BMC
Medical Research Methodology

Attributable risk from distributed lag models

Antonio Gasparrini[1*] and Michela Leone[2]

Abstract

Background: Measures of attributable risk are an integral part of epidemiological analyses, particularly when aimed at the planning and evaluation of public health interventions. However, the current definition of such measures does not consider any temporal relationships between exposure and risk. In this contribution, we propose extended definitions of attributable risk within the framework of distributed lag non-linear models, an approach recently proposed for modelling delayed associations in either linear or non-linear exposure-response associations.

Methods: We classify versions of attributable number and fraction expressed using either a forward or backward perspective. The former specifies the future burden due to a given exposure event, while the latter summarizes the current burden due to the set of exposure events experienced in the past. In addition, we illustrate how the components related to sub-ranges of the exposure can be separated.

Results: We apply these methods for estimating the mortality risk attributable to outdoor temperature in two cities, London and Rome, using time series data for the periods 1993–2006 and 1992–2010, respectively. The analysis provides estimates of the overall mortality burden attributable to temperature, and then computes the components attributable to cold and heat and then mild and extreme temperatures.

Conclusions: These extended definitions of attributable risk account for the additional temporal dimension which characterizes exposure-response associations, providing more appropriate attributable measures in the presence of dependencies characterized by potentially complex temporal patterns.

Keywords: Attributable risk, Attributable fraction, Distributed lag models

Background

Epidemiological studies usually rely on effect summaries based on *ratio measures*, such as relative risk, odds ratio or rate ratio, with the choice depending on the specific study design [1]. Although these measures are ideal for summarizing the association of interest, they offer limited information on the actual impact of the exposure. This information is critical for the planning and evaluation of public health interventions, and it is better provided by *relative excess measures* such as the attributable fraction (AF), or *absolute excess measures* such as the attributable number (AN). Steenland and Armstrong offer a thorough overview on the topic [2]. Here we generally refer to these summaries as *attributable risk measures*.

Problems in the definition of these measures may arise in the presence of *delayed associations*, occurring when

an exposure generates a risk lasting well beyond the exposure period. Researchers in different fields, during the last thirty years, have proposed approaches for modelling this type of association [3-6]. In time series analysis, a popular approach is based on distributed lag models (DLMs) [7,8], generalized to distributed lag non-linear models (DLNMs) when including non-linear exposure-response associations [9,10]. Recently, the DLNM framework has been extended beyond time series data for modelling such dependencies, defined *exposure-lag-response associations*, in different study designs [11]. However, attributable risk measures have not been developed for the DLM and DLNM class, with the result that their current definitions do not take into account the additional temporal structure in the exposure-response association. Previous work investigated the issue in time series analysis, although without producing a general approach [12,13].

In this contribution, we extend this research and attempt to formalize the definition of attributable risk

*Correspondence: antonio.gasparrini@lshtm.ac.uk
[1]Department Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT, London, UK
Full list of author information is available at the end of the article

measures within the DLNM modelling framework. In particular, we illustrate how complex non-linear and temporal patterns can be accounted for in the computation of the attributable risk. Also, we show how attributable components related to different exposure ranges can be separated. We propose an example on the estimation of the health burden attributable to temperature with time series data, a problem which has received quite a lot of interest recently due to climate change predictions. However, the approach can be easily applied to other exposure-lag-response associations. The method is implemented in simple functions developed within the R software package and provided as additional files.

Methods

Attributable risk measures

A general definition of the attributable fraction AF_x and number AN_x for a given exposure x can be provided by:

$$AF_x = 1 - \exp(-\beta_x) , \tag{1a}$$

$$AN_x = n \cdot AF_x , \tag{1b}$$

with n as the total number of cases. The parameter β_x used in Eq. (1a) represents the risk associated with the exposure, and it usually corresponds to the logarithm of a ratio measure such as relative risk, relative rate or odds ratio. It is generally obtained from regression models while adjusting for potential confounders. The general definition of β_x used here refers to the association with a specific exposure intensity x compared to a reference value x_0. For linear exposure-response relationships, the association can also be reported as $\beta \cdot x$, where in this case β refer to a unit increase in x. For binary variables reporting presence/absence of the exposure, Eq. (1a) simplifies to $AF = (RR - 1)/RR$, with RR as relative risk, as reported by Steenland and Armstrong [2]. We keep the more general definition of β_x, which is easily applicable to non-linear exposure-response relationships, throughout the manuscript.

The theoretical nature of these effect measures is based on a *counterfactual*, where the observed condition is compared with a reference state which never occurred. This state postulates that the same population is followed in an identical situation where only the exposure level changes to the reference value x_0. Typically, such a reference is represented by the absence of association, meaning $x_0 = 0$ and $\beta_{x_0} = 0$. However, different counterfactual conditions can be used, for example a lower exposure which can be determined by an intervention. In this case the quantity β_x can be simply re-parameterized as $\beta_x^* = \beta_x - \beta_{x_0}$, and Eq. (1) still applies.

