Funded by Wellcome

Standards for Official Statistics on Climate-Health Interactions (SOSCHI)

Suicides attributed to extreme heat: methodology

Alpha Phase document

Publication date: 12 November 2024

We welcome users' views and expertise on the alpha version of the statistical framework to further develop our work. Please email us at climate.health@ons.gov.uk.

Contents

Introduction to the SOSCHI project

The impacts on health of rising temperatures, wildfires, extreme weather events and other direct and indirect effects of climate change are a major global concern. The most significant hazards and their impacts differ between countries and regions, as do the possibilities and priorities for climate change adaptation. National and local governments and other stakeholders need to have regular, reliable and comparable data to monitor climate impacts and inform adaptation strategies, based on a transparent and globally generalisable statistical framework. The SOSCHI project, led by the UK Office for National Statistics and funded by Wellcome, is developing a framework of indicators based on state-of-the-art statistical methods to measure climate-related health risks. To support global reporting and monitoring, we are also developing a knowledge-sharing platform, open-source tools and R code. Our findings will also help highlight data gaps and help set the agenda for future improvement of data sources and methods.

Project partners

African Institute for Mathematical Sciences, Kigali, Rwanda
Cochrane Planetary Health Thematic Group, University of Alberta, Edmonton, Canada
Office for National Statistics, Newport, United Kingdom
Regional Institute for Population Studies, University of Ghana, Accra, Ghana
UK Health Security Agency, London, United Kingdom
United Nations Global Platform, New York, United States of America

The project is supported by Wellcome grant no. 224682/Z/21/Z

Document information

Document title	Standards for Official Statistics on Climate-Health Interactions (SOSCHI): Suicides attributed to extreme heat: methodology
Document location	Zenodo: https://zenodo.org/communities/soschi
Date and version	12 November 2024 - published version 1.0
Authors	Matt Pearce[1], Ellie Watkins[1] [1]ONS
DOI	https://doi.org/10.5281/zenodo.14050225
Preferred citation	Pearce M, Watkins E. Standards for Official Statistics on Climate-Health Interactions (SOSCHI): Suicides attributed to extreme heat: methodology. Zenodo; 2024. doi: 10.5281/zenodo.14050225.
Related topic documents	Mental Health: Introduction (10.5281/zenodo.14049912) Suicides attributed to extreme heat: metadata sheet (10.5281/zenodo.14051476)
Contact	Climate and Health Team, Health and International Division, Office for National Statistics, Newport NP10 8XG, United Kingdom Climate.health@ons.gov.uk
Copyright	This is an open access document published under a UK Open Government Licence 3.0 which is compatible with the Creative Commons CC-BY-SA 4.0 licence. You are free to copy, publish, distribute and adapt the information, provided you acknowledge the source and link to this publication or the ONS SOSCHI project home page. Statistical data used in the preparation of this document may be confidential under UK or other national laws. Enquiries concerning the availability of data can be made to the project team at climate.health@ons.gov.uk. Institutional logos are the property of the relevant institutions.

Important notes

This document has been published as part of the alpha version of the SOSCHI statistical framework. Therefore, this should be read as a draft document/ document in progress, which does not necessarily represent the final state of the framework. We welcome users' views and expertise to further develop our work. Please email us at climate.health@ons.gov.uk.

Statistics within this document are considered statistics in development. See our Guide to official statistics in development for more information about what this means.

1. Reproducible code

We are in the process of developing an open-source R code library for the calculation of this indicator. Future updates will include a link to this on GitHub, so that it can be adapted to users' own needs and data requirements.

2. Overview

2.1 Purpose

The indicator assesses the short-term impact of extreme heat on suicides by providing estimates of excess risk of suicide attributable to extreme outdoor temperature. This can give insights into how more frequent and intense hot days may be impacting mental health and how this varies regionally. Disaggregation of results can also help provide evidence on risk for the most vulnerable groups in society and therefore inform local adaptation needs.

While suicide is not a direct measure of the incidence or prevalence of mental health conditions, it is a useful proxy for some types of mental health challenges which may be exacerbated by extreme heat.

There are three main outcomes calculated for this indicator:

- Relative risk for the temperature-suicide association for each region
- Number of suicides attributable to extreme heat for each region
- Suicides per 100,000 population attributable to extreme heat for each region

2.2 Concepts & definitions

Attributable Number (AN):

The estimated number of suicides that are attributable to extreme heat, that is, that would not have occurred if there had been no exposure to that heat.

