H2020 program	The Exposome Project for Health and Occupational Research
Grant agreement number	874703
Title	Report on the inventory of Exposure-Time-Response (ETR) models for EPHOR
Date	2022
Responsible author	Dr Susan Peters
E-mail	s.peters@uu.nl
Co-authors	Wenxin Wan (UU)
	Dr. Lützen Portengen (UU)
	Dr. Tanja Krone (TNO)
	Prof. Roel Vermeulen (UU)

Contents

1. Exposure-Time-Response relations in epidemiological research

1.1 Deficiencies of using cumulative exposure

In epidemiologic modelling of exposure-response relations, researchers often summarise a subject's exposure history using the metric of cumulative exposure (CE), i.e., the integral of (time-specific) intensity of exposure over the full exposure history (White et al., 2008). Although CE is by far the most often used metric for epidemiologic analyses of chronic diseases, two assumptions underlying its validity are often ignored: i) stability of effects, i.e., the effects of exposure on disease incidence are the same, irrespective of when the exposure occurred; and ii) independence of effects, i.e., the effects of the exposure incurred at one time point do not affect the effects at other time points.

For studies with a short time span (e.g., as in experimental toxicology), CE may be an appropriate summary measure of the exposure history, but for studies with more sustained exposures, the use of CE will obscure the complicated relations between exposure, dose, and disease risks (de Vocht et al., 2015). For example, a worker subjected to low level exposure over a long period of time (e.g., 1 ppm-year over 20 years) could well be at a different risk than a worker exposed to high exposure levels over a short period (e.g., 20 ppm for just one year) (White et al., 2008). Such differences may reflect a non-linear exposure-response relation (e.g., a no-effect level), but could also be due to possible age-related changes in physiology affecting exposure uptake and metabolism (Kim et al., 2006). Richardson (2008) showed that, conditioned on exposure intensity, the relation between benzene and leukaemia mortality seemed to be much stronger for benzene exposure accrued at older ages (≥45 years) than for that at younger ages. For occupational exposure to sparsely ionizing radiation, several epidemiologic studies have indicated that protracted or fractionated exposures at low dose rates may result in lower risks of cancer than short-lived intense exposures for the same total dose (Jacob et al., 2009).

On the other hand, dependence of effects over several time points, as is likely to occur for allergic diseases (Dotson et al., 2015), could be harder to determine, and such effects are often overlooked with common modelling approaches (Basagana and Barrera-Gomez, 2022). In practice, rather than completely ignoring these time-related aspects, researchers tend to approach these issues in a simple manner, for example, by assessing exposure effect modification by time-since-last-exposure or average exposure intensity using stratified analysis. However, such simple approaches may fail to provide necessary insights into the complex interplay between exposure, time, and response, as the results of these ad-hoc approaches are often based on approximations (e.g., average intensity) and differences between estimated effects are often difficult to comprehensively evaluate (Kriebel et al., 2007, de Vocht et al., 2015).

1.2 Latency and exposure lagging

Exposure latency is the term often applied to the time interval between exposure and disease manifestation or recognition. Rothman defined *latency* as the period from disease initiation to manifestation (Rothman, 1981). This is in line with the idea that there is some period of time when the disease exists in a "hidden" state in an individual. A related concept is *induction time*, which can be defined as the period from first exposure to disease initiation. The sum of the latency and induction intervals is what was referred to as the *empirical induction time* (Rothman, 1981), and by exposure lagging, researchers aim to obtain better estimation of the latency period (or to distinguish the induction and latency periods). As an example, Checkoway et al., (1997) reported results that suggest a stronger association between silica exposure and respiratory disease under a 15-year lag than without any lag. Some techniques to take into account the latency or induction time were applied, including exposure truncation at either some fixed age or fixed time interval since first exposure; and lag exposures by some assumed latency interval (Checkoway et al., 1990, Rothman, 1981). However, these approaches require assumptions on etiologically "relevant" latency periods by excluding "irrelevant periods" but sacrifice information of exposure history either by subject exclusion or exposure truncation. Moreover, the selection of lag periods could be arbitrary, limiting the interpretation of results.

1.3 Time windows

A natural extension of the lagging approach is to estimate effects related to exposures that occur during pre-specified time periods, or time windows. For a particular outcome, exposures occurred at other periods could be considered less important than exposures during a critical time window. An illustration would be that exposures only show effects on birth defects during some particular periods of fetal development (Selevan et al., 2000). The life course approach has been suggested as a conceptual framework for guiding the investigation into the temporal relationship between long-term exposure and risk of chronic diseases (Kuh et al., 2003). The approach highlights the significance of identifying critical periods (a limited time window where an exposure *can* have an effect on the outcome) and sensitive periods (a time window where the effect of exposure is *stronger* than other periods) (Kuh et al., 2003). Identifying different time windows that are relevant in occupational epidemiology, as well as taking a life course approach provides opportunities to systematically explore the complex temporal associations, given complete exposure history data are available (Vineis et al., 2013b).

