equal to the incidence rate in the general population. To estimate this expected number of cases, we need to know how many individuals from a respective migrant group live in the area of the country of residence the registry covers, and multiply that with the corresponding incidence rate in the age-group of the corresponding host population (to do this the size of the population-at-risk, namely migrant population X, needs to be known).

$$SIR = \frac{\sum_x Observed_x}{\sum_x Expected_x} = \frac{\sum_{i,x} IR_{migrant_{i,x}} \times Pop_{migrant_{i,x}}}{\sum_{i,x} IR_{general_i} \times Pop_{migrant_{i,x}}}$$

where

$i$  = age-group

$x$  = migrant population of interest

$Observed_x$  = observed number of cervical cancer cases in migrant population  $x$

$Expected_x$  = expected number of cervical cancer cases in migrant population  $x$  under the assumption that the incidence rate in the group is equal to the incidence rate in the general population

$IR$  = Incidence rate

$Pop$  = size of population

**Proportional Incidence Ratio (PIR):** PIR is the ratio between the observed and expected number of cases in a migrant group of interest. The expected number is estimated by multiplying the total number of all cancer cases (or a selection of cancer types) in each age group from migrant group X by the corresponding age- and cause-specific proportion of cervical cancer in the host population. **The advantage of this method is that the size of the population at risk does not need to be known.** A disadvantage is that this method only reflects how much higher the proportion of cervical cancer cases is in the migrant population than in the host population, but not if the risk of the disease is elevated.

$$PIR = \frac{\sum_x Observed_x}{\sum_x Expected_x} = \frac{\sum_{i,x} Observed \text{ in migrant}_{i,x}}{\sum_{i,x} t_{i,x} \times \frac{r_i^*}{t_i^*}}$$

where

$i$  = age-group

$Observed$  = observed number of cervical cancer cases in migrant population  $x$

$Expected$  = expected number of cervical cancer cases in migrant population  $x$

$t_{i,x}$  = number of cases of cancers at all sites (or a selection of sites) in the age group  $i$  in migrant population  $x$

$r_i^*$  = number of cases of cervical cancer cases in the age group  $i$  in the general population

$t_i^*$  = number of cases of cancer at all sites (or a selection of sites) in the age group  $i$  in general population

Note that the PIR can be considered a proxy for the underlying risk only if the rates of cancer at all sites in both the general and migrant populations are similar—an assumption that is rarely met. In other words, the PIR will be equivalent to the SIR only if the overall cancer rates (excluding skin cancer) are the same in the general and migrant populations, which is unlikely due to inherent differences in cancer risk. In practice, it has been found that  $PIR = SIR/SIR(\text{all cancers})$ . This study will help shed more light on these assumptions and the possibilities of inferring the PIR from the SIR. A better understanding of this relationship will allow to use this Cancer RADAR data collection framework for registries and countries where the population-at-risk is not available.

### Why are IRR, SIR, and PIR estimated instead of just the IRR?

If the age distribution of the migrant population is similar to that of the World Standard Population, we expect the SIR and IRR to be nearly identical. The **IRR is a direct comparison** of two incidence rates (e.g., migrants and non-migrants), both adjusted using a common reference population (in our case, the World Standard Population), which allows for a standardized comparison between groups. In contrast, the **SIR is an indirect method of comparison** and compares the observed number of cases in the study group (e.g., migrants) to the expected number of cases, calculated by applying the age-specific incidence rates from a reference population (such as the host population) to the age distribution of the study group (e.g., migrants). Essentially, the SIR helps

estimate how many cases would be expected if the migrant group experienced the same age-specific risk as the reference population.

In addition, in cases where the population-at-risk data stratified by birth country are not available in the cancer registry database, we will use the PIR. As mentioned, there is an ecological relationship between PIR and SIR, making it useful to estimate the SIR as well, to allow for this inference. This approach helps to better understand the cancer risk dynamics, even when certain data are limited or missing.

