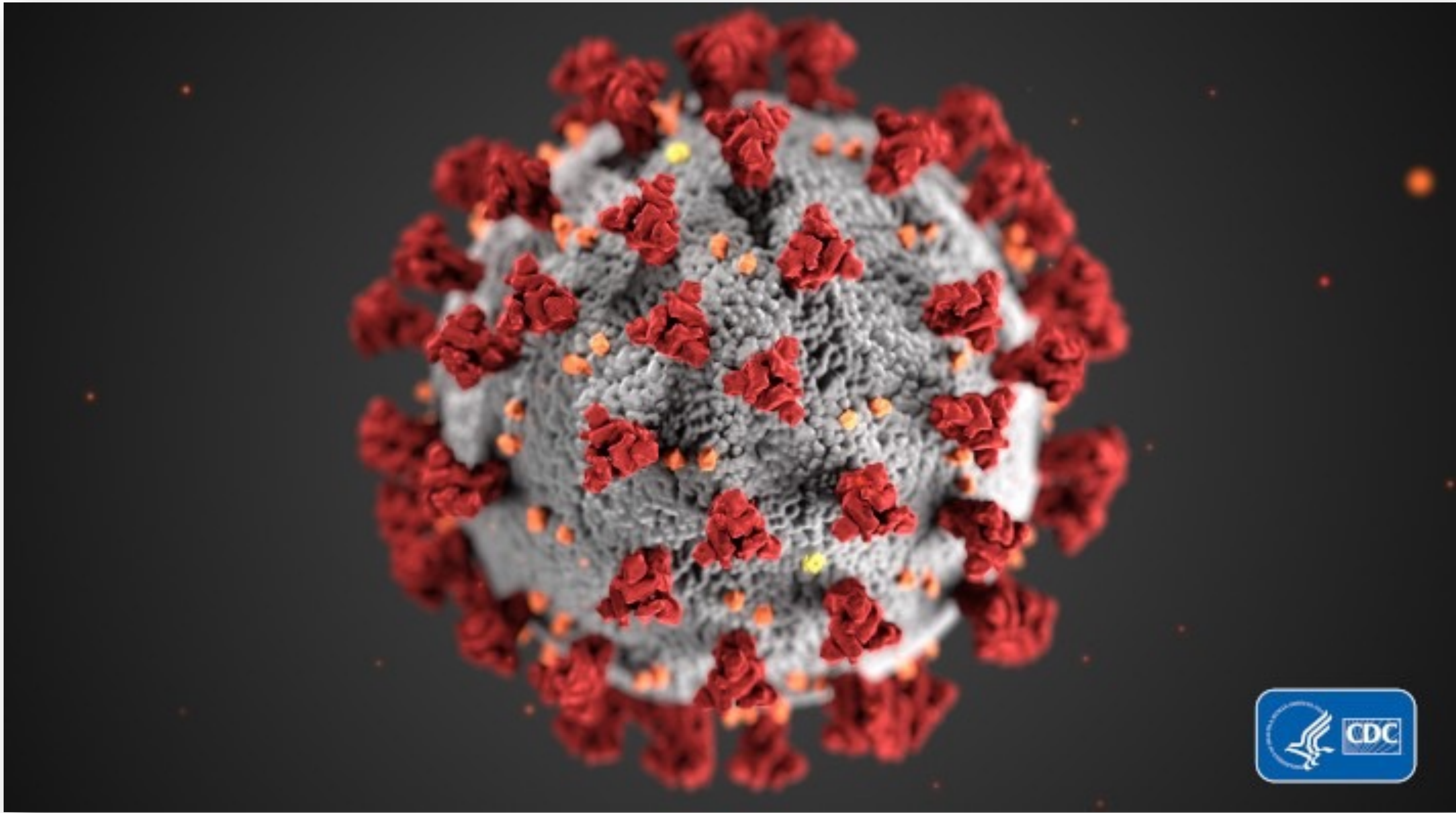
Seminar INPUT platform – 25th March 2022