Eq. (1a) can be extended to define the risk attributable to multiple exposures $x_1, \ldots, x_p$:

$$AF_{x_1,\ldots,x_p} = 1 - \exp\left(-\sum_{i=1}^{p} \beta_{x_i}\right) = 1 - \prod_{i=1}^{p}\left(1 - AF_{x_i}\right) , \tag{2}$$

with $AN_{x_1,\ldots,x_p}$ obtained by substituting Eq. (2) in Eq. (1b) [2]. For the specific form of Eq. (2), it should be noted that $AF_{x_1,\ldots,x_p} \leq AF_{x_1} + \ldots + AF_{x_p}$, *i.e.* the sum of the attributable risk measured for individual exposures is usually higher than their concurrent attributable risk.

A review of the DLNM modelling framework

The basic idea underpinning the development of DLNMs is that the risk at time t can be described as the weighted sum of effects cumulated from a series of exposures $x_{t-\ell_0}, \ldots, x_{t-L}$ experienced in the past over the lag period $\ell = \ell_0, \ldots, L$, with ℓ_0 and L corresponding to minimum and maximum lags, respectively. The risk can be described by the function $f(x)$, determining the exposure-response, and the function $w(\ell)$, specifying the lag-response, related to the weights given to exposures at different lags ℓ. These functions are combined in a bi-dimensional *exposure-lag-response function $f \cdot w(x, \ell)$*. Algebraically, the risk is defined by a function $s(x, t; \boldsymbol{\eta})$, written in terms of parameters $\boldsymbol{\eta}$ as:

$$s(x, t; \boldsymbol{\eta}) = \int_{\ell_0}^{L} f \cdot w(x_{t-\ell}, \ell) \, d\ell$$

$$\approx \sum_{\ell=\ell_0}^{L} f \cdot w(x_{t-\ell}, \ell) = \mathbf{w}_{x,t}^{\mathsf{T}} \boldsymbol{\eta} . \tag{3}$$

The function $s(x, t)$ is computed as the approximate integral of the exposure-lag-response function over the lag dimension, representing the cumulated risk over the lag period. The parameterization in the final step of Eq. (3) is obtained through a *cross-basis*, involving a tensor product between the basis chosen for $f(x)$ and $w(\ell)$, generating the transformed variables $\mathbf{w}_{x,t}$ linearly combined with the parameters $\boldsymbol{\eta}$. Simpler DLMs are defined by Eq. (3) by assuming $f(x)$ as linear. Algebraic details and additional information are provided elsewhere [11]. The cross-basis is specified with a reference value x_0 used later as a centering point for the function $f(x)$, which is used to define the counterfactual condition.

The complex parameterization of exposure-lag-response associations provided by Eq. (3) can be more easily interpreted by computing effect summaries from the original parameters $\boldsymbol{\eta}$. Specifically, the bi-dimensional exposure-lag-response risk surface modelled through $f \cdot w(x, \ell)$ can be expressed by a grid of effect summaries $\beta_{x,\ell}$, each interpreted as the association with an exposure x at lag ℓ versus

the reference value x_0. For a given time t, the cross-basis parameterization in (3) can be re-expressed as:

$$\mathbf{w}_{x,t}^{\mathsf{T}}\boldsymbol{\eta} = \sum_{\ell=\ell_0}^{L} \beta_{x_{t-\ell},\ell} \ . \tag{4}$$

This *overall cumulative* association is composed of the sum of contributions $\beta_{x,\ell}$ from exposures $x_{t-\ell_0}, \ldots, x_{t-L}$ experienced within the lag period. Algebraic definitions have been previously provided [11].

Forward and backward perspectives

The term $\beta_{x,\ell}$ for each intensity x can be interpreted using two complementary perspectives, illustrated graphically in Figure 1. From a *forward* standpoint, looking from current exposure to future risks, the terms $\beta_{x,\ell}$ are the contributions from the exposure x_t occurring at time t to the risk at times $t+\ell_0, \ldots, t+L$, identified by green circles. From a *backward* standpoint, looking from current risk to past exposures, the terms $\beta_{x,\ell}$ are the contributions to the risk at time t from exposures $x_{t-\ell_0}, \ldots, x_{t-L}$ experienced at $t-\ell_0, \ldots, t-L$, identified by yellow squares. The underlying curve in Figure 1 depicting such associations is called the *lag-response curve* related to a given exposure intensity x. The sum of these contributions over the whole lag period can be interpreted as the overall cumulative risk.