The life course approach also takes into account more than just single exposures. Kuh et al., (2003) used several *chain of risk models* to refer to a sequence of linked exposures that lead to increased disease risk because one exposure tends to lead to another. In these causal chain models, exposures could act independently (*simple accumulation model*), but can also be "clustered" via a shared cause (*accumulation with risk clustering model*). In the *additive chain of risk model*, exposure can act both as

mediating factors (exposures in between the pathway of another exposure-response relationship) and *modifying factor*s (exposures having effect on another set of exposure and outcome, differentiating the exposure-response relationships), and in the *trigger chain of risk model*, exposures are described to have this "trigger effect" describing the situation where earlier exposures have no effect on the disease risk without the final link in the chain that precipitates disease onset (Kuh et al., 2003). Although some statistical approaches have been proposed to validate life course models for different research questions (Petersen et al., 2021, Madathil et al., 2018), application in an occupational epidemiological setting has been limited and the interpretation of effect of multiple exposures at multiple time windows could be very complicated (Vineis et al., 2013a).

1.4 Multistage models

Multistage models for carcinogenesis were introduced in the 1950s to explain the pattern of age-specific mortality curves for many adult-onset carcinomas. Results from these models indicated that many age-specific mortality curves are consistent with carcinogenesis being the result of an accumulation of mutations, an observation that was reinforced by recent studies showing a correlation between the number of stem cell divisions and cancer incidence rates for several tissues (Tomasetti et al., 2017, Tomasetti and Vogelstein, 2015). These observations could suggest that accumulation of randomly occurring mutations at critical gene loci could be important in carcinogenesis because mutations mostly occur during cell division. The best known multistage model is the Armitage–Doll model (Armitage and Doll, 1954), which continues to be a framework for understanding spontaneous carcinogenesis and the temporal evolution of disease risk with carcinogenic exposures with varying intensity (Day and Brown, 1980, Brown and Chu, 1983). Richardson (2008) discussed the multistage model in the context of occupational exposure to benzene exposure and risk of leukaemia, where the temporal effects could be modelled as a multistage process. The result suggested benzene exposure affects the penultimate stage in disease induction.

The two-stage clonal expansion (2SCE) model, also known as the Moolgavkar–Venzon–Knudson (MVK) model, is based on three influential ideas from cancer biology. These are, in chronological order: i) the concept of initiation-promotion-progression observed from experiments in chemical carcinogenesis (Berenblum and Shubik, 1947) in the late 1940s; ii) the observation that the age-specific mortality curves of many adult carcinomas could be explained by a multistage model (Moolgavkar and Venzon, 1979); and iii) Knudson's two-hit hypothesis for embryonal tumours, such as retinoblastoma (Knudson et al., 1975, Hethcote and Knudson, 1978). The latter hypothesis suggested that most tumour suppressor genes require both alleles to be inactivated, either through mutations or through epigenetic silencing, to cause a phenotypic change (Knudson et al., 1975). The 2SCE model and its generalizations allow for clonal expansion of intermediate cells on the pathway to cancer via a linear birth–death process, and it was applied to study the individual effects of asbestos and silica on lung cancer (Zeka et

al., 2011), where the results suggested strong evidence for an early effect of asbestos, but for silica, an early and less evidently late effect on lung cancer. Richardson (2009) also applied the 2SCE in the context of occupational exposure to benzene exposure and leukaemia mortality among rubber hydrochloride production workers, where the effect of benzene on leukaemia risk appears to be due to an exposure-induced increase in the proliferation of initiated cells.

1.5 Research objectives

By adopting the life-course concepts and common research topics in occupational epidemiology, we proposed the following research objectives that could potentially be achieved using ETR modelling approaches:

1. assess the effects of different aspects of exposure history (e.g., intensity and duration) on disease risk;
2. investigate the effect of temporal modifiers (e.g., time since last exposure, age at first exposure) on the exposure-response relationship; and
3. evaluate the effect of time-varying exposures on different stages of disease development.

In this report we aimed to provide some background of ETR and an inventory of statistical models for ETR modelling in occupational epidemiology.

2. Development of the methods inventory

To develop an inventory of ETR methods, we first selected 14 studies based on consultation with two experts - both have more than 10 years' experience in biostatistics/occupational epidemiology (LP) and in statistics (TK). Those 14 papers were considered to represent state-of-the-art examples illustrating the development and application of ETR-related approaches in epidemiologic research, covering a wide range of methods, namely: multi-state models (Jackson, 2011, Jackson et al., 2003), multi-stage clonal expansion models (Zeka et al., 2011, Richardson, 2008), hierarchical regression models (Richardson et al., 2011), exposure rate model (Richardson et al., 2012), compartmental hidden Markov model (Chadeau-Hyam et al., 2014), extensions of weighted cumulative index (Lacourt et al., 2017, Mauff et al., 2017, Wagner et al., 2021), trajectory approaches (Lévêque et al., 2020), structured Bayesian regression tree pairs (Mork and Wilson, 2021), "flexible modelling" (Danieli et al., 2019), and the reconstruction of exposure metrics (Wang et al., 2016).