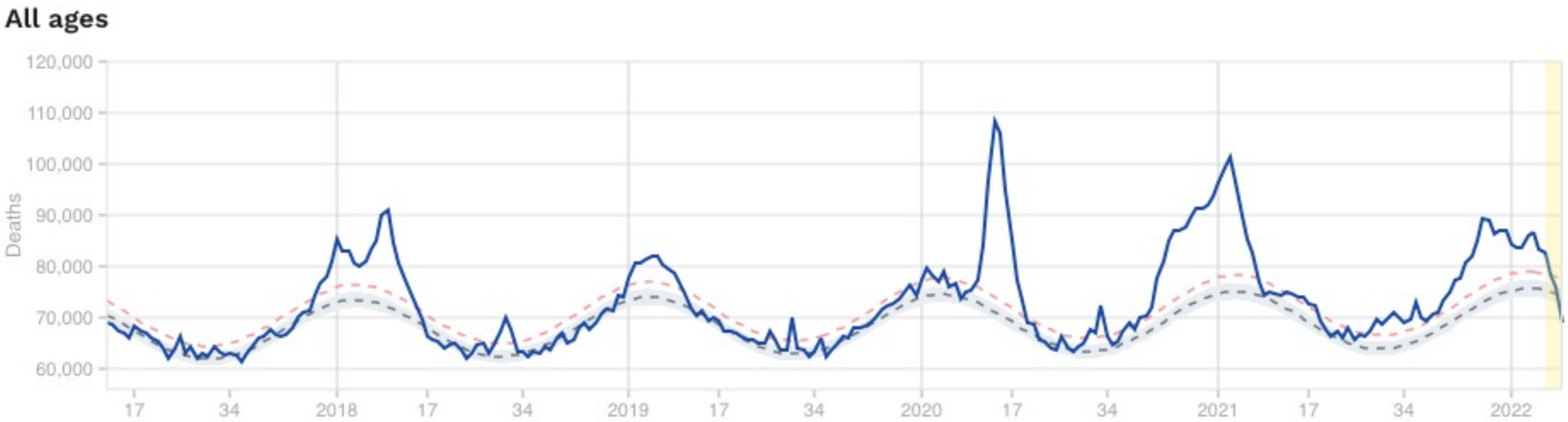
UNIVERSITÄT
BERN

Applications of time series analysis in public health research

Ana M. Vicedo-Cabrera
Head of research group – Climate Change & Health
Institute of Social and Preventive Medicine
Oeschger Center for Climate Change Research
University of Bern

