

# Package ‘PHEindicatormethods’

May 15, 2019

**Type** Package

**Version** 1.1.3

**Title** Common Public Health Statistics and their Confidence Intervals

**Description** Functions to calculate commonly used public health statistics and their confidence intervals using methods approved for use in the production of Public Health England indicators such as those presented via Fingertips (<<http://fingertips.phe.org.uk/>>). It provides functions for the generation of proportions, crude rates, means, directly standardised rates, indirectly standardised rates, standardised mortality ratios, slope and relative index of inequality and life expectancy.

Statistical methods are referenced in the following publications.

Breslow NE, Day NE (1987) <doi:10.1002/sim.4780080614>.

Dobson et al (1991) <doi:10.1002/sim.4780100317>.

Armitage P, Berry G (2002) <doi:10.1002/9780470773666>.

Wilson EB. (1927) <doi:10.1080/01621459.1927.10502953>.

Altman DG et al (2000, ISBN: 978-0-727-91375-3).

Chiang CL. (1968, ISBN: 978-0-882-75200-6).

Newell C. (1994, ISBN: 978-0-898-62451-9).

Eayres DP, Williams ES (2004) <doi:10.1136/jech.2003.009654>.

Silcocks PBS et al (2001) <doi:10.1136/jech.55.1.38>.

Low and Low (2004) <doi:10.1093/pubmed/fdh175>.

**BugReports** <https://github.com/PublicHealthEngland/PHEindicatormethods/issues>

**Depends** R (>= 3.1.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** dplyr, rlang, broom, tidyr, purrr, stats

**Suggests** knitr, readxl, rmarkdown, RODBC, testthat,

**RoxygenNote** 6.1.1

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Anderson Georgina [aut, cre],  
 Fox Sebastian [ctb],  
 Francis Matthew [ctb],  
 Fryers Paul [ctb],  
 Clegg Emma [ctb]

**Maintainer** Anderson Georgina <georgina.anderson@phe.gov.uk>

**Repository** CRAN

**Date/Publication** 2019-05-15 13:30:03 UTC

## R topics documented:

byars_lower . . . . .	2
byars_upper . . . . .	3
DSR_data . . . . .	4
esp2013 . . . . .	4
FindXValues . . . . .	5
LE_data . . . . .	6
na.zero . . . . .	6
PHEindicatormethods . . . . .	6
phe_dsr . . . . .	7
phe_isr . . . . .	8
phe_life_expectancy . . . . .	10
phe_mean . . . . .	13
phe_proportion . . . . .	14
phe_quantile . . . . .	16
phe_rate . . . . .	17
phe_sii . . . . .	19
phe_smr . . . . .	22
prevalence_data . . . . .	24
SimulationFunc . . . . .	24
wilson_lower . . . . .	25
wilson_upper . . . . .	26
<b>Index</b>	<b>28</b>

---

byars\_lower

*byars\_lower*

---

### Description

Calculates the lower confidence limits for observed numbers of events using Byar's method [1].

### Usage

```
byars_lower(x, confidence = 0.95)
```

**Arguments**

x	the observed numbers of events; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Value**

Returns lower confidence limits for observed numbers of events using Byar's method [1]

**Notes**

byars\_lower and [byars\\_upper](#) together return symmetric confidence intervals around counts, therefore for a specified confidence level,  $\alpha$ , the probability that, by chance, the lower limit returned will be above the true underlying value, is  $\alpha/2$ . If the confidence level is very close to 1 or the number of events is very small Byar's method is inaccurate and may return a negative number - in these cases an error is returned.

**References**

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

---



---

byars_upper	<i>byars_upper</i>
-------------	--------------------

---

**Description**

Calculates the upper confidence limits for observed numbers of events using Byar's method [1].

**Usage**

```
byars_upper(x, confidence = 0.95)
```

**Arguments**

x	the observed numbers of events; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Value**

Returns upper confidence limits for observed numbers of events using Byar's method [1]

**Notes**

`byars_lower` and `byars_upper` together return symmetric confidence intervals around counts, therefore for a specified confidence level,  $\alpha$ , the probability that, by chance, the upper limit returned will be below the true underlying value, is  $\alpha/2$ .

**References**

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

---



---

DSR\_data

*SII test datasets - DSR*

---

**Description**

A data table of dummy Directly Standardised Rates by deprivation quintiles

**Usage**

`data(DSR_data)`

**Format**

A data table

**Examples**

DSR\_data

---

esp2013

*European Standard Population 2013*

---

**Description**

A numeric vector containing nineteen 5-year age band populations making up the 2013 European Standard Population ordered from age 0-4, 5-9, 10-14 ... to ... 85-89, 90+. Sorted by increasing age band.