Based on those 14 papers, we widened our search grid by using the freely available online platform "Connected Papers" (Connected Papers, 2022). This platform was used because it was expected to provide relevant ETR model papers in a more efficient way than the keyword-based systematic

literature search, which would be time-consuming and outside the scope of this report. The artificial intelligence (AI) algorithm that it uses is based on the concepts of co-citation and bibliographic coupling to create the similarity metric, and takes into account the fact that papers that do not directly cite each other may have similar bibliographic profiles (Ammar et al., 2018). Connected Papers mines the Semantic Scholar database (covering >200 million academic papers) and provides 41 papers that are judged by the AI algorithm to be similar to the single input paper.

We acknowledge that our review was not meant to be exhaustive, but readers with interest in more technical and conceptual details on the ETR modelling can refer to other papers (Chen et al., 2015, Buckley et al., 2019, Sanchez et al., 2011, Thomas, 1988) as well as the 14 "key" papers cited in the method section.

Using the 14 papers and the "Connected Papers" platform, we obtained 447 unique papers that were considered possibly relevant for ETR analyses. We additionally checked the citations from all 14 original papers to include more potentially relevant papers. We screened the collected papers based on the following eligibility criteria:

a) a statistical approach to allow estimation of the effects of time-resolved exposures could be identified either in the abstract or method section;
b) the statistical approach was applied or showed applicability to model non-communicable diseases as an outcome (e.g., cancer); and
c) only original research articles published in English (including pre-prints).

3. Selected ETR modelling approaches

Based on the aforementioned approach to develop the inventory, we retained 189 papers, covering 24 implementations that could be used to model ETR relationships. For each implementation, we specified the modelling goal, the key reference(s), key developer, the available software for model implementation, and comments on its utility. The inventory can be found at the end of this report (Table 1).

Only a few of the methods have been applied somewhat more widely and no single method can be considered as the standard analytical approach. From a practical point of view, a "good" method should: i) be able to account for the full exposure history in as much detail as possible; ii) produce interpretable results; and iii) be relatively easy to use for non-experts. Among the collected ETR models in this inventory, we would like to highlight two approaches, namely exposure rate models and distributed lag models, that show relatively better utility compared with the others. It is worth noting that method

selection is highly study-specific and should be primarily dependent on the research goals and features of data. In the following section we provide brief descriptions for the two models.

Exposure rate models

Researchers may indirectly assess the effect of temporal factors on the ETR by investigating whether secondary exposure indices (e.g., exposure duration or exposure intensity) or other temporal factors (e.g., time-since-last/first exposure, age-at-first exposure) modify the estimated effect of CE. Such approaches were mostly applied to understand the dynamics aspects of smoking on lung cancer risk (Vlaanderen et al., 2014, Lubin et al., 2007, Lubin and Caporaso, 2006). Using $x_i(t)$ to describe possibly time-varying exposure for subject *i*, and with notation similar to that in Gasparrini (2014), we can describe the excess (absolute or relative) risk at some time *t* as a (possibly non-linear) function $s(.)$ of $x_i(t)$ that is parametrized by an (unknown) vector of parameters η. As an example, for a linear effect of cumulative exposure on the log-hazard rate, $s(.)$ could be defined as:

$$s(x_i(.), t, \eta) = exp\,(\eta \int_0^t x_i(u)\,du) \qquad \text{equation (1)}$$

where $\int_0^t x_i(u)\,du$ corresponds to the usual definition of cumulative exposure. Without loss of generalization and to conform to most practical implementations, we will focus on a time-discretized version of this model for the remainder:

$$s(x_i(.), t, \eta) = f(\eta \sum_{j=0}^{t} x_i(j)) \qquad \text{equation (2)}$$

One way to investigate the effects of different exposure patterns on estimated excess risks is by allowing the slope coefficient for cumulative exposure in this model to depend on individual-level exposure characteristics:

$$s(x_i(.), t, g(.), z_i, \eta) = f(g(z_i, \eta) \sum_{j=0}^{t} x_i(j)) \qquad \text{equation (3)}$$

where $g(.)$ could be a fully parametric or semi-parametric model and z_i the average intensity or time since last exposure. An extension of this model was suggested by Richardson et al. (2012) to account for both between- and within-person variation in exposure over time. A general formulation of the model is given as:

$$s(x_i(.), g(.), t, \eta) = f(\sum_{j=0}^{t} g(x_i(j), \eta) \qquad \text{equation (4)}$$

The proposed model can be implemented using SAS PROC NLP software with codes provided in the paper (Richardson et al., 2012) and they provide a concrete example where the effect of cumulative exposure is modified by exposure intensity as follows:

$$s(x_i(.), t, \eta) = f(\sum_{j=0}^{t} \eta_1 \exp(\eta_2 x_i(j)) * x_i(j)) \qquad \text{equation (5)}$$

Note the similarity to equation (4) when individual exposure is the same for each exposure period. The model was applied to study the effect of radon exposure on lung cancer mortality with a cohort of uranium miners (Richardson et al., 2012). Exposure rate models refine the well-established regression model to efficiently account for exposure increments and their potential impact on disease risk over time by more fully incorporating available information on time-related variation in exposure (Vermeulen and Chadeau-Hyam, 2012), provided that such detailed exposure information is available. It is likely to be useful for researchers that want to understand how differences in exposure intensity may affect the estimated excess risk per unit of exposure.