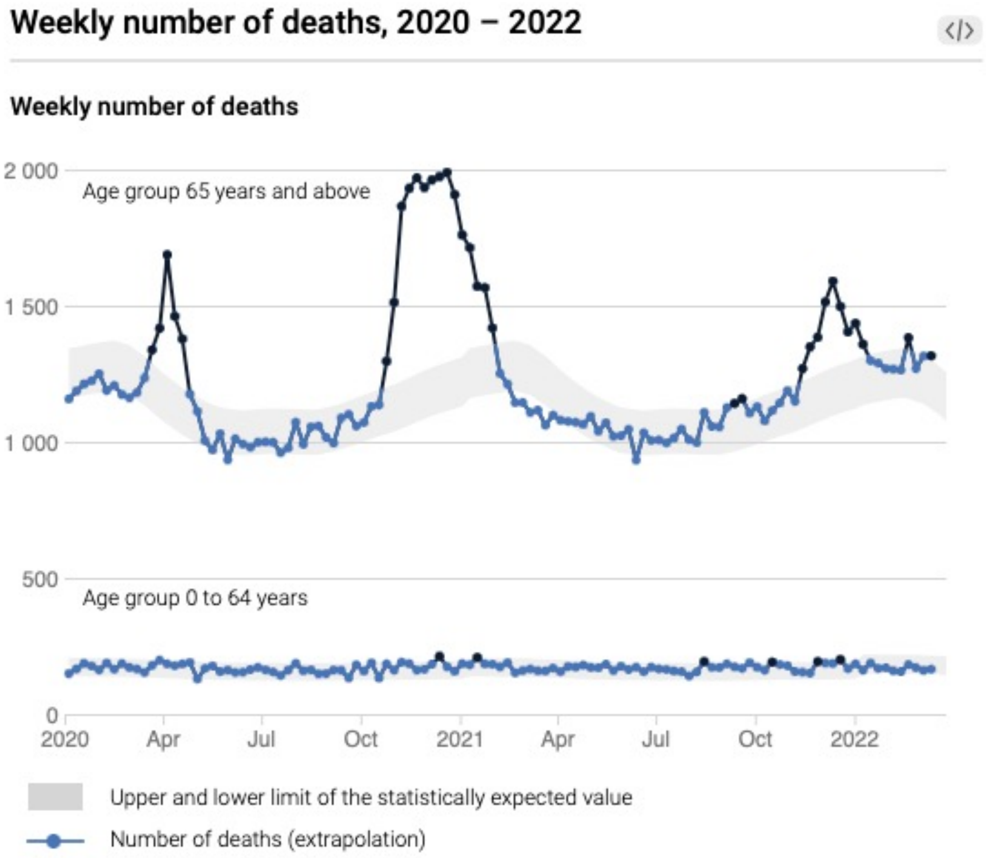

CDC

COVID19 - Excess mortality



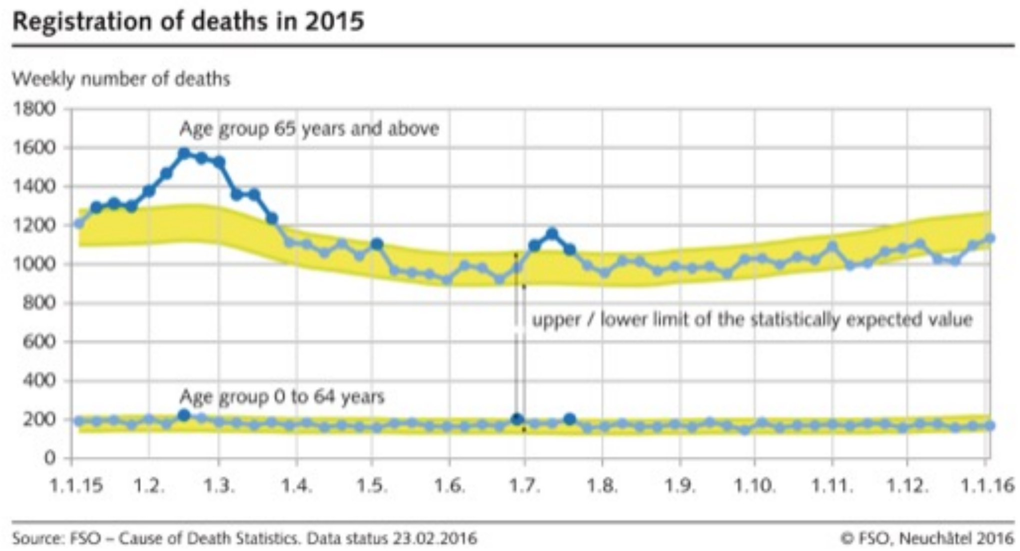
https://www.euromomo.eu/graphs-and-maps/

COVID19 - Excess mortality



https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/graphs.assetdetail.22044006.html