Attributable risk from DLNMs

The effect summaries provided above can be used for defining attributable risk measures within the DLNM framework. The idea is to treat the associations with exposures at different lags as independent contributions to the risk. A neat definition can be developed using a backward perspective, assuming the risk at time t as attributable to a series of exposure events in the past. The backward

attributable fraction b-AF$_{x,t}$ and number b-AN$_{x,t}$ at time t are obtained by substituting Eq. (4) in Eq. (2):

$$\text{b-AF}_{x,t} = 1 - \exp\left(-\sum_{\ell=\ell_0}^{L} \beta_{x_{t-\ell},\ell}\right), \tag{5a}$$

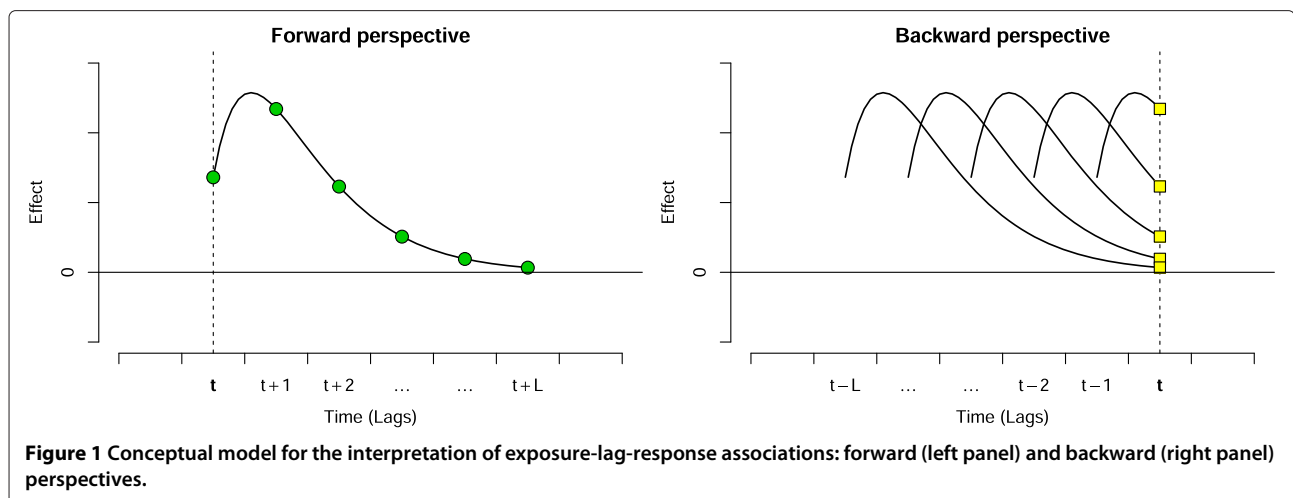$$\text{b-AN}_{x,t} = \text{b-AF}_{x,t} \cdot n_t \ , \tag{5b}$$

with n_t as the number of cases at time t. This structure is consistent with the configuration of the regression model usually applied to fit the data, where the risk at time t is associated with lagged exposures at times $t - \ell$. The definition of backward attributable risk requires an extended version of the counterfactual condition accounting for the additional lag dimension: b-AN$_{x,t}$ and b-AF$_{x,t}$ are interpreted as the number of cases and the related fraction at time t attributable to past exposures to x in the period $t - \ell_0, \ldots, t - L$, compared to a constant exposure x_0 throughout the same period.

An alternative version can be obtained using a forward perspective. Among other possible definitions, forward attributable number f-AN$_{x,t}$ and fraction f-AF$_{x,t}$ can be defined as:

$$\text{f-AF}_{x,t} = 1 - \exp\left(-\sum_{\ell=\ell_0}^{L} \beta_{x_t,\ell}\right), \tag{6a}$$

$$\text{f-AN}_{x,t} = \text{f-AF}_{x,t} \cdot \sum_{\ell=\ell_0}^{L} \frac{n_{t+\ell}}{L - \ell_0 + 1} \ . \tag{6b}$$

This alternative version has some advantages if compared to the backward definition. First, the counterfactual condition is simpler: f-AF$_{x,t}$ and f-AN$_{x,t}$ are interpreted as the fraction and number of future cases in the period $t + \ell_0, \ldots, t+L$ attributable to the single exposure x occurring

Figure 1 Conceptual model for the interpretation of exposure-lag-response associations: forward (left panel) and backward (right panel) perspectives.

at time t, compared to x_0. Moreover, the overall cumulative risk $\sum \beta_{x_t,\ell}$ for a given exposure x_t in (6a) is available also when the bi-dimensional exposure-lag-response is reduced to uni-dimensional exposure-response relationship, a step often needed in multi-site studies [14]. In contrast, all the lag-specific contributions are needed to compute $\sum \beta_{x_{t-\ell},\ell}$ in (5a) for the backward counterpart.

However, the forward version also has an important limitation, related to the fact that the contributions are associated to risks measured at different times. The attributable number f-AN$_{x,t}$ in (6b) is computed by averaging the total counts experienced in the next $\ell_0,\ldots,L$ times, thus only approximating the lag structure of risks. This approximation is likely to produce some bias, which is expected as an underestimation of the attributable number if compared to the backward version.