**Usage**

`esp2013`

**Format**

A numeric vector with 19 elements

**Value**

5000 5500 5500 5500 6000 6000 6500 7000 7000 7000 7000 6500 6000 5500 5000 4000 2500  
1500 1000

**Notes**

The 2013 European Standard Population is modelled and published by Eurostat [1] for use in the production of age-standardised rates. It uses the unweighted average 2010-based population projections of the European Union (x27) and European Free Trade Association (x4) countries for the period 2011-2030 broken down into 5-year age bands from age 0 - age 95+ with the 0-5 age band separated into age 0 and age 1-4. The version provided with this package combines the age 0 and age 1-4 populations into a single 0-4 age band and combines the 90-94 and 95+ populations into a single 90+ age band, giving 19 age bands in total.

**References**

[1] Eurostat Methodologies and Working Papers. Revision of the European Standard Population: Report of Eurostat's Taskforce, 2013.

<http://ec.europa.eu/eurostat/documents/3859598/5926869/KS-RA-13-028-EN.PDF/e713fa79-1add-44e8-b23d>

**Examples**

```
esp2013
```

---

FindXValues

*FindXValues*

---

**Description**

Calculates mid-points of cumulative population for each quantile

**Usage**

```
FindXValues(xvals, no_quantiles)
```

**Arguments**

xvals	field name in input dataset that contains the quantile populations; unquoted string; no default
no_quantiles	number of quantiles supplied in dataset for SII; integer; no default

LE\_data *SII test datasets - Life Expectancy*

---

**Description**

A data table of life expectancy data by area and deprivation decile

**Usage**

```
data(LE_data)
```

**Format**

A data table

**Examples**

```
LE_data
```

---

na.zero *Convert NAs to zeros using na.zero*

---

**Description**

converts NAs to zeros

**Usage**

```
na.zero(y)
```

**Arguments**

y           input vector

---

PHEindicatormethods *PHEindicatormethods: A package for performing standard statistics for public health indicators*

---

**Description**

PHEindicatormethods: A package for performing standard statistics for public health indicators

---

 phe\_dsr

*Calculate Directly Standardised Rates using phe\_dsr*


---

## Description

Calculates directly standardised rates with confidence limits using Byar's method [1] with Dobson method adjustment [2].

## Usage

```
phe_dsr(data, x, n, stdpop = esp2013, stdpoptype = "vector",
         type = "full", confidence = 0.95, multiplier = 1e+05)
```

## Arguments

data	data.frame containing the data to be standardised, pre-grouped if multiple DSRs required; unquoted string; no default
x	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
n	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
stdpop	the standard populations for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of stdpoptype; default = esp2013
stdpoptype	whether the stdpop has been specified as a vector or a field name from data; quoted string "field" or "vector"; default = "vector"
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

## Value

When type = "full", returns a tibble of total counts, total populations, directly standardised rates, lower confidence limits, upper confidence limits, confidence level, statistic and method for each grouping set

**Notes**

User MUST ensure that x, n and stdpop vectors are all ordered by the same standardisation category values as records will be matched by position.

For total counts  $\geq 10$  Byar's method [1] is applied using the [byars\\_lower](#) and [byars\\_upper](#) functions. When the total count is  $< 10$  DSRs are not reliable and will therefore not be calculated.

**References**

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

[2] Dobson A et al. Confidence intervals for weighted sums of Poisson parameters. Stat Med 1991;10:457-62.

**See Also**

Other PHEindicatormethods package functions: [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

**Examples**

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
                 pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_dsr(obs, pop)

## OR

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_dsr(obs, pop, type = "standard")
```

---

phe\_isr

*Calculate Indirectly Standardised Rates using phe\_isr*

---

**Description**

Calculates indirectly standardised rates with confidence limits using Byar's [1] or exact [2] CI method.



**Usage**