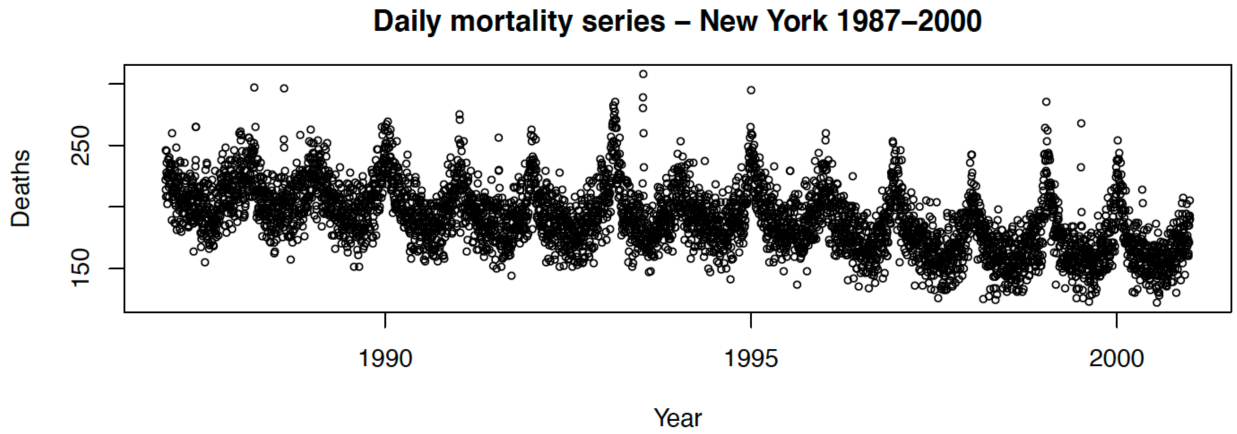
Excess mortality during extreme heat events

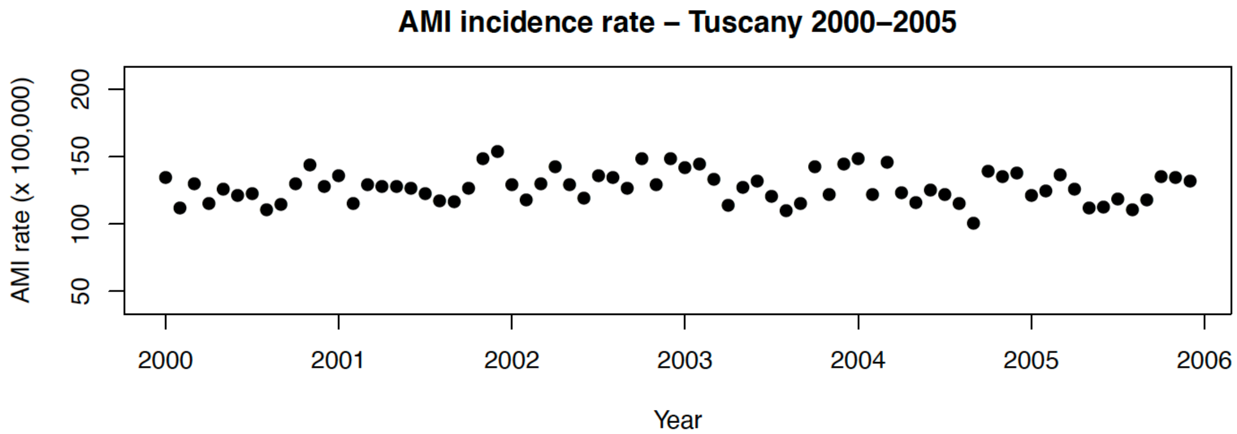


<u>Time series</u>: "*a collection of observations sampled at equally-spaced and ordered time points*" (Chateld 2004).

Usual time series plot → initial exploration of the data.

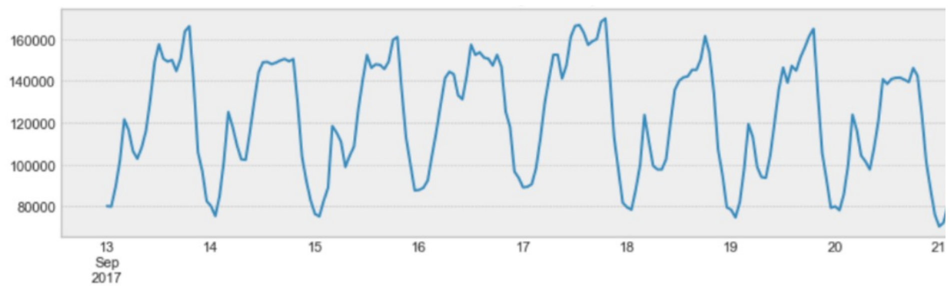
X-axis: time variable
Y-axis: outcome, measure, etc.