Separating attributable components

The definitions provided in Eq. (5)–(6) can be extended to separate the attributable components related to specific exposures or exposure ranges. This will be used later in the example to single out the contributions from cold and heat in temperature-health associations. Let's define a range $r = [\,l, h\,]$ between low and high exposure limits l and h. The definition of forward attributable number f-AN$_{x,t}^r$ and fraction f-AF$_{x,t}^r$ limited to exposures within the range r is clear-cut, as they are either equal to the quantities reported in Eq. (6) if $x \in r$ or zero otherwise. Adopting a backward perspective, a similar definition of b-AF$_{x,t}^r$ can be obtained by modifying Eq. (5a) as:

$$\text{b-AF}_{x,t}^r = 1 - \exp\left(-\sum_{\ell=\ell_0}^{L} I\left(x_{t-\ell} \in r\right) \beta_{x_{t-\ell},\ell} \right), \quad (7)$$

simply selecting the risk contributions from past exposures included in the range r. The related attributable number b-AN$_{x,t}^r$ is computed by substituting Eq. (7) into Eq. (5b). Attributable components referring to different ranges can be summed up, as all are defined using the same counterfactual condition of a constant exposure $x_\ell = x_0$ for the whole lag period $\ell = \ell_0,\ldots,L$.

The forward version has the additional advantage that for two non-overlapping ranges r_1 and r_2 the sum of the components is equal to the overall attributable risk, namely f-AF$_{x,t}^{r_1+r_2}$ = f-AF$_{x,t}^{r_1}$+f-AF$_{x,t}^{r_2}$. In contrast, adopting a backward perspective b-AF$_{x,t}^{r_1+r_2} \leq$ b-AF$_{x,t}^{r_1}$ + b-AF$_{x,t}^{r_2}$, as the risks are simultaneously computed for the same time t in the like of Eq. (2).

Total attributable risk

The attributable risk measures provided above can be computed for each of the $i = 1,\ldots,m$ observations in

a data set. An estimate of the total attributable number AN$_{tot}$ and fraction AF$_{tot}$ is provided by:

$$\text{AN}_{tot} = \sum_{i=1}^{m} \text{AN}_{x,t_i} \,, \quad (8a)$$

$$\text{AF}_{tot} = \text{AN}_{tot} / \sum_{i=1}^{m} n_{t_i} \,. \quad (8b)$$

The equations above can be applied either to forward or backward attributable risk and to separate components, simply substituting the related attributable numbers in Eq. (8a).

Computing uncertainty intervals

Analytical formulae for confidence intervals of attributable risk measures are not easily produced [15], and this also applies to the extended versions developed here. Although approximated estimators have been proposed [15,16], in this context the most straightforward approach is to rely on interval estimation obtained empirically through Monte Carlo simulations [17,18]. Basically, we take random samples $\boldsymbol{\eta}^{(j)}$ of the original parameters $\boldsymbol{\eta}$ of the cross-basis in Eq. (3) from the assumed multivariate normal distribution with point estimate $\hat{\boldsymbol{\eta}}$ and (co)variance matrix $V(\hat{\boldsymbol{\eta}})$ derived from the regression model. These samples $\boldsymbol{\eta}^{(j)}$ are used to compute $\beta_{x,\ell}^{(j)}$ for $\ell = \ell_0,\ldots,L$ and each intensity x, empirically reconstructing the distributions of the attributable measures defined in Eq. (5)–(8). The related 2.5[th] and 97.5[th] percentiles of such distributions are interpreted as 95% empirical confidence intervals (eCI).

Results

The methods illustrated in the previous section are applied to estimate the all-cause mortality risk attributable to temperature, using daily time series from two cities, London and Rome, in the periods 1993-2006 and 1992-2010 respectively. R scripts and data implementing the method and partly replicating the results are provided as Additional files 1, 2, 3, 4, 5 and 6.

Modelling strategy

We fitted a standard time series Poisson model allowing for overdispersion, controlling for seasonal and long term trends and day of the week, using a 10 df/year spline and indicator variables, respectively. Model selection is still an issue of current research within the DLNM framework, although simulation studies indicate a good performance of methods based on the Akaike Information criterion (AIC) [11]. Considering the illustrative purpose of the

example, we selected a-priori the cross-basis function in Eq. (3) for representing the association between mean daily temperature and mortality, basing our choice on previous analyses. Specifically, the cross-basis is composed of a quadratic B-spline with two equally-spaced knots as the exposure-response function $f(x)$, and a natural cubic B-spline with three equally-spaced knots in the log-scale as the lag-response function $w(\ell)$ over lags 0–25.

In the specific case of temperature where a null exposure condition cannot be defined, a reasonable choice is to center the cross-basis in Eq. (3) to the temperature of minimum risk, as suggested in previous publications [13]. This optimal temperature corresponds to 20°C and 21°C for London and Rome respectively, and it represents the reference point x_0 for the computation of the attributable risk measures. These are obtained for the whole temperature range, and then for cold and heat contributions by separating the associations with temperatures lower or higher than x_0. In addition, the attributable components are separated further in mild and extreme cold and heat by selecting as cut-off values the 1st and 99th percentiles of city-specific distributions, corresponding to 0.4°C and 23.7°C in London and 2.6°C and 28.6°C in Rome.

We derived empirical confidence intervals for backward total attributable numbers and fractions, computed overall and for separated components, by simulating 5,000 samples from the assumed distribution of $\hat{\eta}$.

