# Standardised rates

# Tỷ suất chuẩn hóa

This page will show you two ways to standardize an outcome, such as hospitalizations or mortality, by characteristics such as age and sex.

Chương này sẽ hướng dẫn hai cách để chuẩn hóa một kết quả như là tỷ suất tử vong hoặc nhập viện theo các đặc điểm của tuổi và giới.

* Using **dsr** package
* Using **PHEindicatormethods** package
* Sử dung gói lệnh **dsr**
* Sử dụng gói lệnh **PHEindicatormethods**

We begin by extensively demonstrating the processes of data preparation/cleaning/joining, as this is common when combining population data from multiple countries, standard population data, deaths, etc.

Chúng ta bắt đầu bằng cách làm rõ quá trình chuẩn bị/làm sạch/kết nối số liệu, vì việc này thường được thực hiện khi kết nối số liệu dân số từ nhiều quốc gia, số liệu dân số chuẩn, tử vong,.v.v.

## Overview

## Tổng quan

There are two main ways to standardize: direct and indirect standardization. Let’s say we would like to the standardize mortality rate by age and sex for country A and country B, and compare the standardized rates between these countries.

Có hai cách để chuẩn hóa: trực tiếp và gián tiếp. Giả sử chúng ta muốn chuẩn hóa tỷ suất tử vong theo tuổi và giới của quốc gia A và quốc gia B, và so sánh tỷ lệ giữa hai quốc gia này.

* For direct standardization, you will have to know the number of the at-risk population and the number of deaths for each stratum of age and sex, for country A and country B. One stratum in our example could be females between ages 15-44.
* For indirect standardization, you only need to know the total number of deaths and the age- and sex structure of each country. This option is therefore feasible if age- and sex-specific mortality rates or population numbers are not available. Indirect standardization is furthermore preferable in case of small numbers per stratum, as estimates in direct standardization would be influenced by substantial sampling variation.
* Đối với chuẩn hóa trực tiếp, chúng ta phải biết dân số nguy cơ và số tử vong theo từng tầng của tuổi và giới của quốc gia A và B. Một tầng trong ví dụ của chúng tôi là phụ nữ trong nhóm tuổi từ 15-44.
* Đối với chuẩn hóa gián tiếp, chúng ta chỉ cần biết tổng số tử vong và cấu trúc tuổi và giới của mỗi quốc gia. Vì thế, phương án này là khả khi tỷ suất tử vong hoặc dân số theo tuổi và giới không có sẵn. Hơn nữa, chuẩn hóa gián tiếp được ưu chuộng hơn khi có số lượng nhỏ trong các tầng, vì các ước tính trong chuẩn hóa trực tiếp sẽ bị ảnh hưởng bởi các thay đổi quan trọng khi lấy mẫu.

## Preparation

## Chuẩn bị

To show how standardization is done, we will use fictitious population counts and death counts from country A and country B, by age (in 5 year categories) and sex (female, male). To make the datasets ready for use, we will perform the following preparation steps:

Để chỉ dẫn cách thực hiện chuẩn hóa, chúng tôi tạo ra dân số và số tử vong của quốc gia A và B theo tuổi (trong các nhóm tuổi của mỗi 5 tuổi) và giới (nữ, nam). Để các bộ số liệu có thể sử dụng được, chúng ta sẽ thực hiện các bước chuẩn bị như sau:

1. Load packages
2. Load datasets
3. Join the population and death data from the two countries
4. Pivot longer so there is one row per age-sex stratum
5. Clean the reference population (world standard population) and join it to the country data
6. Tải các gói lệnh
7. Tải các bộ số liệu
8. Kết nối số liệu dân số và tử vong của hai quốc gia
9. Chuyển đổi thành bộ số liệu dọc/dài hơn để có một hàng là một tầng theo tuổi-giới
10. Làm sạch dân số tham chiếu (dân số chuẩn) và kết nối nó với số liệu các quốc gia trên

In your scenario, your data may come in a different format. Perhaps your data are by province, city, or other catchment area. You may have one row for each death and information on age and sex for each (or a significant proportion) of these deaths. In this case, see the pages on [Grouping data](#grouping-data), [Pivoting data](#pivoting-data), and [Descriptive tables](#descriptive-tables) to create a dataset with event and population counts per age-sex stratum.