Distributed lag models

A general model that considers the possible different effects of exposures accrued at different timepoints can be formulated as follows:

$$s(x_i(.), t, w(), \eta) = f(\sum_{j=0}^{t} g(x_i(j), w(x_i(j), \eta))) \qquad \text{equation (6)}$$

with $w()$ a suitably specified weighing function. Assuming weights that only depend on the timepoint and exposure effects that depend linearly on the exposure intensity this model can be simplified as:

$$s(x_i(.), t, w(), \eta) = f(\sum_{j=0}^{t} w(j, \eta) x_i(j)) \qquad \text{equation (7)}$$

With a limited number of discrete time periods and relatively low between-period exposure correlations, the model can be fitted using standard software that does not put any restrictions on the time-specific regression weights, but in most other cases regularization will be needed to avoid imprecise and unstable effect estimates (Schildcrout and Heagerty, 2005). This can be achieved using basis expansion methods and software that allows for penalized estimation.

This model is most often used to investigate exposure latency, which requires defining the timepoints in terms of the number of time units (e.g., years) preceding the health outcome assessment, i.e.:

$$s(x_i(.), l, w(), \eta) = f(\sum_{l=0}^{t} w(l, \eta) x_i(t - l)) \quad \text{equation (8)}$$

This model is called the dynamic lag model and was originally developed in econometrics (Almon, 1965) but has also become popular for studying possible delayed health effects associated with environmental/occupational risk factors (Muggeo and Hajat, 2009, McClean et al., 2007, Jung et al., 2022, Gasparrini, 2014, Schwartz, 2000).

The assumption of a linear effect of exposure on the outcome can be relaxed, by using a similar re-parametrisation of the model structure shown in equation (8), resulting in the class of distributed lag non-linear models (DLNMs). An extensive overview of the different model structures and approaches to estimation is provided by (Gasparrini et al., 2010).

Since its development, it has been used relatively widely to assess relationships between environmental exposures and possibly delayed effects. For example, with an occupational cohort, researchers applied DLNM to explore the temporal associations between silica exposure and lung cancer mortality (Neophytou et al., 2018). The interpretation of results from DLNM models is made easier thanks to the well-developed R package *dlnm* (Gasparrini, 2011), that includes many built-in visualisations including a 3D graph to show the full estimated non-linear exposure-lag-response surface.

4. Summary

Between the two modelling structures, exposure rate modelling could be a useful tool to account for the interplay between exposure intensity and duration (the first proposed research objective), and the DLM framework could incorporate exposure lags and age at exposure, and showed relatively good applicability and flexibility to be applied in evaluating the effect of temporal factors (the second/third research objective).

We also noted that the multistage models, with the purpose of evaluating the effect of time-varying exposures on different stages of disease development (the fourth research objective), usually require detailed data on different disease status, and the type of outcome with the included biological models is usually restricted to cancer (e.g., with 2SCE). Moreover, only very few studies have applied biology-based models in occupational epidemiologic analyses, making it very difficult to judge their applicability for the purpose of this report. Interested readers could refer to Moolgavkar and Luebeck (2020) for more detailed introduction.

Overall, this report provides a background in ETR modelling, an inventory of ETR models, and a brief discussion of the most relevant model structures. Based on this report, researchers have an overview of

ETR methods that are listed in Table 1 and select the appropriate one(s) based on different research goals and dataset features. We refrained from providing more general recommendations because of the complex nature of ETR modelling and the limited number of studies in which these methods were used or validated for occupational epidemiology. Validation of the utility of DLM models using real-world datasets will be carried out in 2023 and will likely provide more insights into the ways ETR models could be applied within the scope of EPHOR and the working-life exposome.