Risk attributable to temperature

The estimated associations between temperature and all-cause mortality in the two cities are illustrated in Figure 2. The left panels show the bi-dimensional exposure-lag-response surfaces, while the right panels display the overall cumulative exposure-response curves, interpreted as the risk cumulated over the entire lag period of 0–25 days. The associations are represented in the RR scale, with the centering point and the cut-off values for defining extreme cold and heat displayed as dotted and dashed vertical lines respectively.

Table 1 reports the estimates of the total backward and forward attributable fraction b-AF$_{tot}$ and f-AF$_{tot}$, with 95%eCI. The backward version indicates that overall 13.59% and 12.58% deaths are attributable to temperature in London and Rome within the study periods, respectively. As expected, the corresponding estimates computed forward are affected by a certain degree of negative bias associated to the averaging of future mortality within the lag period, as described above. Nonetheless, the difference is not substantial in this case.

Cold, heat and extreme components

The total backward attributable risk is then separated into components due to cold and hot temperatures, defined as those below and above the optimal temperature, respectively. The estimates, computed using Eq. (7), are reported
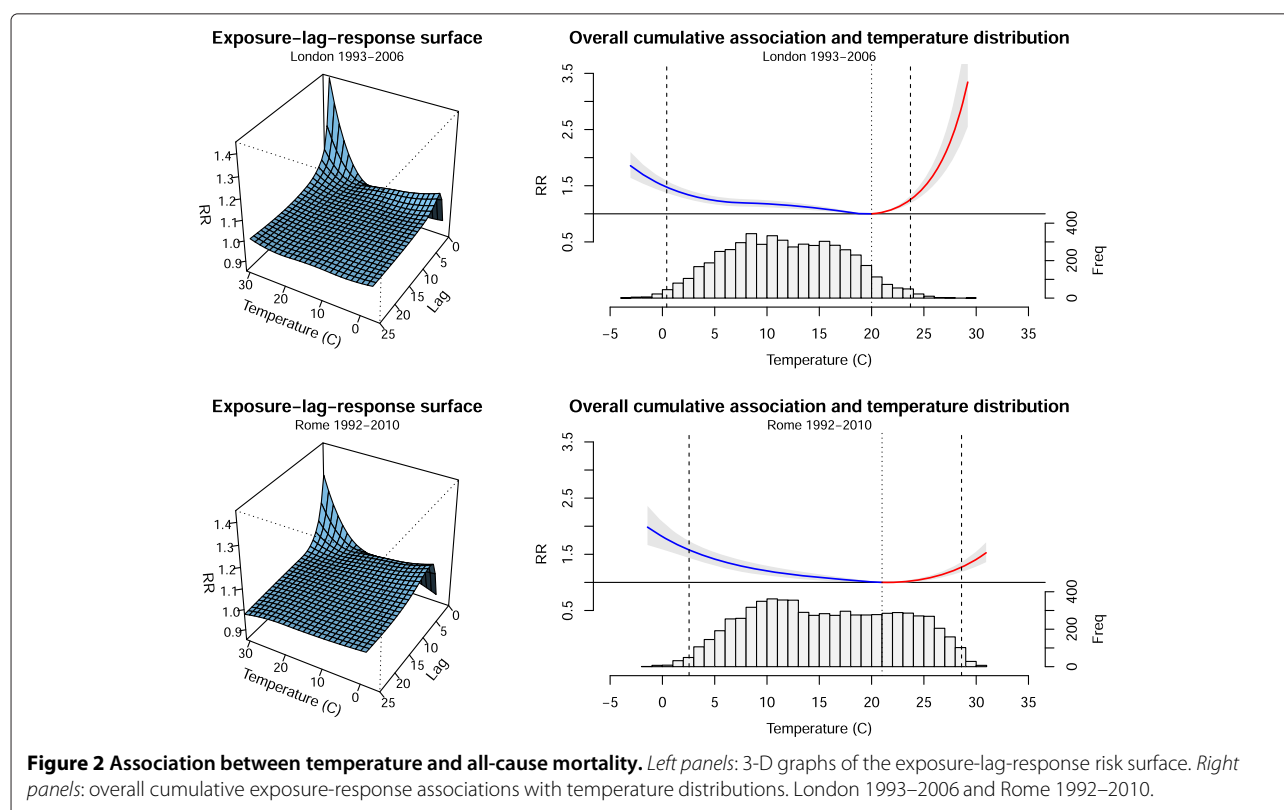

Figure 2 Association between temperature and all-cause mortality. *Left panels*: 3-D graphs of the exposure-lag-response risk surface. *Right panels*: overall cumulative exposure-response associations with temperature distributions. London 1993–2006 and Rome 1992–2010.