Trong trường hợp này, số liệu của chúng ta có thể có các định dạng khác nhau. Có lẽ số liệu của chúng ta theo tỉnh, thành phố, hoặc khu vực dịch vụ khác. Chúng ta có thể có một hàng cho mỗi ca tử vong và các thông tin về tuổi và giới cho mỗi trường hợp (hoặc một tỷ lệ có ý nghĩa) của các ca tử vong này. Trong trường hợp này, hãy xem các chương về [Nhóm số liệu](#grouping-data), [Chuyển đổi số liệu](#pivoting-data), và [Bảng mô tả số liệu](#descriptive-tables) để tạo một bộ số liệu với số lượng tử vong và dân số theo tầng của tuổi-giới tính.

We also need a reference population, the standard population. For the purposes of this exercise we will use the world\_standard\_population\_by\_sex. The World standard population is based on the populations of 46 countries and was developed in 1960. There are many “standard” populations - as one example, the website of [NHS Scotland](https://www.opendata.nhs.scot/dataset/standard-populations) is quite informative on the European Standard Population, World Standard Population and Scotland Standard Population.

Chúng ta cũng cần một dân số tham chiếu, dân số chuẩn. Với mục đích của bài tập này, chúng ta sẽ sử dụng bộ sối liệu world\_standard\_population\_by\_sex. Dân số chuẩn này dựa trên đân số của 46 quốc gia và được tạo ra vào năm 1960. Có rất nhiều dân số “chuẩn” – ví dụ, trang web của [NHS Scotland](https://www.opendata.nhs.scot/dataset/standard-populations) cung cấp khá nhiều thông tin về Dân số Chuẩn của châu Âu, Dân số Chuẩn Thế giới, Dân số Chuẩn của Scotland.

### Load packages

### Tải gói lệnh

This code chunk shows the loading of packages required for the analyses. In this handbook we emphasize p\_load() from **pacman**, which installs the package if necessary and loads it for use. You can also load installed packages with library() from **base** R. See the page on [R basics](#r-basics) for more information on R packages.

Đoạn mã này hiển thị việc tải các gói lệnh cần thiết cho phân tích. Trong cuốn sổ tay này, chúng tôi nhấn mạnh đến hàm p\_load()trong gói lệnh **pacman**, cài đặt gói lệnh nếu cần thiết và tải để sử dụng. Các gói lệnh đã cài đặt cũng có thể được tải bằng library()từ **base** R. Xem thêm thông tin các gói lệnh của R trong chương [Cơ](#r-basics) bản về R.

**CAUTION:** If you have a newer version of R, the **dsr** package cannot be directly downloaded from CRAN. However, it is still available from the CRAN archive. You can install and use this one.

**LƯU Ý:** Nếu đang sử dụng một phiên bản mới hơn của R, chúng ta không thể tải gói lệnh **dsr** trực tiếp từ CRAN. Tuy nhiên, nó vẫn còn trong mục lưu trữ của CRAN. Chúng ta có thể cài đặt và sử dụng nó.

For non-Mac users:

Cho người dùng không sử dụng hệ điều hành Mac

For Mac users:

Cho người dùng hệ điều hành Mac

### Load population data

### Tải số liệu dân số

See the [Download handbook and data](#download-handbook-and-data) page for instructions on how to download all the example data in the handbook. You can import the Standardisation page data directly into R from our Github repository by running the following import() commands:

First we load the demographic data (counts of males and females by 5-year age category) for the two countries that we will be comparing, “Country A” and “Country B”.

Xem chương [Tải sổ tay và số liệu](#download-handbook-and-data) để biết các hướng dẫn về cách tải tất cả các bộ số liệu mẫu trong cuốn số tay này. Chúng ta có thể nhập trực tiếp số liệu Chuẩn hóa vào R từ kho lưu trữ trong Github của chúng tôi bằng cách thực hiện các lệnh import():

Trước tiên, chúng ta tải số liệu nhân khẩu (số lượng nam và nữ theo nhóm mỗi 5 tuổi) của hai quốc gia mà chúng ta sẽ so sánh, “Quốc gia A” và “Quốc gia B”.