Table 1. Inventory of the Exposure-Time-Response (ETR) models

Model type	Algorithm	Goal	Reference (DOI)	First Author	Implementation	Comments
Exposure Rate Models	Maximum Likelihood	Modification of cumulative exposure effects by duration and timing of exposure	10.1097/EDE.0b013e31826c3149	Richardson, D.B.	SAS script	
Linear Distributed Lag Model	Piecewise constant/Bilinear/Exponential decay	Latency analysis/Critical exposure windows for a single exposure	10.1002/(sici)1097-0274(199903)35:3<246::aid-ajim4>3.0.co;2-6; 10.1136/oem.2004.017368; 10.1007/s10654-020-00658-9	Langholz, B.	Epicure scripts available on request	
	Hierarchical	Latency analysis/Critical exposure windows for a single exposure	10.1093/aje/kwq387	Richardson, D.B.	SAS script	
	Regression spline + constrained optimization	Latency analysis/Critical exposure windows for a single exposure	10.1111/j.0006-341x.2000.01105.x; 10.1002/sim.3701; 10.1136/oemed-2016-104133	Hauptmann, M.	-	
	BKMR	Latency analysis/Critical exposure windows for a single exposure	10.48550/arXiv.1904.12417	Wilson, A.	R package regimes (Github)	Only allows for gaussian outcomes
	Bayesian	Latency analysis/Critical exposure windows for a single exposure, allowing for interaction with a categorical variable	10.1093/biostatistics/kxx002	Wilson, A.	R package regimes (Github)	Only allows for gaussian outcomes
	Bayesian+postprocessing to identify critical window	Similar to a tree-based approach	10.1097/EDE.0000000000001428	Johnson, M.	-	Not much detail provided on how the method is implemented.
	Combination of mixed model & spline WCIE	Standard WCIE approach with modelled exposure in a longitudinal stiudy design for the outcome	10.1186/s12874-021-01403-w	Wagner, M.	R code (Github)	The WCIE is estimated using standard natural regression splines
Non-Linear Distributed Lag Model	Tensor-product splines	Identify critical exposure window of single exposure, allowing for non-linear exposure effect	10.1002/sim.3354	Berhane, K.	Epicure scripts available on request	
	Cross-basis regression splines		10.1002/sim.3940; 10.18637/JSS.V043.I08	Gasparrini, A.	R package dlnm (CRAN)	
	Cross-basis penalized splines		10.1111/biom.12645	Gasparrini, A.	R package dlnm (CRAN)	
	Functional model		10.48550/arXiv.2103.12822	Lenart, P.	uses package fda	Only allows for gaussian outcomes. Package refund allows binomial outcomes and seems easier.
Distributed Lag Mixture Models	Random forest	Identify critical exposure window of single exposure, allowing for non-linear exposure effect	10.1111/biom.13568	Mork, D.	R package dlmtree (CRAN)	Allows for gaussian outcomes

Model type	Algorithm	Goal	Reference (DOI)	First Author	Implementation	Comments
	Shared trees/Gaussian process	Identify critical exposure window of single exposure, allowing for between-subject heterogeneity	10.48550/arXiv.2109.13763	Mork, D.	R package dlmtree (CRAN)	
	Bayesian	Identify critical exposure windows of 2 exposures, allowing for non-linear exposure effect and interactions	10.1111/rssc.12297	Yin-Hsiu Chen	R package glm+rjags (CRAN)	
	Gaussian Processes	Variable selection	10.1093/biostatistics/kxz006	Warren, J.L.	R package CWVSmix (Github)	Needs extensive tuning
	Bayesian	Estimate exposure-response functions and identify critical exposure windows of chemical mixtures	arXiv:2107.14567	Antonelli, J.	R package BayesianDLAG (Github)	Only allows for gaussian outcomes
	BKMR	Identify critical exposure windows of chemical mixtures, allowing for non-linear exposure effect and interactions	10.1093/biostatistics/kxx036	Liu, S.H.	R code (Github)	Only allows for gaussian outcomes. Code needs substantial work.
	BKMR	Identify critical exposure windows of chemical mixtures, allowing for non-linear exposure effect and interactions	10.1214/21-AOAS1533	Wilson, A.	R package regimes (Github)	Only allows for gaussian outcomes. Default output only includes a plot of the weight function.
Joint models	Bayesian	Include weighted cumulative effect in joint models	10.1002/sim.7385; 10.1002/sim.7027; 10.18637/jss.v072.i07	Mauff, K.	R package JMbayes (CRAN)	
	Two-stage models + latent class	Identify exposure trajectories associated with risk	10.1371/journal.pone.0236736	Lévêque, E.	R package lcmm (CRAN)	
Multi-stage	Two-stage clonal expansion	Evaluate effect of exposure on different stages of carcinogenesis	10.1093/aje/kwn284	Richardson, D.B.	SAS script	
	Multi-stage clonal expansion		10.1093/aje/kwn285	Richardson, D.B.	package msce (CRAN)	
Multi-state models	Hidden Markov Model	Evaluate the effect of covariates (exposures) on the transition of different disease states	10.1097/EDE.0000000000000032	Chadeau-Hyam, M.	C++ code	

References

ALMON, S. 1965. The Distributed Lag between Capital Appropriations and Expenditures. *Econometrica,* 33**,** 178-196.