Table 1 Total mortality fraction (%) attributable to temperature, computed backward (b-AF$_{tot}$) and forward (f-AF$_{tot}$), reported as overall, hot and cold components with 95% empirical confidence intervals (eCI)

	Deaths		Overall	Cold	Hot
London	845,215	b-AF$_{tot}$	13.59 (10.04–17.09)	12.95 (9.32–16.38)	0.66 (0.52–0.80)
		f-AF$_{tot}$	13.41 (9.72–16.87)	12.84 (9.38–16.33)	0.57 (0.45–0.68)
Rome	395,691	b-AF$_{tot}$	12.58 (9.30–15.64)	10.84 (7.37–14.23)	1.74 (1.12–2.37)
		f-AF$_{tot}$	12.27 (8.94–15.41)	10.72 (7.19–14.00)	1.55 (0.95–2.13)

London 1993–2006 and Rome 1992–2010.

in Table 1. The comparison of the two contributions clearly indicates that cold is responsible for most of the mortality attributable to temperature, with b-AF$_{tot}$ equal to 12.95% and 10.84%, compared to 0.66% and 1.74% for heat, in the two cities. Estimates of forward attributable risk are very similar, and as expected their sum is equal to the overall burden, differently than for the backward version.

The analysis is extended by further separating the attributable components into contributions from mild and extreme temperatures, with results summarized in Table 2. For cold temperatures, the contribution from extreme days accounts only for a minimal part of the mortality burden, while still 12.38%–12.48% and 10.27%–10.37% of deaths in London and Rome, respectively, are attributable to mild cold days. This result is expected when inspecting the right panels of Figure 2, with the definition of mild cold including the majority of days in the series for both cities.

In contrast, the comparison between the two cities is rather different for the components attributable to mild and extreme hot temperatures. In spite of the stronger risk in London, the attributable fraction is similar for extreme heat and even higher in Rome for mild heat (1.32%–1.45% versus 0.25%–0.33%). This apparent contradiction is explained by the different temperature distribution, and in particular the percentile corresponding to the optimal temperature, corresponding to 93.6[th] and 72.5[th] in London and Rome. This result suggests the hypothesis that the population in Rome is more adapted to the range of temperatures corresponding to extreme hot if compared to London, where the population experienced only a few days of unusually high temperatures.
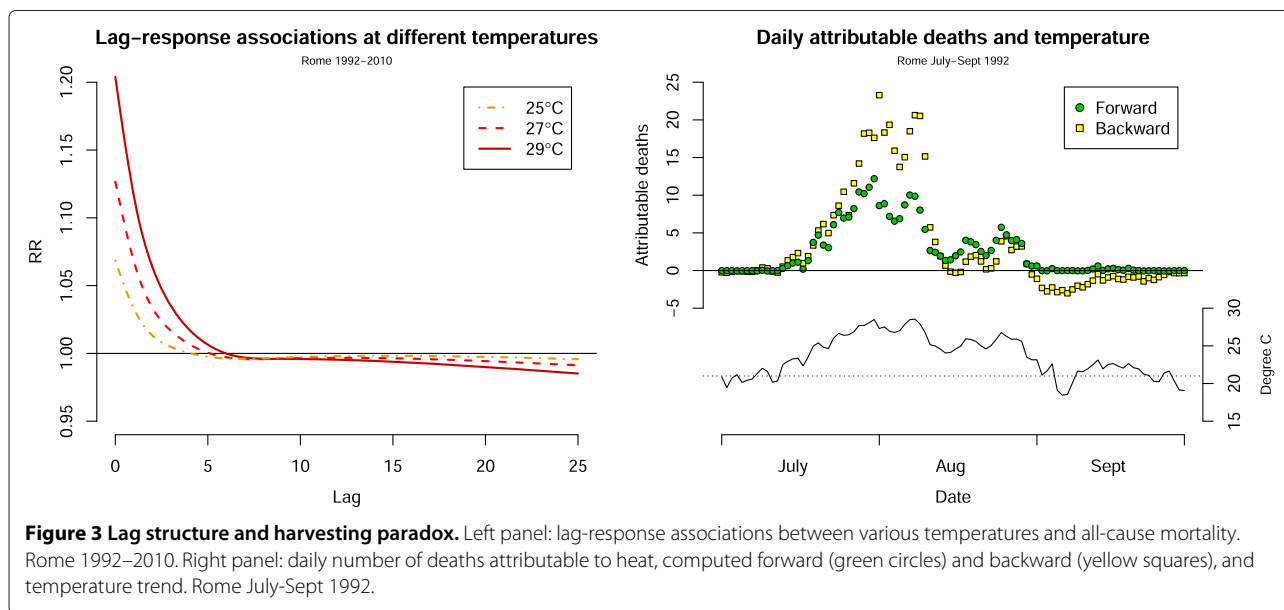
The harvesting paradox

Accounting for the additional lag dimension in exposure-lag-response associations involves further complexities in the interpretation of attributable risk measures. We now focus our attention to the association with hot temperature in Rome. The left panel of Figure 3 shows the estimated lag-response curves at various temperatures, computed versus the reference optimal temperatures of 21°C. The graph indicates a strong risk in the first lags, then followed by a protective association at longer lags. This pattern is consistent with the harvesting hypothesis, which assumes that the initial risk is partly discounted by the depletion of the pool of susceptible after an extreme heat event [19]. Also, the extent of harvesting seems more pronounced for extreme temperatures.