```
phe_isr(data, x, n, x_ref, n_ref, refpoptype = "vector", type = "full",
        confidence = 0.95, multiplier = 1e+05)
```

**Arguments**

data	data.frame containing the data to be standardised, pre-grouped if multiple ISRs required; unquoted string; no default
x	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
n	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
x_ref	the observed number of events in the reference population for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of refpoptype; no default
n_ref	the reference population for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of refpoptype; no default
refpoptype	whether x_ref and n_ref have been specified as vectors or a field name from data; quoted string "field" or "vector"; default = "vector"
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

**Value**

When type = "full", returns a tibble of observed events, expected events, indirectly standardised rate, lower confidence limit, upper confidence limit, confidence level, statistic and method for each grouping set

**Notes**

User MUST ensure that x, n, x\_ref and n\_ref vectors are all ordered by the same standardisation category values as records will be matched by position.

For numerators  $\geq 10$  Byar's method [1] is applied using the [byars\\_lower](#) and [byars\\_upper](#) functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

## References

- [1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.
- [2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

## See Also

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

## Examples

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
                 pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))

refdf <- data.frame(refcount = sample(200, 19, replace = TRUE),
                   refpop = sample(10000:20000, 19, replace = TRUE))

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_isr(obs, pop, refdf$refcount, refdf$refpop)

## OR

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_isr(obs, pop, refdf$refcount, refdf$refpop, type="standard", confidence=99.8)
```

---

phe\_life\_expectancy     *Calculate Life Expectancy using phe\_life\_expectancy*

---

## Description

Compute life expectancy for a given age, and its standard error

**Usage**

```

phe_life_expectancy(data, deaths, population, startage,
  age_contents = c(0L, 1L, 5L, 10L, 15L, 20L, 25L, 30L, 35L, 40L, 45L,
  50L, 55L, 60L, 65L, 70L, 75L, 80L, 85L, 90L), le_age = "all",
  type = "full", confidence = 0.95)

```

**Arguments**

data	data.frame or tbl containing the deaths and population data
deaths	field name from data containing the number of deaths within age band; unquoted string; no default
population	field name from data containing the population within age band; unquoted string; no default
startage	field name from data containing the age band; no default
age_contents	vector; describes the contents of startage in the ascending order. This vector is used to check whether each group in data contains the complete set of age bands for the calculation to occur. It is also used to reorder the data based on the startage field
le_age	the age band to return the life expectancy for. The default is "all", where the function returns the life expectancy values for all ages appended onto the input table. Any other value (or vector of values) must be age bands described by the age_contents input
type	type of output; can be "standard" or "full" (full contains added details on the calculation within the dataframe); quoted string; default full
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Details**

This function aligns with the methodology in Public Health England's [Life Expectancy Excel Tool](#). The function is for an abridged life table using 5 year age intervals with a final age interval of 90+. The table has been completed using the methods described by Chiang.[1],[2] This age structure and methodology is used by The Office for National Statistics to produce life expectancy at national and local authority level.[3]

This function includes an adjustment to the method for calculating the variance of the life expectancy estimate to include a term for the variance associated with the final age interval. In the Chiang method the variance of the life expectancy is the weighted sum of the variance of the probability of survival across all the age intervals. For the final age interval the probability of survival is, Chiang argues, zero and has zero variance. However, Silcocks et al argue[4] that in the case of the final age interval the life expectancy is dependent not on the probability of survival but on the mean length of survival ( $1/M < sub > omega < /sub >$ ). Therefore the variance associated with the final age interval depends on the age-specific mortality rate  $M < sub > omega < /sub >$ .

Life expectancy cannot be calculated if the person-years in any given age interval is zero. It will also not be calculated if the total person-years is less than 5,000 as this is considered to be the minimum size for robust calculation of life expectancy.[5] Zero death counts are not a problem, except for

the final age interval - there must be at least one death in the 90+ interval for the calculations to be possible.

The methodology used in this function, along with discussion of alternative options for life expectancy calculation for small areas, were described Eayres and Williams.[6]

### Author(s)

Sebastian Fox, <sebastian.fox@phe.gov.uk>

### References

- [1] Chiang CL. The Life Table and its Construction. In: Introduction to Stochastic Processes in Biostatistics. New York, John Wiley & Sons, 1968:189-214.
- [2] Newell C. Methods and Models in Demography. Chichester, John Wiley & Sons, 1994:63-81
- [3] Office for National Statistics Report. Life expectancy at birth by health and local authorities in the United Kingdom, 1998 to 2000 (3-year aggregate figures.) Health Statistics Quarterly 2002;13:83-90
- [4] Silcocks PBS, Jenner DA, Reza R. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. J Epidemiol Community Health 2001;55:38-43
- [5] Toson B, Baker A. Life expectancy at birth: methodological options for small populations. National Statistics Methodological Series No 33. HMSO 2003.
- [6] Eayres DP, Williams ES. Evaluation of methodologies for small area life expectancy estimation. J Epidemiol Community Health 2004;58:243-249