Attributable Rate (AR):

The estimated number of suicides per 100,000 population that are attributable to extreme heat, that is, that would not have occurred if there had been no exposure to that heat.

Relative Risk (RR):

The likelihood of an individual committing suicide during, or shortly after, exposure to a certain daily temperature. When a temperature has an RR of 1, this means there is neither an increase nor a decrease in the likelihood of the individual committing suicide during, or shortly after, exposure to that temperature. An RR greater than 1 indicates an increased likelihood of suicide compared with the reference temperature.

Extreme heat:

There is no global standardised definition, which this indicator draws on due to the wide variation of temperatures observed between countries. Instead, the indicator will determine extreme heat using percentiles.

Case-crossover:

A design commonly used in environmental epidemiology. It compares differences in exposure on case days (on which the health event occurred) with that on control days (usually within strata defined by day of the week, month, and year) that might explain differences in the number of cases[1].

Distributed lag non-linear model (DLNM):

A model often used to describe an exposure-time-response relationship where the exposure effect is non-linear and can have delayed effects[2].

Spline function:

A function defined piecewise by polynomials, where each polynomial is joined to another at a fixed point, a knot.

Suicide:

Suicide is the common term for death due to intentional self-harm, where the deceased intended to end their life. The requirements for certification of a death as being due to intentional self-harm vary between countries, with the determination potentially being made by a judicial investigation (coroner), by the attending medical practitioner, or by the police or another authority.

Quasi-Poisson regression:

A quasi-Poisson distribution is used for count data, where the variance of the data is assumed to be a linear function of the mean. This distribution tends to be used when a count variable is over-dispersed.

3. Data and variables

3.1 Data sources

Official data sources are used where possible:

- Daily mortality data on suicides with demographic information: from national administrative records (civil registration)
- Daily temperature and other climate data: from governmental meteorological offices

The following high-quality global datasets are given as proxy data sources in case of missing administrative data:

- ERA5[3]: for temperature and other climate data
    - It is a source of global ground level temperature data which provides hourly estimates of many atmospheric, land and oceanic climate variables. The data are of high accuracy, even though they are partially modelled. ERA5 provides reanalysis data which combines observations with modelled data to have spatial coverage in the grid. The is especially useful where ground data is patchy.
- Air Quality Modelling Products | 4 Earth Intelligence[4]: for air pollution data
    - Provides satellite-derived modelled air quality data at 20m resolution globally.

For each data used, it is important to note how the data are gathered, and thus, their limitations and accuracy.

Deaths due to ‘intentional self-harm’ are normally indicated in mortality data by appropriate underlying cause of death codes in the International Classification of Diseases (ICD) and often also by a specific manner of death and/or legal verdict code. In some countries deaths coded as ‘event of undetermined intent’ are also included in national suicide statistics based on evidence that these are often undetected suicides; the appropriate national practice should be followed for this indicator.

3.2 Data template

The data template (Table 1) provides a list of the variables needed for the indicator, identifying those which are mandatory for the method to work, and those which are optional. Their format is also provided to align with the data requirements for the indicator code provided in GitHub. The codes and variable groupings are, where appropriate, aligned with the Inter-agency and Expert Group on Sustainable Development Goal Indicators' Statistical Data and Metadata Exchange[5].

Table 1: Data template

Variable name	Description	Format	Priority	Notes
healthoutcome	Daily number of suicides	Integer	Mandatory	ICD-10 codes: • X60-X84 Intentional self-harm • Y10-Y34 Undetermined intent ICD-11 codes: • PB80-PD3Z Intentional self-harm • PF40-PH8Z Undetermined intent
dod	Date of death	Date	Mandatory	
tmax	Maximum daily temperature (°C)	Numeric	Mandatory	
humidity	Daily relative humidity (%)	Numeric	Optional	Controlled for as a possible confounder
sunshine	Duration of daily bright sunshine (hours)	Numeric	Optional	Controlled for as a possible confounder
precipitation	Daily total precipitation (mm)	Numeric	Optional	Controlled for as a possible confounder
airpollution			Optional	*To be defined*
regnames	Region names	Character	Optional	For geographical disaggregation
More to be added on disaggregation and population totals for attributable rates				

The method uses daily data for measures of possible confounders between the temperature and suicide relationship, namely, for precipitation, air pollution, humidity, and sunshine hours. These can all impact psychological state and its association with temperature.

The data should also contain variables for the number of suicides and the population of every disaggregation. For example, if we wish to disaggregate by sex, we need variables containing the number of male suicides, the size of the male population, the number of female suicides and the size of the female population.