### Load death counts

### Tải số liệu tử vong

Conveniently, we also have the counts of deaths during the time period of interest, by age and sex. Each country’s counts are in a separate file, shown below.

Để thuận tiện cho việc thực hành, chúng tôi cũng có số lượng tử vong theo tuổi và giới trong khoảng thời gian quan tâm. Số lượng tử vong của mỗi quốc gia nằm trong một tập tin riêng, được hiện thị bên dưới.

Deaths in Country A

Tử vong trong Quốc gia A

Deaths in Country B

Tử vong trong Quốc gia B

### Clean populations and deaths

### Làm sạch số liệu dân số và tử vong

We need to join and transform these data in the following ways:

Chúng ta cần kết nối và chuyển đổi các số liệu này theo những cách sau:

* Combine country populations into one dataset and pivot “long” so that each age-sex stratum is one row
* Combine country death counts into one dataset and pivot “long” so each age-sex stratum is one row
* Join the deaths to the populations
* Liên kết dân số của các quốc gia vào một bộ số liệu và chuyển đổi thành bộ số liệu dạng “dọc/dài” để mỗi tầng tuổi-giới là một hàng
* Liên kết số liệu tử vong của các quốc gia vào một bộ số liệu và chuyển đổi thành bộ số liệu dạng “dọc/dài” để mà mỗi tầng tuổi-giới là một hàng
* Nối số liệu tử vong với số liệu dân số

First, we combine the country populations datasets, pivot longer, and do minor cleaning. See the page on [Pivoting data](#pivoting-data) for more detail.

Đầu tiên, chúng ta liên kết các bộ số liệu dân số quốc gia, chuyển đổi thành số liệu dạng dọc/dài hơn, và làm sạch số liệu. Xem chương [Chuyển đổi số liệu](#pivoting-data) để biết thêm chi tiết.

Số liệu dân số đã liên kết bây giờ trông giống như sau (nhấp qua để xem cột quốc gia A và B):

Và bây giờ, chúng ta thực hiện các thao tác tương tự trên hai bộ số liệu tử vong.

The deaths data now look like this, and contain data from both countries:

Số liệu tử vong bây giờ trông giống như sau và chứa số liệu của cả hai quốc gia:

We now join the deaths and population data based on common columns Country, age\_cat5, and Sex. This adds the column Deaths.

Bây giờ, chúng ta nối số liệu tử vong và số liệu dân số dựa trên các cột chung Country, age\_cat5, và Sex. Thao tác này sẽ thêm cột Deaths.

We can now classify Sex, age\_cat5, and Country as factors and set the level order using fct\_relevel() function from the **forcats** package, as described in the page on [Factors](#factors). Note, classifying the factor levels doesn’t visibly change the data, but the arrange() command does sort it by Country, age category, and sex.

Bây giờ, chúng ta có thể phân loại các biến số Sex, age\_cat5, và Country dưới dạng factors và đặt thứ tự các giá trị của biến số bằng cách sử dụng hàm fct\_relevel() từ gói lệnh **forcats**, như được mô tả trong chương [Factors](#factors). Lưu ý rằng việc phân loại giá trị biến không thay đổi số liệu một cách rõ ràng, nhưng lệnh arrange() sẽ sắp xếp nó theo Quốc gia, nhóm tuổi và giới tính.

**CAUTION:** If you have few deaths per stratum, consider using 10-, or 15-year categories, instead of 5-year categories for age.

**Lưu ý:** Nếu mỗi tầng có it ca tử vong trong, hay cân nhắc sử dụng nhóm tuổi mỗi 10 hoặc 15 tuổi thay vì mỗi 5 tuổi cho các nhóm.

### Load reference population

### Tải dân số tham chiếu

Lastly, for the direct standardisation, we import the reference population (world “standard population” by sex)

Cuối cùng, để chuẩn hóa trực tiếp, chúng ta nhập dân số tham chiếu (“dân số chuẩn” thế giới theo giới tính)

### Clean reference population

### Làm sach dân số tham chiếu

The age category values in the country\_data and standard\_pop\_data data frames will need to be aligned.