### See Also

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

### Examples

```
library(dplyr)

## A simple example
df <- data.frame(startage = c(0L, 1L, 5L, 10L, 15L, 20L, 25L, 30L, 35L, 40L, 45L, 50L, 55L,
                             60L, 65L, 70L, 75L, 80L, 85L, 90L),
                 pops = c(7060L, 35059L, 46974L, 48489L, 43219L, 38561L, 46009L, 57208L,
                          61435L, 55601L, 50209L, 56416L, 46411L, 39820L, 37978L,
                          37039L, 33288L, 23306L, 11936L, 11936L),
                 deaths = c(17L, 9L, 4L, 8L, 20L, 15L, 24L, 33L, 50L, 71L, 100L, 163L,
                            263L, 304L, 536L, 872L, 1390L, 1605L, 1936L, 1937L))
```

```

phe_life_expectancy(df, deaths, pops, startage)

## OR

phe_life_expectancy(df, deaths, pops, startage, le_age = c(5, 25), type = "standard")

## Unordered age bands example
df <- data.frame(startage = c("0", "1-4", "5-9", "10 - 14", "15 - 19", "20 - 24", "25 - 29",
                             "30 - 34", "35 - 39", "40 - 44", "45 - 49", "50 - 54",
                             "55 - 59", "60 - 64", "65 - 69", "75 - 79", "80 - 84",
                             "85 - 89", "90 +", "70 - 74"),
                 pops = c(7060L, 35059L, 46974L, 48489L, 43219L, 38561L, 46009L, 57208L,
                          61435L, 55601L, 50209L, 56416L, 46411L, 39820L, 37039L,
                          23306L, 11936L, 11936L, 37978L, 33288L),
                 deaths = c(17L, 9L, 4L, 8L, 20L, 15L, 24L, 33L, 50L, 71L, 100L, 163L,
                             263L, 304L, 872L, 1605L, 1936L, 1937L, 536L, 1390L))
phe_life_expectancy(df, deaths, pops, startage,
                   age_contents = c("0", "1-4", "5-9",
                                     "10 - 14", "15 - 19",
                                     "20 - 24", "25 - 29",
                                     "30 - 34", "35 - 39",
                                     "40 - 44", "45 - 49",
                                     "50 - 54", "55 - 59",
                                     "60 - 64", "65 - 69",
                                     "70 - 74", "75 - 79",
                                     "80 - 84", "85 - 89",
                                     "90 +"))

df <- data.frame(area = c(rep("Area 1", 20), rep("Area 2", 20)),
                 startage = rep(c(0L, 1L, 5L, 10L, 15L, 20L, 25L, 30L, 35L, 40L, 45L, 50L, 55L,
                                   60L, 65L, 70L, 75L, 80L, 85L, 90L), 2),
                 pops = rep(c(7060L, 35059L, 46974L, 48489L, 43219L, 38561L, 46009L, 57208L,
                              61435L, 55601L, 50209L, 56416L, 46411L, 39820L, 37978L,
                              37039L, 33288L, 23306L, 11936L, 11936L), 2),
                 deaths = rep(c(17L, 9L, 4L, 8L, 20L, 15L, 24L, 33L, 50L, 71L, 100L, 163L,
                               263L, 304L, 536L, 872L, 1390L, 1605L, 1936L, 1937L), 2))
df %>%
  group_by(area) %>%
  phe_life_expectancy(deaths, pops, startage)

```

---

phe\_mean

*Calculate Means using phe\_mean*


---

### Description

Calculates means with confidence limits using Student's t-distribution method.

### Usage

```
phe_mean(data, x, type = "full", confidence = 0.95)
```

**Arguments**

data	a data.frame containing the data to calculate means for, pre-grouped if multiple means required; unquoted string; no default
x	field name from data containing the values to calculate the means for; unquoted string; no default
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Value**

When type = "full", returns a data.frame of value\_sum, value\_count, stdev, value, lowercl, uppercl, confidence, statistic and method for each grouping set

**See Also**

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

**Examples**

```
library(dplyr)
df <- data.frame(values = c(30,40,50,60))
phe_mean(df, values)

## OR

df2 <- data.frame(area = rep(c("Area1", "Area2"),each=3),
                  values = c(20,30,40,200,300,400)) %>%
  group_by(area)
phe_mean(df2, values)
phe_mean(df2, values, type="standard", confidence=0.998)
```

---

phe\_proportion

*Calculate Proportions using phe\_proportion*

---

**Description**

Calculates proportions with confidence limits using Wilson Score method [1,2].