AMMAR, W., GROENEVELD, D., BHAGAVATULA, C., BELTAGY, I., CRAWFORD, M., DOWNEY, D., DUNKELBERGER, J., ELGOHARY, A., FELDMAN, S., HA, V. A., KINNEY, R. M., KOHLMEIER, S., LO, K., MURRAY, T. C., OOI, H.-H., PETERS, M. E., POWER, J. L., SKJONSBERG, S., WANG, L. L., WILHELM, C., YUAN, Z., VAN ZUYLEN, M. & ETZIONI, O. Construction of the Literature Graph in Semantic Scholar. NAACL, 2018.

ARMITAGE, P. & DOLL, R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer,* 8**,** 1-12.

BASAGANA, X. & BARRERA-GOMEZ, J. 2022. Reflection on modern methods: visualizing the effects of collinearity in distributed lag models. *Int J Epidemiol,* 51**,** 334-344.

BERENBLUM, I. & SHUBIK, P. 1947. A New, Quantitative, Approach to the Study of the Stages of Chemical Carcinogenesis in the Mouse's Skin. *British Journal of Cancer,* 1**,** 383-391.

BROWN, C. C. & CHU, K. C. 1983. Implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. *J Natl Cancer Inst,* 70**,** 455-63.

BUCKLEY, J. P., HAMRA, G. B. & BRAUN, J. M. 2019. Statistical Approaches for Investigating Periods of Susceptibility in Children's Environmental Health Research. *Current Environmental Health Reports,* 6**,** 1-7.

CHADEAU-HYAM, M., TUBERT-BITTER, P., GUIHENNEUC-JOUYAUX, C., CAMPANELLA, G., RICHARDSON, S., VERMEULEN, R., DE IORIO, M., GALEA, S. & VINEIS, P. 2014. Dynamics of the risk of smoking-induced lung cancer: a compartmental hidden Markov model for longitudinal analysis. *Epidemiology,* 25**,** 28-34.

CHECKOWAY, H., HEYER, N. J., SEIXAS, N. S., WELP, E. A., DEMERS, P. A., HUGHES, J. M. & WEILL, H. 1997. Dose-response associations of silica with nonmalignant respiratory disease and lung cancer mortality in the diatomaceous earth industry. *Am J Epidemiol,* 145**,** 680-8.

CHECKOWAY, H., PEARCE, N., HICKEY, J. L. & DEMENT, J. M. 1990. Latency analysis in occupational epidemiology. *Arch Environ Health,* 45**,** 95-100.

CHEN, Y. H., FERGUSON, K. K., MEEKER, J. D., MCELRATH, T. F. & MUKHERJEE, B. 2015. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. *Environmental Health,* 14.

CONNECTED PAPERS. 2022. Available: https://www.connectedpapers.com/about [Accessed Sept. 22 2022].

DANIELI, C., COHEN, S., LIU, A., PILOTE, L., GUO, L., BEAUCHAMP, M. E., MARELLI, A. J. & ABRAHAMOWICZ, M. 2019. Flexible Modeling of the Association Between Cumulative Exposure to Low-Dose Ionizing Radiation From Cardiac Procedures and Risk of Cancer in Adults With Congenital Heart Disease. *Am J Epidemiol,* 188**,** 1552-1562.

DAY, N. E. & BROWN, C. C. 1980. Multistage models and primary prevention of cancer. *J Natl Cancer Inst,* 64**,** 977-89.

DE VOCHT, F., BURSTYN, I. & SANGUANCHAIYAKRIT, N. 2015. Rethinking cumulative exposure in epidemiology, again. *Journal of Exposure Science and Environmental Epidemiology,* 25**,** 467-473.

DOTSON, G. S., MAIER, A., SIEGEL, P. D., ANDERSON, S. E., GREEN, B. J., STEFANIAK, A. B., CODISPOTI, C. D. & KIMBER, I. 2015. Setting Occupational Exposure Limits for Chemical Allergens--Understanding the Challenges. *J Occup Environ Hyg,* 12 Suppl 1**,** S82-98.

GASPARRINI, A. 2011. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *J Stat Softw,* 43**,** 1-20.

GASPARRINI, A. 2014. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med,* 33**,** 881-99.

GASPARRINI, A., ARMSTRONG, B. & KENWARD, M. G. 2010. Distributed lag non-linear models. *Stat Med,* 29**,** 2224-34.