Các giá trị của biến tuổi trong bộ số liệu country\_data và standard\_pop\_data cần phải được căn chỉnh.

Currently, the values of the column age\_cat5 from the standard\_pop\_data data frame contain the word “years” and “plus”, while those of the country\_data data frame do not. We will have to make the age category values match. We use str\_replace\_all() from the **stringr** package, as described in the page on [Characters and strings](#characters-and-strings), to replace these patterns with no space "".

Hiện tại, các giá trị của cột biến phân loại age\_cat5 trong bộ số liệu standard\_pop\_data có chứa từ “years” và “plus”, trong khi không có các từ này trong trong bộ số liệu country\_data. Chúng ta cần phải chuyển đổi các giá trị này giống nhau. Chúng ta sẽ sử dụng hàm str\_replace\_all() trong gói lệnh **stringr**, như được mô tả trong chương [Ký tự và chuỗi](#characters-and-strings), để thay thế thành những kiễu mẫu không có khoảng trắng "".

Furthermore, the package **dsr** expects that in the standard population, the column containing counts will be called "pop". So we rename that column accordingly.

Hơn nữa, gói lệnh **dsr** đòi hỏi cột chứa số lượng trong bộ số liệu dân số chuẩn phải được gọi là "pop". Vì vậy, chúng ta đổi tên cột này cho phù hợp.

**CAUTION:** If you try to use str\_replace\_all() to remove a plus symbol, it won’t work because it is a special symbol. “Escape” the specialnes by putting two back slashes in front, as in str\_replace\_call(column, "\\+", "").

**THẬN TRỌNG:** Nếu chúng ta cố gắng sử dụng hàm str\_replace\_all() để loại bỏ biểu tương dấu cộng, hàm này sẽ không thực hiện bởi vì nó là một biểu tượng đặc biệt. “Xử lý” các biểu tượng này bằng cách đặt hai dấu gạch chéo phía trước như trong lệnh str\_replace\_call(column, "\\+", "").

### Create dataset with standard population

### Tạo bộ số liệu với dân số chuẩn

Finally, the package **PHEindicatormethods**, detailed [below](#standard_phe), expects the standard populations joined to the country event and population counts. So, we will create a dataset all\_data for that purpose.

Cuối cùng, sử dụng gói lệnh **PHEindicatormethods**, được trình bày chi tiết [bên dưới](#standard_phe), để nối dân số chuẩn vào biến tử vong và dân số của các quốc gia. Vì vậy, chúng ta sẽ tạo một bộ số liệu với tên all\_data cho mục đich đó.

This complete dataset looks like this:

Bộ số liệu hoàn chỉnh trong như thế này:

## **dsr** package

## Gói lệnh dsr

Below we demonstrate calculating and comparing directly standardized rates using the **dsr** package. The **dsr** package allows you to calculate and compare directly standardized rates (no indirectly standardized rates!).

Bên dưới, chúng tôi trình bày việc tính toán và so sánh các tỷ suất được chuẩn hóa trực tiếp bằng các sử dụng gói lệnh **dsr**. Gói lệnh **dsr** cho phép chúng ta tính toán và so sánh các tỷ suất được chuẩn hóa trực tiếp (không phải các tỷ suất được chuẩn hóa gián tiếp!).

In the data Preparation section, we made separate datasets for country counts and standard population:

Trong phần Chuẩn bị số liệu, chúng ta đã tạo các bộ số liệu riêng biệt cho các quốc gia và dân số chuẩn:

1. the country\_data object, which is a population table with the number of population and number of deaths per stratum per country
2. the standard\_pop\_clean object, containing the number of population per stratum for our reference population, the World Standard Population
3. đối tượng country\_data, mà là một bảng dân số với số lượng dân số và số tử vong theo tầng cho mỗi quốc gia
4. đối tượng standard\_pop\_clean, bao gồm số lượng dân số theo tầng cho dân số tham chiếu, Dân số Chuẩn Thế giới

We will use these separate datasets for the **dsr** approach.

Chúng ta sẽ sử dụng các bộ số liệu riêng rẽ này để tiếp cận gói lệnh **dsr**

### Standardized rates

### Tỷ suất chuẩn hóa

Below, we calculate rates per country directly standardized for age and sex. We use the dsr() function.