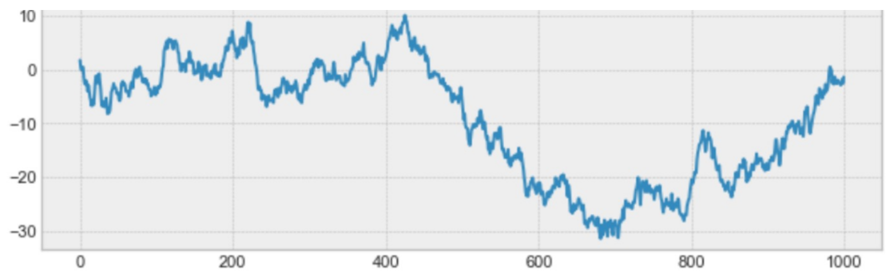


Some characteristics of the TS…

- a finite sample from a *time-series stochastic process*
- each element of the time series is treated as a random variable with a probability distribution
- serial correlation (*autocorrelation*) – similarity between observations as a function of the time between lags.
- *Stationarity* (or not) – if its statistical properties do not change over time (constant mean and variance, and covariance is independent of time)



Example of a stationary process



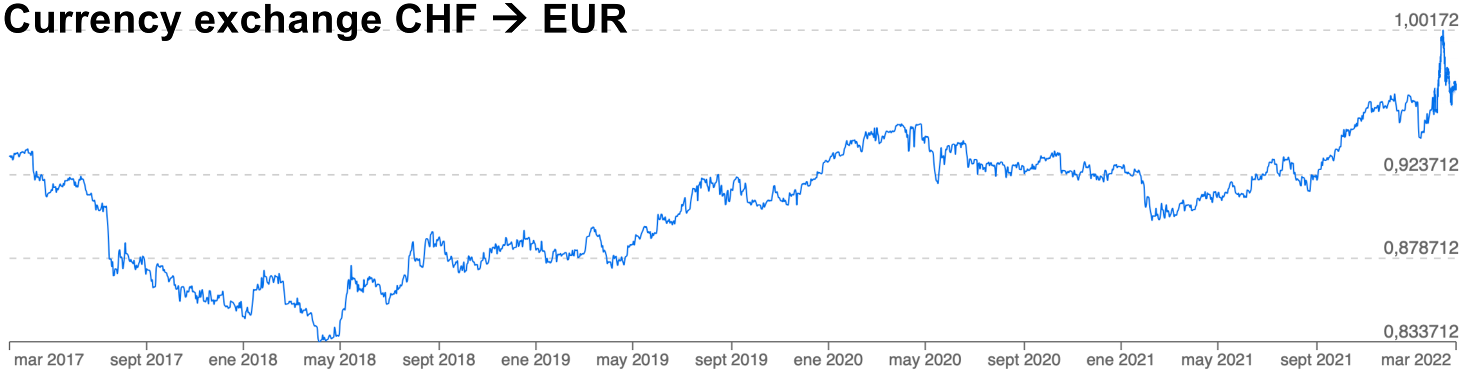
Example of a non-stationary process

Time series analysis

The aim of our statistical analysis is to *use the information contained in the sample to infer properties of the underlying distribution of the time-series process (such as the covariances) from the sample of available observations*.

✓ In their original application (econometrics), time series analysis is applied to forecast future observations given an observed series.