This phenomenon has interesting implications. An example is offered by the right panel of Figure 3, illustrating the estimated daily deaths b-AN$_{x,t}$ and f-AN$_{x,t}$ attributable to heat, computed backward and forward for the first summer in the time series for Rome, with related temperature trend. As expected, a substantial number of deaths are attributable to temperatures above the optimal value, represented by the horizontal dotted line, in the period mid-July to mid-August. The trend of forward attributable deaths f-AN$_{x,t}$ closely follows the daily temperatures, consistently with the definition of number of deaths attributable to the temperature in day t cumulated in the next L days. In contrast, the backward attributable number b-AN$_{x,t}$ decreases to zero and even becomes negative in late summer days, although the overall cumulative exposure-response in Figure 2 (bottom-right panel) does not show a RR below 1 for any temperature.

Table 2 Total mortality fraction (%) attributable to temperature, computed backward (b-AF$_{tot}$) and forward (f-AF$_{tot}$), reported as components from mild and extreme hot and cold contributions with 95% empirical confidence intervals (eCI)

		Extreme cold	Mild cold	Mild hot	Extreme hot
London	b-AF$_{tot}$	0.55 (0.45–0.64)	12.48 (8.86–15.88)	0.31 (0.23–0.38)	0.36 (0.29–0.43)
	f-AF$_{tot}$	0.47 (0.40–0.53)	12.38 (8.98–15.78)	0.29 (0.22–0.35)	0.28 (0.23–0.33)
Rome	b-AF$_{tot}$	0.59 (0.47–0.70)	10.37 (6.88–13.63)	1.45 (0.89–2.01)	0.33 (0.25–0.40)
	f-AF$_{tot}$	0.47 (0.39–0.54)	10.27 (6.69–13.50)	1.32 (0.75–1.85)	0.25 (0.19–0.30)

London 1993–2006 and Rome 1992–2010.

Figure 3 Lag structure and harvesting paradox. Left panel: lag-response associations between various temperatures and all-cause mortality. Rome 1992–2010. Right panel: daily number of deaths attributable to heat, computed forward (green circles) and backward (yellow squares), and temperature trend. Rome July-Sept 1992.

This paradox is explained by the counterfactual condition associated with the backward perspective. Specifically, each b-AN$_{x,t}$ compares the association with the observed temperatures in the past L days to a constant exposure x_0. In the presence of harvesting, the observed population becomes 'healthier' than the counterfactual population after a series of heat days, due to the depletion of the susceptible pool. This explains the negative attributable numbers for specific combinations of lagged exposures. This fact emphasises that harvesting should not be interpreted as a true protective association at longer lags, but rather as an artefact due to a change in the underlying population following a stress, which affects the counterfactual condition. This issue is relevant when using backward attributable risk measures b-AN$_{x,t}$ and b-AF$_{x,t}$ to assess the contribution of specific days. However, similarly to the net overall cumulative risk, the total attributable number b-AN$_{tot}$ and fraction b-AN$_{tot}$, produced by Eq. (8) and reported in Tables 1 and 2, account for the discount by summing the contributions over the whole series.

Discussion and conclusions

In this contribution we illustrate an extended definition of attributable risk measures based on the DLNM framework. Consistently with this class of models, such a definition accounts for the complex pattern of potentially non-linear and delayed associations described through exposure-lag-response associations.

Two alternative definitions of attributable risk are proposed, assuming backward or forward perspectives. The former provides more consistent estimators which naturally arise from the structure of the regression model,

where distributed lag terms at times $t - \ell$ contributes to the risk at time t. The forward attributable measures, in contrast, are affected from a negative bias related to the averaging of future counts, which nonetheless is likely to be relatively low. On the other hand, the forward version is well suited for separating the risk in components attributable to different ranges, as their sum matches the overall risk. Furthermore the forward perspective, looking from current exposure to future risk, seems more appropriate for quantifying the health burden due to specific exposure occurrences, as it is based on a more coherent counterfactual condition. Corrections have been proposed in previous works on risk attributable to multiple exposures [20-22], and can be applied to the backward version.

Strictly speaking, the definition given in Eq. 1a is interpreted as the attributable fraction among the sub-population of exposed subjects. In the setting of time series analysis for environmental stressors, the whole population is usually considered as exposed, and this definition can be more generally interpreted as the population attributable fraction. If only a subset is instead exposed, Eq. (5)–(8) can be easily extended using the equations proposed by Steenland and Armstrong [2] for population attributable risk.

Previous papers suggested approaches for producing attributable risk from distributed lag models when applied to heat-mortality associations. Baccini and colleagues applied DLMs and computed attributable risk measures, specifically addressing the issue of harvesting [12]. Honda and colleagues illustrated an analysis on the mortality burden due to heat using DLNMs [13]. However both approaches are limited, as the former assumes a linear

threshold form of the exposure-response, while the latter averages the non-linear risk across the whole temperature range. In this paper we offer a formal and more consistent definition of such attributable risk measures.

An advantage of the proposed method is the provision of estimates for separate components of the attributable risk, associated with different exposure ranges. In the specific case of temperature-health associations, this allows the separation of attributable risks from cold and heat, and further from mild and extreme temperatures. The estimates reported in the example highlights how the simple analysis of exposure-response curves can be misleading in the attribution of risk, and that most of the mortality in the two cities is in fact attributable to mild cold temperatures, in spite of the relatively low RR.