Bên dưới, chúng ta tính toán các tỷ suất theo quốc gia được chuẩn hóa trực tiếp theo tuổi và giới. Chúng ta sử dụng hàm dsr()

Of note - dsr() expects one data frame for the country populations and event counts (deaths), and a ***separate*** data frame with the reference population. It also expects that in this reference population dataset the unit-time column name is “pop” (we assured this in the data Preparation section).

Lưu ý - dsr() đòi hỏi một bộ số liệu dân số của các quốc gia và số lượng sự kiện (tử vong), và một bộ số liệu ***riêng rẽ*** với dân số tham chiếu. Nó cũng đòi hỏi trong bộ số liệu của dân số tham chiếu này tên cột đơn vị-thời gian là “pop” (chúng ta đã đảm bảo điều này trong phần Chuẩn bị số liệu).

There are many arguments, as annotated in the code below. Notably, event = is set to the column Deaths, and the fu = (“follow-up”) is set to the Population column. We set the subgroups of comparison as the column Country and we standardize based on age\_cat5 and Sex. These last two columns are not assigned a particular named argument. See ?dsr for details.

Có nhiều đối số, như được chú thích bên dưới. Đáng chú ý, đối số event = được đặt thành cột Deaths, và đối số fu = (“follow-up”) được đặt thành cột Population. Chúng ta đặt các nhóm so sánh là cột Country và chuẩn hóa dựa trên age\_cat5 và Sex. Hai cột cuối cùng này không được gán với một đối số có tên cụ thể. Gõ ?dsr để biết thêm chi tiết.

Above, we see that while country A had a lower crude mortality rate than country B, it has a higher standardized rate after direct age and sex standardization.

Bên trên, chúng ta thấy rằng tỷ suất tử vong thô của quốc gia A thấp hơn quốc gia B, nhưng tỷ suất được chuẩn hóa trực tiếp theo tuổi và giới thì cao hơn.

### Standardized rate ratios

### Tỷ số của các tỷ suất chuẩn hóa

The standardized mortality rate is 1.22 times higher in country A compared to country B (95% CI 1.17-1.27).

Tỷ suất tử vong chuẩn hóa của quốc A bằng 1.22 lần so với quốc gia B (KTC 95% 1.17-1.27).

### Standardized rate difference

### Khác biệt của các tỷ suất chuẩn hóa

Country A has 4.24 additional deaths per 100.000 population (95% CI 3.24-5.24) compared to country A.

Quốc gia A có hơn 4,24 ca tử vong trên 100,000 dân (KTC 95% 3.24-5.24) so với quốc gia B.

## **PHEindicatormethods** package

## Gói lệnh **PHEindicatormethods**

Another way of calculating standardized rates is with the **PHEindicatormethods** package. This package allows you to calculate directly as well as indirectly standardized rates. We will show both.

Một cách khác để tính tỷ suất chuẩn hóa là dùng gói lệnh **PHEindicatormethods**. Gói lệnh này cho phép chúng ta tính toán tỷ suất chuẩn hóa trực tiếp cũng như gián tiếp. Chúng tôi sẽ hướng dẫn cả hai.

This section will use the all\_data data frame created at the end of the Preparation section. This data frame includes the country populations, death events, and the world standard reference population. You can view it [here](#standard_all).

Phần này chúng ta sẽ sử dụng bộ số liệu all\_data được tạo ở cuối phần Chuẩn bị. Bộ số liệu này bao gồm dân số quốc gia, số tử vong và dân số tham chiếu. Chúng ta có thể xem ở [đây](#standard_all).

### Directly standardized rates

### Tỷ suất chuẩn hóa trực tiếp

Below, we first group the data by Country and then pass it to the function phe\_dsr() to get directly standardized rates per country.

Đầu tiên, chúng ta nhóm số liệu theo Quốc gia như bên dưới và sau đó chuyển số liệu vào hàm phe\_dsr() để có được tỷ suất chuẩn hóa trực tiếp theo quốc gia.