HETHCOTE, H. W. & KNUDSON, A. G. 1978. Model for the incidence of embryonal cancers: application to retinoblastoma. *Proceedings of the National Academy of Sciences,* 75**,** 2453-2457.
JACKSON, C. 2011. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software,* 38**,** 1-28.
JACKSON, C., SHARPLES, L., THOMPSON, S., DUFFY, S. & COUTO, E. 2003. Multistate Markov models for disease progression with classification error. *The Statistician,* 52**,** 193-209.
JACOB, P., RÜHM, W., WALSH, L., BLETTNER, M., HAMMER, G. & ZEEB, H. 2009. Is cancer risk of radiation workers larger than expected? *Occupational and Environmental Medicine,* 66**,** 789-796.
JUNG, J., LEE, E. M., MYUNG, W., KIM, H., KIM, H. & LEE, H. 2022. Burden of dust storms on years of life lost in Seoul, South Korea: A distributed lag analysis. *Environ Pollut,* 296**,** 118710.
KIM, S., VERMEULEN, R., WAIDYANATHA, S., JOHNSON, B. A., LAN, Q., SMITH, M. T., ZHANG, L., LI, G., SHEN, M., YIN, S., ROTHMAN, N. & RAPPAPORT, S. M. 2006. Modeling human metabolism of benzene following occupational and environmental exposures. *Cancer Epidemiol Biomarkers Prev,* 15**,** 2246-52.
KNUDSON, A. G., HETHCOTE, H. W. & BROWN, B. W. 1975. Mutation and childhood cancer: a probabilistic model for the incidence of retinoblastoma. *Proceedings of the National Academy of Sciences,* 72**,** 5116-5120.
KRIEBEL, D., CHECKOWAY, H. & PEARCE, N. 2007. Exposure and dose modelling in occupational epidemiology. *Occup Environ Med,* 64**,** 492-8.
KUH, D., BEN-SHLOMO, Y., LYNCH, J., HALLQVIST, J. & POWER, C. 2003. Life course epidemiology. *J Epidemiol Community Health,* 57**,** 778-83.
LACOURT, A., LÉVÊQUE, E., GUICHARD, E., GILG SOIT ILG, A., SYLVESTRE, M. P. & LEFFONDRÉ, K. 2017. Dose-time-response association between occupational asbestos exposure and pleural mesothelioma. *Occup Environ Med,* 74**,** 691-697.
LÉVÊQUE, E., LACOURT, A., PHILIPPS, V., LUCE, D., GUÉNEL, P., STÜCKER, I., PROUST-LIMA, C. & LEFFONDRÉ, K. 2020. A new trajectory approach for investigating the association between an environmental or occupational exposure over lifetime and the risk of chronic disease: Application to smoking, asbestos, and lung cancer. *PLoS One,* 15**,** e0236736.
LUBIN, J. H., CAPORASO, N., WICHMANN, H. E., SCHAFFRATH-ROSARIO, A. & ALAVANJA, M. C. 2007. Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiology,* 18**,** 639-48.
LUBIN, J. H. & CAPORASO, N. E. 2006. Cigarette smoking and lung cancer: modeling total exposure and intensity. *Cancer Epidemiol Biomarkers Prev,* 15**,** 517-23.
MADATHIL, S., JOSEPH, L., HARDY, R., ROUSSEAU, M. C. & NICOLAU, B. 2018. A Bayesian approach to investigate life course hypotheses involving continuous exposures. *Int J Epidemiol,* 47**,** 1623-1635.
MAUFF, K., STEYERBERG, E. W., NIJPELS, G., VAN DER HEIJDEN, A. & RIZOPOULOS, D. 2017. Extension of the association structure in joint models to include weighted cumulative effects. *Stat Med,* 36**,** 3746-3759.
MCCLEAN, M. D., RINEHART, R. D., SAPKOTA, A., CAVALLARI, J. M. & HERRICK, R. F. 2007. Dermal exposure and urinary 1-hydroxypyrene among asphalt roofing workers. *Journal of Occupational and Environmental Hygiene,* 4**,** 118-126.
MOOLGAVKAR, S. & LUEBECK, G. 2020. Multistage carcinogenesis: a unified framework for cancer data analysis. *Statistical Modeling for Biological Systems.* Springer.
MOOLGAVKAR, S. H. & VENZON, D. J. 1979. Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Mathematical Biosciences,* 47**,** 55-77.
MORK, D. & WILSON, A. 2021. Estimating perinatal critical windows of susceptibility to environmental mixtures via structured Bayesian regression tree pairs. *Biometrics*.
MUGGEO, V. M. & HAJAT, S. 2009. Modelling the non-linear multiple-lag effects of ambient temperature on mortality in Santiago and Palermo: a constrained segmented distributed lag approach. *Occupational and Environmental Medicine,* 66**,** 584-591.
NEOPHYTOU, A. M., PICCIOTTO, S., BROWN, D. M., GALLAGHER, L. E., CHECKOWAY, H., EISEN, E. A. & COSTELLO, S. 2018. Exposure-Lag-Response in Longitudinal Studies:

Application of Distributed-Lag Nonlinear Models in an Occupational Cohort. *Am J Epidemiol,* 187**,** 1539-1548.
PETERSEN, A. H., OSLER, M. & EKSTROM, C. T. 2021. Data-Driven Model Building for Life-Course Epidemiology. *Am J Epidemiol,* 190**,** 1898-1907.
RICHARDSON, D. B. 2008. Temporal variation in the association between benzene and leukemia mortality. *Environ Health Perspect,* 116**,** 370-4.
RICHARDSON, D. B. 2009. Multistage modeling of leukemia in benzene workers: a simple approach to fitting the 2-stage clonal expansion model. *Am J Epidemiol,* 169**,** 78-85.
RICHARDSON, D. B., COLE, S. R. & LANGHOLZ, B. 2012. Regression models for the effects of exposure rate and cumulative exposure. *Epidemiology,* 23**,** 892-9.
RICHARDSON, D. B., MACLEHOSE, R. F., LANGHOLZ, B. & COLE, S. R. 2011. Hierarchical latency models for dose-time-response associations. *Am J Epidemiol,* 173**,** 695-702.
ROTHMAN, K. J. 1981. Induction and latent periods. *Am J Epidemiol,* 114**,** 253-9.
SANCHEZ, B. N., HU, H. W., LITMAN, H. J. & TELLEZ-ROJO, M. M. 2011. Statistical Methods to Study Timing of Vulnerability with Sparsely Sampled Data on Environmental Toxicants. *Environmental Health Perspectives,* 119**,** 409-415.
SCHILDCROUT, J. S. & HEAGERTY, P. J. 2005. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics,* 6**,** 633-52.
SCHWARTZ, J. 2000. The distributed lag between air pollution and daily deaths. *Epidemiology,* 11**,** 320-326.
SELEVAN, S. G., KIMMEL, C. A. & MENDOLA, P. 2000. Identifying critical windows of exposure for children's health. *Environ Health Perspect,* 108 Suppl 3**,** 451-5.
THOMAS, D. C. 1988. Models for exposure-time-response relationships with applications to cancer epidemiology. *Annu Rev Public Health,* 9**,** 451-82.
TOMASETTI, C., LI, L. & VOGELSTEIN, B. 2017. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science,* 355**,** 1330-1334.
TOMASETTI, C. & VOGELSTEIN, B. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science,* 347**,** 78-81.
VERMEULEN, R. & CHADEAU-HYAM, M. 2012. Dynamic aspects of exposure history-do they matter? *Epidemiology,* 23 6**,** 900-1.
VINEIS, P., VAN VELDHOVEN, K., CHADEAU-HYAM, M. & ATHERSUCH, T. J. 2013a. Advancing the application of omics-based biomarkers in environmental epidemiology. *Environ Mol Mutagen,* 54**,** 461-7.
VINEIS, P., VAN VELDHOVEN, K., CHADEAU-HYAM, M. & ATHERSUCH, T. J. 2013b. Advancing the application of omics-based biomarkers in environmental epidemiology. *Environmental and Molecular Mutagenesis,* 54**,** 461-467.
VLAANDEREN, J., PORTENGEN, L., SCHUZ, J., OLSSON, A., PESCH, B., KENDZIA, B., STUCKER, I., GUIDA, F., BRUSKE, I., WICHMANN, H. E., CONSONNI, D., LANDI, M. T., CAPORASO, N., SIEMIATYCKI, J., MERLETTI, F., MIRABELLI, D., RICHIARDI, L., GUSTAVSSON, P., PLATO, N., JOCKEL, K. H., AHRENS, W., POHLABELN, H., TARDON, A., ZARIDZE, D., FIELD, J. K., T MANNETJE, A., PEARCE, N., MCLAUGHLIN, J., DEMERS, P., SZESZENIA-DABROWSKA, N., LISSOWSKA, J., RUDNAI, P., FABIANOVA, E., STANESCU DUMITRU, R., BENCKO, V., FORETOVA, L., JANOUT, V., BOFFETTA, P., FORASTIERE, F., BUENO-DE-MESQUITA, B., PETERS, S., BRUNING, T., KROMHOUT, H., STRAIF, K. & VERMEULEN, R. 2014. Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: a flexible method applied to cigarette smoking and lung cancer in the SYNERGY Study. *Am J Epidemiol,* 179**,** 290-8.
WAGNER, M., GRODSTEIN, F., LEFFONDRE, K., SAMIERI, C. & PROUST-LIMA, C. 2021. Time-varying associations between an exposure history and a subsequent health outcome: a landmark approach to identify critical windows. *BMC Med Res Methodol,* 21**,** 266.
WANG, M., LIAO, X., LADEN, F. & SPIEGELMAN, D. 2016. Quantifying risk over the life course - latency, age-related susceptibility, and other time-varying exposure metrics. *Stat Med,* 35**,** 2283-95.

WHITE, E., ARMSTRONG, B. K., SARACCI, R., WHITE, E., ARMSTRONG, B. K. & SARACCI, R. 2008. Exposure measurement. *Principles of Exposure Measurement in Epidemiology: Collecting, Evaluating, and Improving Measures of Disease Risk Factors.* Oxford University Press.

ZEKA, A., GORE, R. & KRIEBEL, D. 2011. The two-stage clonal expansion model in occupational cancer epidemiology: results from three cohort studies. *Occup Environ Med,* 68**,** 618-24.