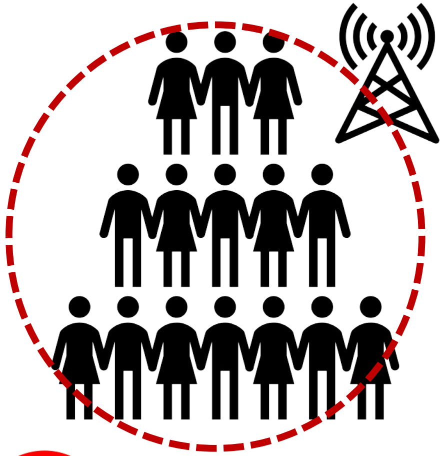
Currency exchange CHF → EUR



✓ In the context of **biomedical research**:

- predict future or estimate past outcomes
- define the association between a predictor and an outcome, both of them showing temporal variability
- assess the impact of interventions (e.g., benefits of public health policies)

Time series analysis – in epidemiology



Unit of analysis: **time**! Not individual.

By design, **time-invariant confounders** are accounted for (yeah!)

Type of **ecological design** – exposure and outcome averaged across a geographical area.

→ limitations: no causality, misclassification bias, limited to short-term effects.

Despite its limitations… **<u>great potential in public health research</u>**:

- Do not require a great amount of (health) data (e.g., individual characteristics)
- Rely on available sources of data (e.g. registries) – cheap, population-based.
- Health impact assessments – quick, flexible, simple.

The statistical characteristics of time series data often *violate the assumptions of conventional statistical methods* → analyzing time series data requires a **unique set of tools and methods**, collectively known as time series analysis

Time series analysis – general modelling approach

Basic regression models – typically, Poisson regression (count data)
Usually with generalized linear models (GLMs) or generalized additive model (GAMs).

$$g(y_t) = \alpha + f(x_t;\boldsymbol{\theta}) + \sum_{j=1}^{J} s_j(t;\gamma_j) + \sum_{p=1}^{P} h_p(z_{pt};\boldsymbol{\eta}) + \epsilon_t$$

function to model the association with hazard x

functions to model time trends (t)

functions controlling for confounders z varying on the same timescale of x

1. Assessment of past trends / prediction of future estimates

Aim: to estimate excess burden (mortality, etc.) – attributed (or not) to a (transient) hazard

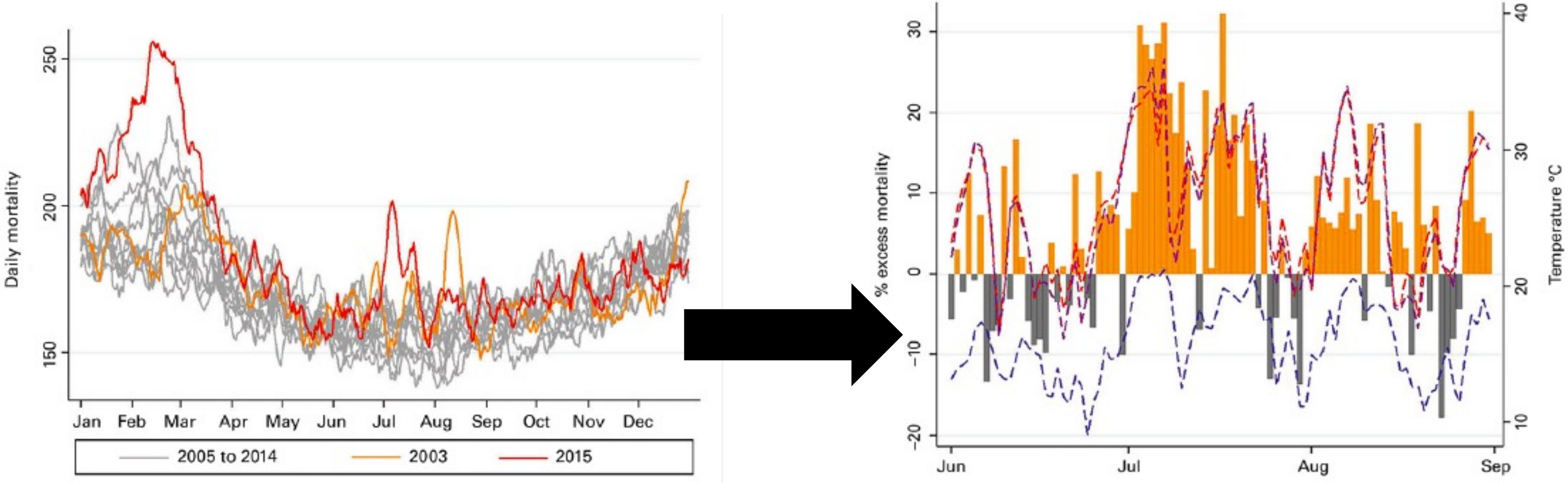
Applications:
→ Monitoring of the incidence of a health outcome in a specific population.
→ Quantification of health burden attributed to a hazards (e.g., COVID-19) – overcome biases in reporting (e.g., underestimation of mortality burden due to COVID-19 at the beginning of the pandemic)