The availability of attributable risk measures, complementary to estimates of exposure-response associations, is essential for the identification and planning of public health interventions. Their extension to exposure-lag-response associations allows the computation of such measures from dependencies showing potentially complex non-linear and temporal patterns.

Additional files

> **Additional file 1: R script implementing the function to compute attributable risk measures.**
>
> **Additional file 2: Documentation of the attrdl function.**
>
> **Additional file 3: R script for fitting the regression models in the example.**
>
> **Additional file 4: R script for computing attributable risk in the example.**
>
> **Additional file 5: R script for producing the graphs in the example.**
>
> **Additional file 6: File with data used in the example.**

Abbreviations

AF: Attributable fraction; AN: Attributable number of cases; RR: Relative risk; DLM: Distributed lag models; DLNM: Distributed lag non-linear models; eCI: empirical confidence interval.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AG conceived the idea of attributable risk measures for DLNMs and worked out their algebraic definitions. AG and ML developed the final version of the measures, planned the example and carried out the analysis, drafted the final version of the manuscript. AG provided the software implementation through the R scripts. Both authors read and approved the final manuscript.

Author details

[1] Department Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, WC1E 7HT, London, UK. [2] Department of Epidemiology, Lazio Regional Health Service, Via di Santa Costanza 53, 00198 Rome, Italy.

References

1. Rothman KJ, Greenland S, Lash TL: *Modern Epidemiology*. Philadelphia: Lipcott Williams & Wilkins; 2008.
2. Steenland K, Armstrong B: **An overview of methods for calculating the burden of disease due to specific risk factors.** *Epidemiology* 2006, **17**(5):512–519.
3. Thomas DC: **Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks.** *Scand J Work Environ Health* 1983, **9**(4):353–366.
4. Breslow NL, Day NE: *Statistical Methods in Cancer Research. Vol. II: The Desing and Analysis of Cohort Studies*. Lyon: International Agency for Reasearch on Cancer (IARC); 1987:232–271. Chap. 6: Modelling the relationship between risk, dose and time.
5. Sylvestre MP, Abrahamowicz M: **Flexible modeling of the cumulative effects of time-dependent exposures on the hazard.** *Stat Med* 2009, **28**(27):3437–3453.
6. Richardson DB: **Latency models for analyses of protracted exposures.** *Epidemiology* 2009, **20**(3):395–399.
7. Almon S: **The distributed lag between capital appropriations and expenditures.** *Econometrica* 1965, **33**:178–196.
8. Schwartz J: **The distributed lag between air pollution and daily deaths.** *Epidemiology* 2000, **11**(3):320–326.
9. Armstrong B: **Models for the relationship between ambient temperature and daily mortality.** *Epidemiology* 2006, **17**(6):624–631.
10. Gasparrini A, Armstrong B, Kenward MG: **Distributed lag non-linear models.** *Stat Med* 2010, **29**(21):2224–2234.
11. Gasparrini A: **Modeling exposure-lag-response associations with distributed lag non-linear models.** *Stat Med* 2014, **33**(5):881–899.
12. Baccini M, Kosatsky T, Biggeri A: **Impact of summer heat on urban population mortality in Europe during the 1990s: an evaluation of years of life lost adjusted for harvesting.** *PloS One* 2013, **8**(7):69638.
13. Honda Y, Kondo M, McGregor G, Kim H, Guo YL, Hijioka Y, Yoshikawa M, Oka K, Takano S, Hales S, Kovats RS: **Heat-related mortality risk model for climate change impact projection.** *Environ Health Prev Med* 2013. (Epub ahead of print. doi:10.1007/s12199-013-0354-6).
14. Gasparrini A, Armstrong B: **Reducing and meta-analyzing estimates from distributed lag non-linear models.** *BMC Med Res Methodol* 2013, **13**(1):1.
15. Graubard BI, Fears TR: **Standard errors for attributable risk for simple and complex sample designs.** *Biometrics* 2005, **61**(3):847–855.
16. Cox C, Li X: **Model-based estimation of the attributable risk: a loglinear approach.** *Comput Stat Data Anal* 2012, **56**(12):4180–4189.
17. Greenland S: **Interval estimation by simulation as an alternative to and extension of confidence intervals.** *Int J Epidemiol* 2004, **33**(6):1389–1397.
18. Wood SN: *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman & Hall/CRC; 2006.
19. Schwartz J: **Is there harvesting in the association of airborne particles with daily deaths and hospital admissions?** *Epidemiology* 2001, **12**(1):55–61.
20. Eide GE, Gefeller O: **Sequential and average attributable fractions as aids in the selection of preventive strategies.** *J Clin Epidemiol* 1995, **48**(5):645–655.
21. Land M, Vogel C, Gefeller O: **Partitioning methods for multifactorial risk attribution.** *Stat Methods Med Res* 2001, **10**(3):217–230.
22. Hamel JF, Fouquet N, Ha C, Goldberg M, Roquelaure Y: **Software for unbiased estimation of attributable risk.** *Epidemiology* 2012, **23**(4):646–647.