Of note - the reference (standard) population can be provided as a **column within the country-specific data frame** or as a **separate vector**. If provided within the country-specific data frame, you have to set stdpoptype = "field". If provided as a vector, set stdpoptype = "vector". In the latter case, you have to make sure the ordering of rows by strata is similar in both the country-specific data frame and the reference population, as records will be matched by position. In our example below, we provided the reference population as a column within the country-specific data frame.

Lưu ý, dân số tham chiếu (chuẩn) có thể được cung cấp dưới dạng một cột bên trong số liệu của một quốc gia cụ thể hoặc dưới dạng một véc tơ riêng biệt. Nếu được cung cấp bên trong số liệu của một quốc gia cụ thể, chúng ta phải đặt stdpoptype = "field". Nếu được cung cấp như một véc tơ, đặt stdpoptype = "vector". Trong trường hợp thứ hai, chúng ta phải đảm bảo thứ tự của các hàng theo tầng tương tự nhau trong cả khung số liệu quốc gia và dân số tham chiếu, vì vị trí các bản ghi/các hàng phải khớp với nhau. Trong ví dụ bên dưới, chúng tôi đã cung cấp dân số tham chiếu dưới dạng một cột bên trong khung số liệu của một quốc gia cụ thể.

See the help with ?phr\_dsr or the links in the References section for more information.

Xem trợ giúp bằng cách gõ ?phr\_dsr hoặc các liên kết trong phần Tài liệu tham khảo để biết thêm thông tin.

### Indirectly standardized rates

### Tỷ suất chuẩn hóa gián tiếp

For indirect standardization, you need a reference population with the number of deaths and number of population per stratum. In this example, we will be calculating rates for country A using country B as the reference population, as the standard\_pop\_clean reference population does not include number of deaths per stratum.

Để chuẩn hóa gián tiếp, chúng ta cần dân số tham chiếu có số lượng tử vong và dân số theo tầng. Trong ví dụ này, chúng ta sẽ tính tỷ suất của quốc gia A bằng cách *sử dụng quốc gia B làm dân số tham chiếu*, vì dân số tham chiếu trong bộ số liệu standard\_pop\_clean không có số lượng tử vong theo tầng.

Below, we first create the reference population from country B. Then, we pass mortality and population data for country A, combine it with the reference population, and pass it to the function phe\_isr(), to get indirectly standardized rates. Of course, you can do it also vice versa.

Trước tiên, chúng ta tạo dân số tham chiếu từ quốc gia B như bên dưới. Sau đó, chúng ta chuyển số liệu tử vong và dân số của quốc gia A, kết nối nó với dân số tham chiếu và chuyển nó vào hàm phe\_isr()để có được tỷ suất chuẩn hóa gián tiếp. Dĩ nhiên, chúng ta cũng có thể làm ngược lại.

Of note - in our example below, the reference population is provided as a separate data frame. In this case, we make sure that x =, n =, x\_ref = and n\_ref = vectors are all ordered by the same standardization category (stratum) values as that in our country-specific data frame, as records will be matched by position.

Lưu ý - trong ví dụ bên dưới, dân số tham chiếu được cung cấp dưới dạng một khung số liệu riêng rẽ. Trong trường hợp này, chúng ta cần đảm bảo rằng các véc tơ x =, n =, x\_ref = và n\_ref = tất cả được sắp đặt cùng giá trị với loại chuẩn hóa (theo đúng tầng) như được sắp đặt trong khung số liệu quốc gia, vì vị trí các bản ghi/các hàng phải khớp với nhau.

See the help with ?phr\_isr or the links in the References section for more information.

Xem trợ giúp bằng cách gõ ?phr\_dsr hoặc các liên kết trong phần Tài liệu tham khảo để biết thêm thông tin.

## Resources

## Nguồn

If you would like to see another reproducible example using **dsr** please see [this vignette](https://mran.microsoft.com/snapshot/2020-02-12/web/packages/dsr/vignettes/dsr.html)

For another example using **PHEindicatormethods**, please go to [this website](https://mran.microsoft.com/snapshot/2018-10-22/web/packages/PHEindicatormethods/vignettes/IntroductiontoPHEindicatormethods.html)

See the **PHEindicatormethods** [reference pdf file](https://cran.r-project.org/web/packages/PHEindicatormethods/PHEindicatormethods.pdf)