Table 2 describes the variables by which to disaggregate the indicator's output. Within these groups there are particularly vulnerable populations that extreme heat can disproportionately affect and therefore important to understand the differences in trends. Whether this is achievable will depend on data availability and may in some cases require data linkage to other sources.

The Framework Introduction document describes the recommended classifications to use for each disaggregation. The Relative Risk, Attributable Number and Attributable Rate are calculated for each grouping to identify changes in risk and suicide between groups. This can be done by changing the dependent and population variables for the analysis to be the suicides and the population size for the relevant group.

Table 2: Indicator disaggregation

Disaggregation	Application
Age	Using the classification described in the Framework Introduction for age at time of death
Sex	Using the classification described in the Framework Introduction for sex of the deceased
Urbanisation	Using the classification described in the Framework Introduction for the degree of urbanisation of the reference area of the deceased's place of death or residence
Socio-economic group	Using the classification described in the Framework Introduction for the socio-economic group of the deceased or for the deceased's place of residence
Ethnic group	Using the classification described in the Framework Introduction for the deceased's ethnicity
Occupation	Using the classification described in the Framework Introduction for the deceased occupation

3.3 Data preparation

Care should be taken to count suicide deaths according to national practice, which may vary because of limitations in the civil registration data, the range of ICD codes used, and for non-technical reasons (see under Data Limitations).

All data, the suicide, demographic, and climate data, are joined to create a daily timeseries dataset by region. Each variable is checked for duplicates, errors and missing entries and is formatted as described in the data template.

To deal with missing climate data, the data can be imputed. Dropping observations is not appropriate as it would remove a point in the timeseries. An appropriate imputation method is interpolation for each region as it takes autocorrelation of timeseries data into account. If an optional variable has very few or no observations for a region, this variable can be dropped from the equation or imputed with the daily mean of the non-missing regional values.

3.4 Data limitations

Reporting of suicide as a cause of death varies globally not only because of national differences in legal systems for the investigation and certification of sudden deaths, but also because the subject of suicide is surrounded by cultural and religious issues. It is important to make comparisons of suicide in different periods only within the same country, as comparisons between countries are likely to be affected by such legal and cultural differences. Care should also be taken to account for any time-trend differences within countries that could be explained by changes in the legal system or reporting practices.

There is often a delay between the occurrence of a death and its recording in the national civil registration system. Providing the most timely statistics can compromise the completeness of the mortality data, as the most recent deaths to occur may not yet have been registered. This is especially likely to be the case with suicides, which like other unexpected or violent deaths require a legal investigation in many countries. A balance needs to be found between timeliness and completeness.

As this analysis specifically explores the relationship between suicide and temperature, it is necessary to use daily counts (or other short time periods) of suicides. These numbers can be low because suicide is relatively infrequent as a cause of death in the population as a whole. The statistical power of the analysis and width of the confidence intervals should be considered in interpretation of the results, particularly if numbers of deaths per period are further reduced by disaggregation.

The available local temperature data may be based on estimates taken from nearby areas, thus reducing the accuracy of such data. If the temperature data contain estimates, this should be recognised in the reporting of the output.

The method includes adjustment for relatively short-term and proximal confounding factors such as sunshine, which are known to have independent effects on suicide rates. It does not specifically allow for broader economic and social factors which may affect the rate of suicide over time, for example because of loss of livelihoods. These should have limited effect on the indicator because of the case-crossover design, but should nevertheless be considered when interpreting the results.

4. Analytical methodology

4.1 Introduction to methodology

The statistical approach for this indicator is based on a combination of two methods. The first is a case-crossover design which is widely used in environmental epidemiology. It compares differences in exposure on case days with that on control days, within strata defined by day of the week, month, and year, that might explain differences in the number of cases.

The second has been adapted from a method developed by Gasparrini et al.[2] which uses distributed lag non-linear models (DLNMs) to describe the exposure-time-response relationship where the exposure effect is non-linear and can have delayed effects.

This approach was chosen for this indicator because it aligns with recommendations made by Massazza et al.[6] on quantitative methods for climate change and mental health research.

4.2 Descriptive statistics

Certain basic descriptive statistics for the analysis' variables provide insights into the quality and quantity of the data.