**Usage**

```
phe_proportion(data, x, n, type = "full", confidence = 0.95,
               multiplier = 1)
```

**Arguments**

data	a data.frame containing the data to calculate proportions for, pre-grouped if proportions required for group aggregates; unquoted string; no default
x	field name from data containing the observed numbers of cases in the sample meeting the required condition (the numerator for the proportion); unquoted string; no default
n	field name from data containing the number of cases in the sample (the denominator for the proportion); unquoted string; no default
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100 = percentage); numeric; default 1

**Value**

When type = "full", returns the original data.frame with the following appended: proportion, lower confidence limit, upper confidence limit, confidence level, statistic and method

**Notes**

Wilson Score method [1,2] is applied using the [wilson\\_lower](#) and [wilson\\_upper](#) functions.

The percentage argument was deprecated in v1\_1\_0, please use multiplier argument instead

**References**

- [1] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc; 1927; 22. Pg 209 to 212.
- [2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.

**See Also**

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

**Examples**

```
# ungrouped data frame
df <- data.frame(area = rep(c("Area1", "Area2", "Area3", "Area4"), each=3),
                 numerator = c(NA, 82, 9, 48, 6500, 8200, 10000, 10000, 8, 7, 750, 900),
                 denominator = rep(c(100, 10000, 10000, 10000), each=3))

phe_proportion(df, numerator, denominator)
```

```

phe_proportion(df, numerator, denominator, confidence=99.8)
phe_proportion(df, numerator, denominator, type="standard")

# grouped data frame
library(dplyr)
dfg <- df %>% group_by(area)
phe_proportion(dfg, numerator, denominator, multiplier=100)

```

---

phe\_quantile

*Assign Quantiles using phe\_quantile*

---

### Description

Assigns small areas to quantiles based on numeric data rankings.

### Usage

```

phe_quantile(data, values, highergeog = NULL, nquantiles = 10L,
  invert = TRUE, inverttype = "logical", type = "full")

```

### Arguments

data	a data.frame containing the base (and optionally higher) geography data and quantitative data for assigning quantiles, pre-grouped if quantiles required for breakdowns other than the defined higher geographies; unquoted string; no default
values	field name from data containing the numeric values for each base geography area to rank data by and assign quantiles from; unquoted string; no default
highergeog	field name from data containing the higher geographies to assign separate quantile categories within if required; unquoted string; default = NULL
nquantiles	the number of quantiles to assign per higher geography; numeric; default=10L
invert	whether the quantiles should be directly (FALSE) or inversely (TRUE) related to the numerical value order; unquoted string referencing logical values as either a single logical value or field name from data depending on value of inverttype; default = TRUE (ie highest values assigned to quantile 1)
inverttype	whether the invert argument has been specified as a single logical value or a field name from data; quoted string "field" or "logical"; default = "logical"
type	defines whether to include metadata columns in output to reference the arguments passed; can be "standard" or "full"; quoted string; default = "full"

### Value

When type = "full", returns the original data.frame with quantile (quantile value), nquantiles (number of quantiles requested) and invert (indicating direction of quantile assignment) fields appended.



**Notes**

See [PHE Technical Guide - Assigning Deprivation Quintiles](https://fingertips.phe.org.uk/profile/guidance) for methodology. In particular, note that this function strictly applies the algorithm defined but some manual review, and potentially adjustment, is advised in some cases where multiple small areas with equal rank fall across a natural quantile boundary.

**See Also**

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_rate](#), [phe\\_sii](#), [phe\\_smr](#)

**Examples**

```
df <- data.frame(region = as.character(rep(c("Region1", "Region2", "Region3", "Region4"), each=250)),
  smallarea = as.character(paste0("Area", seq_along(1:1000))),
  vals = as.numeric(sample(200, 1000, replace = TRUE)),
  stringsAsFactors=FALSE)

# assign small areas to deciles within regions - method 1: assign grouping set within function
phe_quantile(df, vals, region)

# assign small area to deciles within regions - method 2: pre-group input dataframe
library(dplyr)
df_grp <- df %>% group_by(region)
phe_quantile(df_grp, vals)

# assign smallareas to decile across whole data frame (ignoring region)
phe_quantile(df, vals)

# assign smallareas to quintiles within regions, where high val = lowest quantile
phe_quantile(df, vals, region, invert=FALSE)
```

---

phe\_rate

*Calculate Rates using phe\_rate*

---

**Description**

Calculates rates with confidence limits using Byar's [1] or exact [2] CI method.