## Literature

1. McAuliffe, M. and A. Triandafyllidou (eds.), 2021. World Migration Report 2022. International Organization for Migration (IOM), Geneva.
2. World report on the health of refugees and migrants. Geneva: World Health Organization; 2022.
3. European Commission, Consumers, Health, Agriculture and Food Executive Agency, Report on the health status of newly arrived migrants and refugees in EU/EEA, Publications Office, 2020.,
4. Report on the health of refugees and migrants in the WHO European Region: no public health without refugee and migrant health (2018) ISBN: 978-92-890-5384-6.
5. Seedat F, Hargreaves S, Nellums LB, Ouyang J, Brown M, Friedland JS. How effective are approaches to migrant screening for infectious diseases in Europe? A systematic review. *The Lancet Infectious Diseases* 2018; **18**(9): e259-e71.
6. Pabla BS, Shah SC, Corral JE, Morgan DR. Increased Incidence and Mortality of Gastric Cancer in Immigrant Populations from High to Low Regions of Incidence: A Systematic Review and Meta-Analysis. *Clin Gastroenterol Hepatol* 2020; **18**(2): 347-59 e5.
7. Arnold M, Razum O, Coebergh JW. Cancer risk diversity in non-western migrants to Europe: An overview of the literature. *Eur J Cancer* 2010; **46**(14): 2647-59.
8. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet Global Health* 2020; **8**(2): e180-e90.
9. IARC Helicobacter pylori Working Group (2014). Helicobacter pylori Eradication as a Strategy for Preventing Gastric Cancer. Lyon, France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8).
10. Global health sector strategies on, respectively, HIV, viral hepatitis and sexually transmitted infections for the period 2022-2030. Geneva: World Health Organization; 2022.
11. Mitchell T, Nayagam JS, Dusheiko G, Agarwal K. Health inequalities in the management of chronic Hepatitis B Virus infection in patients from sub-Saharan Africa in high income countries. *JHEP Reports* 2022.
12. Brotherton J, Davies C, Committee IP. IPVS policy statement. Equity in cervical cancer prevention: for all and not just for some. *Papillomavirus Res* 2020; **9**: 100192.
13. Marques P, Nunes M, Antunes MDL, Heleno B, Dias S. Factors associated with cervical cancer screening participation among migrant women in Europe: a scoping review. *Int J Equity Health* 2020; **19**(1): 160.
14. Global Migration Indicators, Global Migration Data Analysis Centre. *International Organization for Migration, Berlin* 2018.
15. Graetz V, Rechel B, Groot W, Norredam M, Pavlova M. Utilization of health care services by migrants in Europe-a systematic literature review. *Br Med Bull* 2017; **121**(1): 5-18.
16. Noori T, Hargreaves S, Greenaway C, et al. Strengthening screening for infectious diseases and vaccination among migrants in Europe: What is needed to close the implementation gaps? *Travel Med Infect Dis* 2021; **39**: 101715.
17. Bruni L, Albero G, Serrano B, Mena M, Collado JJ, Gómez D, Muñoz J, Bosch FX, de Sanjosé S. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in the World. Summary Report 22 October 2021.
18. Sheena BS, Hiebert L, Han H, et al. Global, regional, and national burden of hepatitis B, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Gastroenterology & Hepatology* 2022; **7**(9): 796-829.
19. Hooi JK, Lai WY, Ng WK, et al. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* 2017; **153**(2): 420-9.
20. Bray F, Colombet M, Mery L, et al. Cancer incidence in five continents. Lyon: International Agency for Research on Cancer 2017; **11**.
21. Vardell E. Global health observatory data repository. *Medical reference services quarterly* 2020; **39**(1): 67-74.
22. von Karsa L, Anttila A, Ronco G, et al. Cancer screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening: European Commission; 2008.

23. Bruni L, Saura-Lázaro A, Montoliu A, et al. HPV vaccination introduction worldwide and WHO and UNICEF estimates of national HPV immunization coverage 2010–2019. *Preventive Medicine* 2021; **144**: 106399.
24. Bruni L, Serrano B, Roura E, et al. Cervical cancer screening programmes and age-specific coverage estimates for 202 countries and territories worldwide: a review and synthetic analysis. *The Lancet Global Health* 2022; **10**(8): e1115-e27.
25. WHO/UNICEF Joint Reporting Form on Immunization (JRF). Immunization dashboard available from [immunizationdata.who.int/index.html](http://immunizationdata.who.int/index.html).
26. United Nations, Department of Economic and Social Affairs. Population Division (2020). International Migrant Stock 2020 (United Nations database, POP/DB/MIG/Stock/Rev.2020).
27. Wiemken TL, Kelley RR. Machine Learning in Epidemiology and Health Outcomes Research. *Annual review of public health* 2019; **41**: 21-36.
28. Bonjour M, Charvat H, Franco EL, et al. Global estimates of expected and preventable cervical cancers among girls born between 2005 and 2014: a birth cohort analysis. *The Lancet Public Health* 2021; **6**(7): e510-e21.
29. Boyle P, Parkin DM. Chapter 11 . Statistical methods for registries. 2006; 2006.