- The **minimum and maximum** of numerical variables. To indicate the credibility of the variable's range.
- The **proportion of missing entries**. To indicate whether imputation is needed to complete the timeseries or whether too many data are missing for a variable to be informative.
- **Correlation coefficients**

    - **between daily temperature and suicides**. To provide insight into the strength of the relationship between suicides and temperature.
    - **between potential confounders and between effect modifiers**. To provide insight into how the indicator's outputs can be interpreted.
- **Timeseries plots with moving averages** for deaths and the independent variables. To show high-level patterns and long-term trends.

4.3 Approach to analysis

Covering:

- Step-by-step description of the analysis
- Why a particular level of detail was chosen
- What methodological decisions have been made

For each region, the RR of the temperature-suicide association is determined, and the AN and the AR of suicides from extreme heat in a year are calculated. There are three main steps to reach these outputs as outlined below.

Stage 1: Time-stratified case-crossover analysis with DLNM

This first stage is used to examine the short-term association between temperature and suicide in each region. Initially, the case-crossover design is established by creating non-overlapping strata are defined by day of the week, month and year. By matching case days and control days across this reference window long-term trends, seasonality of unmeasured time-varying confounders and the influence of the day of the week are adjusted for[1]. Strata with no suicide events are excluded from the analysis to increase the statistical power.

A distributed lag non-linear model (DLNM) is then applied to determine the short-term association between temperature and suicide, reported as RR. A conditional quasi-Poisson regression is fitted to account for over-dispersion.

Possible confounders are controlled for at this stage. Sensitivity analysis will indicate which independent variables to include in the final model, as described in section 3.4.

Specifically, a cross-basis matrix is used, which is the parameterisation of a composite of two functions: a temperature-response and a lag-response function. The cross-basis uses natural cubic B-splines and 4 degrees of freedom for both temperature and lag. The lag-response function also uses a lag of 3 days. This is guided by existing literature[7].

The temperature-mortality association is subsequently estimated for each region by calculating a predicted RR with the fitted DLNM where the reference value is taken to be the average of the timeseries' daily temperatures in that region.

Stage 2: Meta-analysis

Detail on the steps here are to be added once more testing has been undertaken on this part of the method.

Stage 3: Calculating ANs and ARs

Detail on the steps here are to be added once more testing has been undertaken on this part of the method.

4.4 Sensitivity analysis

Test the following modelling choices, detailed in Table 3, through sensitivity analysis:

Table 3: Sensitivity analysis

Modelling choice	Description	How
Independent variables in the DLNM formula	Test whether each of the following possible independent variables should be included in the DLNM formula in stage 1, if data is available: • Humidity • Air pollution • Rainfall • Sunshine hours	The QAIC of the meta-analytical model fitted in stage 2 can be compared between combinations of independent variables included in the DLNM. The combination of independent variables with the lowest QAIC is chosen as the final input for the DLNM.

4.5 Communication & Presentation

The results from this indicator are best presented by a combination of statistical tables and graphs. Table 4 describes how each output from this indicator can be presented:

Table 4: Presentation and interpretation of analytical outputs

Output	Presentation
The RR of suicide by temperature	Present the RRs in statistical tables. Their respective estimates should be included for each region and for each disaggregation for the whole country. Line graphs can be used to represent differences in RR patterns between regions and disaggregation groups: • A line graph of the RR by region.

	• A line graph of the RR by disaggregation group. Headline results could be quoted as: • "Very high temperatures had higher suicide risk, with temperatures above *xx* degrees Celsius representing the greatest risk, in *reference area* in years *yy* to *yy*." • "The highest mortality risk was in *reference area* for temperatures exceeding *xx* degrees Celsius, where mortality risk was *xx* times the risk at optimal temperatures." *Example charts to be added*
The AN and AR of suicides from extreme temperature	Present the ARs and ANs in statistical tables. Their respective estimates and confidence intervals should be included for each region and for each disaggregation for the whole country. Timeseries plots and choropleth maps can be used to represent changes over time: • A line or dot plot of the AR over the years by region or for the country. • A choropleth map of the AR in each region over the years. Headline results could be quoted as: • **"**An estimated *xx* suicides (95% confidence interval: *xx* to *xx*) in *reference area* were associated with extreme temperature from *yy* to *yy."* • "There was some indication that suicides from extreme temperatures have increased over recent years; in *output year* an estimated *xx* suicides (95% confidence interval: *xx* to *xx*) were associated with extreme temperatures in *reference area*." *Example charts to be added*

4.6 Limitations & Interpretation

The first main outcome for this indicator is the RR of suicide by temperature. It indicates the likelihood of an individual committing suicide during, or shortly after, exposure to a certain temperature compared to the likelihood of suicide at the MMT. When a temperature has an RR of 1, this means there is neither an increase nor a decrease in the likelihood of the individual committing suicides during, or shortly after, exposure to that temperature. An RR of 1.1 denotes a 10% increase in suicide risk.