**Usage**

```
phe_rate(data, x, n, type = "full", confidence = 0.95,
  multiplier = 1e+05)
```

**Arguments**

data	the data.frame containing the data to calculate rates for, pre-grouped if proportions required for group aggregates; unquoted string; no default
x	field name from data containing the rate numerators (eg observed number of events); unquoted string; no default
n	field name from data containing the rate denominators (eg populations); unquoted string; no default
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

**Value**

When type = "full", returns the original data.frame with the following appended: rate, lower confidence limit, upper confidence limit, confidence level, statistic and method

**Notes**

For numerators  $\geq 10$  Byar's method [1] is applied using the [byars\\_lower](#) and [byars\\_upper](#) functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

**References**

- [1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.
- [2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

**See Also**

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_sii](#), [phe\\_smr](#)

**Examples**

```
# ungrouped data frame
df <- data.frame(area = rep(c("Area1", "Area2", "Area3", "Area4"), each=3),
                 obs = c(NA, 82, 9, 48, 6500, 8200, 10000, 10000, 8, 7, 750, 900),
                 pop = rep(c(100, 10000, 10000, 10000), each=3))

phe_rate(df, obs, pop)
phe_rate(df, obs, pop, type="standard")
```

```

phe_rate(df, obs, pop, confidence=99.8, multiplier=100)

# grouped data frame
library(dplyr)
dfg <- df %>% group_by(area)
phe_rate(dfg, obs, pop)

```

---

phe\_sii

*Calculate Slope Index of Inequality using phe\_sii*

---

### Description

phe\_sii returns the slope index of inequality (SII) statistic for each subgroup of the inputted dataframe, with lower and upper confidence limits based on the specified confidence. The Relative Index of Inequality (RII) can also be returned via an optional argument.

### Usage

```

phe_sii(data, quantile, population, x = NULL, value = NULL,
        value_type = 0, lower_cl = NULL, upper_cl = NULL, se = NULL,
        multiplier = 1, repetitions = 1e+05, confidence = 0.95,
        rii = FALSE, reliability_stat = FALSE, type = "full")

```

### Arguments

data	data.frame containing the data to calculate slope index of inequality from, pre-grouped if multiple DSRs required; unquoted string; no default
quantile	field name within data that contains the quantile label (e.g. decile). The number of quantiles should be between 5 and 100; unquoted string; no default
population	field name within data that contains the quantile populations (ie, denominator). Non-zero populations are required for all quantiles to calculate SII for an area; unquoted string; no default
x	(for indicators that are proportions) field name within data that contains the members of the population with the attribute of interest (ie, numerator). This will be divided by population to calculate a proportion as the indicator value (if value field is not provided); unquoted string; no default
value	field name within data that contains the indicator value (this does not need to be supplied for proportions if count and population are given); unquoted string; no default
value_type	indicates the indicator type (1 = rate, 2 = proportion, 0 = other); integer; default 0
lower_cl	field name within data that contains 95% lower confidence limit of indicator value (to calculate standard error of indicator value). This field is needed if the se field is not supplied; unquoted string; no default

upper_cl	field name within data that contains 95% upper confidence limit of indicator value (to calculate standard error of indicator value). This field is needed if the se field is not supplied; unquoted string; no default
se	field name within data that contains the standard error of the indicator value. If not supplied, this will be calculated from the 95% lower and upper confidence limits (i.e. one or the other of these fields must be supplied); unquoted string; no default
multiplier	factor to multiply the SII and SII confidence limits by (e.g. set to 100 to return prevalences on a percentage scale between 0 and 100). If the multiplier is negative, the inverse of the RII is taken to account for the change in polarity; numeric; default 1;
repetitions	number of random samples to perform to return confidence interval of SII; numeric; default 100,000
confidence	confidence level used to calculate the lower and upper confidence limits of SII; numeric between 0.5 and 0.9999 or 50 and 99.99; default 0.95
rii	option to return the Relative Index of Inequality (RII) with associated confidence limits as well as the SII; logical; default FALSE
reliability_stat	option to carry out the SII confidence interval simulation 10 times instead of once and return the Mean Average Difference between the first and subsequent samples (as a measure of the amount of variation). Warning: this will significantly increase run time of the function and should first be tested on a small number of repetitions; logical; default FALSE
type	"full" output includes columns in the output dataset specifying the parameters the user has input to the function (value_type, multiplier, CI_confidence, CI_method); character string either "full" or "standard"; default "full"

## Details

The SII and RII are two measures of health inequality. They show the relation between the level of health or frequency of a health problem in different population groups and the ranking of these groups on the social scale.

The input dataframe should be grouped before passing to the function if an SII/RII for each subgroup is required, and quantiles ordered from least to most advantaged.

## Value

The SII with lower and upper confidence limits for each subgroup of the inputted data.frame.