The AN and the AR are the estimated number of suicides and rate per 100,000 population, respectively, that would not have occurred if there had been no exposure to extreme temperatures within the determined time lag. These are also the AN and AR of suicides which cannot be explained by the independent variables that have been accounted for in the model. However, if a true confounder is not included in the model, the suicides attributed to extreme temperatures can be biased.

The AN and AR are based on the temperatures experienced in a particular year and any change climate towards more extreme temperatures would likely lead to an increase in attributable deaths. For example, if there are more days with temperatures exceeding thresholds where the suicide risk is high, the AN and AR will rise. Therefore, a decrease in AN or AR does not necessarily indicate better adaptation to extreme temperatures from one year to the next. It may simply be due to there having been less exposure to extreme temperatures in that year. It is, therefore, important to interpret this indicator alongside the framework’s exposure indicators.

When making comparisons between localities and disaggregated groups, the ARs should be used since it considers the differences in population sizes.

Comparing the AR of disaggregated groups can help identify possible social confounding factors. However, some disaggregation variables may be correlated with one another, e.g. ethnicity, socio-economic group and occupation. Strong correlations should be reported in the output to inform possible reasons for differences and therefore, most urgent adaptation needs. This may relate to their respective lifestyles, health statuses, and exposure to extreme temperatures.

A limitation to note with regards to the disaggregation is data availability and low counts of suicides. Not all localities will have the information within the available data sources to disaggregate by each variable. The power of the model will also be challenged as numbers of suicides tend to be low, especially at the daily level. Disaggregating further by these groups is likely to be challenging but important to understand.

To account for limitations in data completeness and accuracy, and the ensuing uncertainty in the indicator estimates, it is important to report their uncertainty (the confidence intervals) and to consider them in the interpretation of variations in the output. Among many possible reasons, such variations may be related to differences in data collection or data sources.

A limitation with the method is that outdoor temperature does not capture a population’s true exposure to extreme temperatures. For example, personal levels of exposure to heat in the home, which may vary depending on the presence of cooling systems, are not captured through ambient temperature data.

A limitation of this methodology is the requirement for daily data, which may not be available in all localities. In the next phase of the indicator development we will test the methodology using weekly data.

Conceptually, use of administrative sources only gives some insight into the overall impact of climate on mental health. There are much wider impacts on levels of anxiety, depression and PTSD, for example, that are challenging to measure at this scale without population level surveys with regular assessments and post-event follow ups.

5. Further reading

Space-Time-Stratified Case-Crossover Design in Environmental Epidemiology Study – Health Data Science

Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis – BMC Medical Research Methodology

Time series regression studies in environmental epidemiology – International Journal of Epidemiology

Reducing and meta-analysing estimates from distributed lag non-linear models – BMC Medical Research Methodology

6. References

(1) Tobias A, Kim Y and Madaniyazi L. Time-stratified case-crossover studies for aggregated data in environmental epidemiology: a tutorial. *International Journal of Epidemiology*. 2024;53(2):dyae020. doi:10.1093/ije/dyae020.

(2) Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*. 2015;386(9991):369-375. doi:10.1016/S0140-6736(14)62114-0.

(3) Copernicus Climate Change Service (C3S). *Climate Data Store*. https://cds.climate.copernicus.eu/datasets [Accessed 9th October 2024].

(4) 4 Earth Intelligence. *Products*. https://www.4earthintelligence.com/products/ [Accessed 9th October 2024].

(5) SDMX Global Registry. *Data Structure Definitions*. https://registry.sdmx.org/data/datastructure.html [Accessed 9th October 2024].

(6) Massazza A, Teyton A, Charlson F, Benmarhnia T and Augustinavicius JL. Quantitative methods for climate change and mental health research: current trends and future directions. *The Lancet Planetary Health*. 2022;6(7):613-627. doi:10.1016/S2542-5196(22)00120-6.

(7) Sim K, Kim Y, Hashizume M, Gasparrini A, Armstrong B, Sera F, et al. Nonlinear temperature-suicide association in Japan from 1972 to 2015: Its heterogeneity and the role of climate, demographic, and socioeconomic factors. *Environment International*. 2020;142. doi:10.1016/j.envint.2020.105829.