## Calculation

The SII is calculated using linear regression [1]. To allow for differences in population size between quantiles (e.g. deprivation deciles), each is given a rank score (or relative rank) based on the midpoint of its range in the cumulative distribution of the total area population. The quantiles are first ordered (e.g from 1 most deprived to 10 least deprived for deprivation deciles). If quantile 1 then contains 12% of the total population, its relative rank is  $0.12/2=0.6$ . If quantile 2 includes 10% of the population, its relative rank is  $0.12+(0.10/2)=0.17$ . A square root transformation is applied

to the regression to account for heteroskedasticity (the tendency for the variances of the quantile values to be related to the size of the values, ie larger values will tend to have larger variances). A regression model is fitted to the transformed data:  $Y * \sqrt{a} = \sqrt{a} + b * \sqrt{a}$ , where Y is the value of the indicator for the quantile, a is the proportion of the total population in the quantile and b is the relative rank.

The SII is the gradient of the resulting fitted line, and could be positive or negative according to the indicator polarity. Since the relative ranks, by definition, range from 0 to 1, the SII is the difference between the fitted value at  $x=1$  and  $x=0$ .

The RII is the ratio of the fitted value at  $x=1, Y1$  and the fitted value at  $x=0, Y0$ . which can be calculated as:  $RII = (Y0 + SII)/Y0$

### Function arguments

The indicator type can be specified via the `value_type` parameter. Transformations can be applied to the indicator value and its confidence limits before calculating the standard error in cases where the confidence interval around the indicator value is likely to be non-symmetric. This is a log transformation for rates, and logit for proportions.

If the standard error is supplied directly to the function from the input dataset, this is used instead of calculating one from the indicator confidence limits.

### Warning

The SII calculation assumes a linear relationship between indicator value and quantile, and small populations within quantiles can make it unstable.

This function does not include checks for linearity or stability; it is the user's responsibility to ensure the input data is suitable for the SII calculation.

### Author(s)

Emma Clegg, <emma.clegg@phe.gov.uk>

### References

[1] Low A & Low A. Measuring the gap: quantifying and comparing local health inequalities. *Journal of Public Health*; 2004;26:388-395.

### See Also

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_smr](#)

phe\_smr

*Calculate Standardised Mortality Ratios using phe\_smr***Description**

Calculates standard mortality ratios (or indirectly standardised ratios) with confidence limits using Byar's [1] or exact [2] CI method.

**Usage**

```
phe_smr(data, x, n, x_ref, n_ref, refpoptype = "vector", type = "full",
         confidence = 0.95, refvalue = 1)
```

**Arguments**

data	data.frame containing the data to be standardised, pre-grouped if multiple SMRs required; unquoted string; no default
x	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
n	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
x_ref	the observed number of events in the reference population for each standardisation category (eg age band); unquoted numeric vector or field name from data depending on value of refpoptype; no default
n_ref	the reference population for each standardisation category (eg age band); unquoted numeric vector or field name from data depending on value of refpoptype; no default
refpoptype	whether x_ref and n_ref have been specified as vectors or a field name from data; quoted string "field" or "vector"; default = "vector"
type	defines the data and metadata columns to be included in output; can be "value", "lower", "upper", "standard" (for all data) or "full" (for all data and metadata); quoted string; default = "full"
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
refvalue	the standardised reference ratio, numeric, default = 1

**Value**

When type = "full", returns a tibble of observed events, expected events, standardised mortality ratios, lower confidence limits, upper confidence limits, confidence level, statistic and method for each grouping set

## Notes

User MUST ensure that x, n, x\_ref and n\_ref vectors are all ordered by the same standardisation category values as records will be matched by position.

For numerators  $\geq 10$  Byar's method [1] is applied using the [byars\\_lower](#) and [byars\\_upper](#) functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

## References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

[2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

## See Also

Other PHEindicatormethods package functions: [phe\\_dsr](#), [phe\\_isr](#), [phe\\_life\\_expectancy](#), [phe\\_mean](#), [phe\\_proportion](#), [phe\\_quantile](#), [phe\\_rate](#), [phe\\_sii](#)

## Examples

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
                 pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))

refdf <- data.frame(refcount = sample(200, 19, replace = TRUE),
                   refpop = sample(10000:20000, 19, replace = TRUE))

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_smr(obs, pop, refdf$refcount, refdf$refpop, type="standard")

## OR

df %>%
  group_by(indicatorid, year, sex) %>%
  phe_smr(obs, pop, refdf$refcount, refdf$refpop, confidence=99.8, refvalue=100)
```

---

```
prevalence_data      SII test datasets - Prevalence
```

---

**Description**

A data table of example prevalence data by area and deprivation decile

**Usage**

```
data(prevalence_data)
```

**Format**

A data table

**Examples**

```
prevalence_data
```

---

```
SimulationFunc      SimulationFunc
```

---

**Description**

Function to simulate SII range through random sampling of the indicator value for each quantile, based on the associated mean and standard error

**Usage**

```
SimulationFunc(data, value, value_type = 0, se, repeats = 1e+05,
  confidence = 0.95, sqrt_a, b_sqrt_a, rii = FALSE,
  reliability_stat = FALSE)
```

**Arguments**

data	data.frame containing the data to calculate SII confidence limits from; unquoted string; no default
value	field name within data that contains the indicator value; unquoted string; no default
value_type	indicates the indicator type (1 = rate, 2 = proportion, 0 = other); integer; default 0
se	field name within data that contains the standard error of the indicator value; unquoted string; no default



repeats	number of random samples to perform to return confidence interval of SII; numeric; default 100,000
confidence	confidence level used to calculate the lower and upper confidence limits of SII; numeric between 0.5 and 0.9999 or 50 and 99.99; default 0.95
sqrt_a	field name within dataset containing square root of a values; unquoted string; no default
b_sqrt_a	field name within dataset containing square root of a values multiplied by b values; unquoted string; no default
rii	option to return the Relative Index of Inequality (RII) with associated confidence limits as well as the SII; logical; default FALSE
reliability_stat	option to carry out the SII confidence interval simulation 10 times instead of once and return the Mean Average Difference between the first and subsequent samples (as a measure of the amount of variation); logical; default FALSE

**Value**

returns lower and upper SII confidence limits according to user specified confidence

---

wilson_lower	<i>wilson_lower</i>
--------------	---------------------

---

**Description**

Calculates lower confidence limits for observed numbers of events using the Wilson Score method [1,2].

**Usage**

```
wilson_lower(x, n, confidence = 0.95)
```

**Arguments**

x	the observed numbers of cases in the samples meeting the required condition; numeric vector; no default
n	the numbers of cases in the samples; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Value**

Returns lower confidence limits for observed numbers of events using the Wilson Score method [1,2]

**Notes**

wilson\_lower and [wilson\\_upper](#) together return symmetric confidence intervals, therefore for a specified confidence level,  $\alpha$ , the probability that, by chance, the lower limit returned will be above the true underlying value, is  $\alpha/2$ .#

**References**

- [1] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc; 1927; 22. Pg 209 to 212.
- [2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.
- 

---

wilson\_upper

*wilson\_upper*

---

**Description**

Calculates upper confidence limits for observed numbers of events using the Wilson Score method [1,2].

**Usage**

```
wilson_upper(x, n, confidence = 0.95)
```

**Arguments**

x	the observed numbers of cases in the samples meeting the required condition; numeric vector; no default
n	the numbers of cases in the samples; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

**Value**

Returns upper confidence limits for observed numbers of events using the Wilson Score method [1,2]

**Notes**

[wilson\\_lower](#) and [wilson\\_upper](#) together return symmetric confidence intervals, therefore for a specified confidence level,  $\alpha$ , the probability that, by chance, the upper limit returned will be below the true underlying value, is  $\alpha/2$ .#

## References

- [1] Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*; 1927; 22. Pg 209 to 212.
  - [2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). *Statistics with confidence* (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.
-

# Index

## \*Topic **datasets**

DSR\_data, 4

esp2013, 4

LE\_data, 6

prevalence\_data, 24

byars\_lower, 2, 4, 8, 9, 18, 23

byars\_upper, 3, 3, 8, 9, 18, 23

DSR\_data, 4

esp2013, 4

FindXValues, 5

LE\_data, 6

na.zero, 6

phe\_dsr, 7, 10, 12, 14, 15, 17, 18, 21, 23

phe\_isr, 8, 8, 12, 14, 15, 17, 18, 21, 23

phe\_life\_expectancy, 8, 10, 10, 14, 15, 17,  
18, 21, 23

phe\_mean, 8, 10, 12, 13, 15, 17, 18, 21, 23

phe\_proportion, 8, 10, 12, 14, 14, 17, 18, 21,  
23

phe\_quantile, 8, 10, 12, 14, 15, 16, 18, 21, 23

phe\_rate, 8, 10, 12, 14, 15, 17, 17, 21, 23

phe\_sii, 8, 10, 12, 14, 15, 17, 18, 19, 23

phe\_smr, 8, 10, 12, 14, 15, 17, 18, 21, 22

PHEindicatormethods, 6

PHEindicatormethods-package  
(PHEindicatormethods), 6

prevalence\_data, 24

SimulationFunc, 24

wilson\_lower, 15, 25, 26

wilson\_upper, 15, 26, 26