Approach:
Model the underlying temporal trend of an outcome to estimate/predict "what we would have expected at that point in time (in absence of an "event"), based on past observations?"

$$g(y_t) = \alpha + f(x_t;\boldsymbol{\theta}) + \sum_{j=1}^{J} s_j(t;\gamma_j) + \sum_{p=1}^{P} h_p(z_{pt};\boldsymbol{\eta}) + \epsilon_t$$

Note: other methods (less used in epidemiology) are moving average or ARIMA models (and the whole family)

1. Assessment of past trends / prediction of future estimates



"804 excess deaths (5.4%, 95% confidence interval [CI] 3.0-7.9%) were estimated for summer 2015, with the highest percentage obtained in July (11.6%, 95% CI 3.7-19.4%)".

Excess mortality during the warm summer of 2015 in Switzerland. SMW, 2016 – Vicedo-Cabrera et al.

1. Assessment of past trends / prediction of future estimates

Respiratory season	Rec.	Sentinel model		Influenza mortality model			Pneumonia & Influenza mortality model		
		Est.	95% CI	Est.	95% CI	Ratio*	Est.	95% CI	Ratio*
88–89	227	1125	789–1698	597	329–1298	1.88	2017	1250–2867	0.56
89–90	1059	2963	1560–4804	3062	1814–5336	0.97	4648	3061–6747	0.64
90–91	198	799	576–1050	511	301–911	1.56	2108	1375–3321	0.38
91–92	380	1153	845–1653	979	573–1685	1.18	2726	1845–4336	0.42
92–93	309	1243	822–1813	789	376–1737	1.58	2834	1747–4841	0.44
93–94	349	1424	923–2196	906	501–1669	1.57	2841	1713–4335	0.5
94–95	177	1338	829–1905	449	241–874	2.98	2073	1349–3039	0.65
95–96	219	1406	974–1984	593	216–1469	2.37	1683	965–2628	0.84
96–97	318	1958	1148–3112	878	335–1866	2.23	2393	1297–4774	0.82
97–98	363	1903	1198–3080	957	371–2111	1.99	2431	1432–4087	0.78
98–99	260	1969	1295–2823	688	274–1479	2.86	2286	1247–3773	0.86
1988–99	3859	17282	12177–23022	10409	7976–13674	1.66	28040	22959–35043	0.62

* estimates from the Sentinel ILI model divided by those of the current model

"Only counting official influenza deaths underestimated influenza-attributable mortality in Switzerland by a factor of two to three"

Influenza-attributable mortality among the elderly in Switzerland. SMW 2019 Brinkhof et al.

2. Assessment of short-term associations

Aim: to investigate the association between short-term variation in exposure (hazards) and outcome, controlling for long-term trends, seasonal effects and other time-varying variables.
Hazard and outcome → time-varying, share variation on a similar temporal scale.

Applications:
→ To estimate how much the risk of a specific outcome changes according to levels of a exposure.
→ Extensions towards health impact assessments

→ **Approach:**
Model the association using different functions (from simple indicator terms, to complex non-linear functions and accounting for delayed effects).
Temporal decomposition - underlying trends are filtered (smoothed) out from the series, allowing the inspection of associations with variation at a shorter timescale

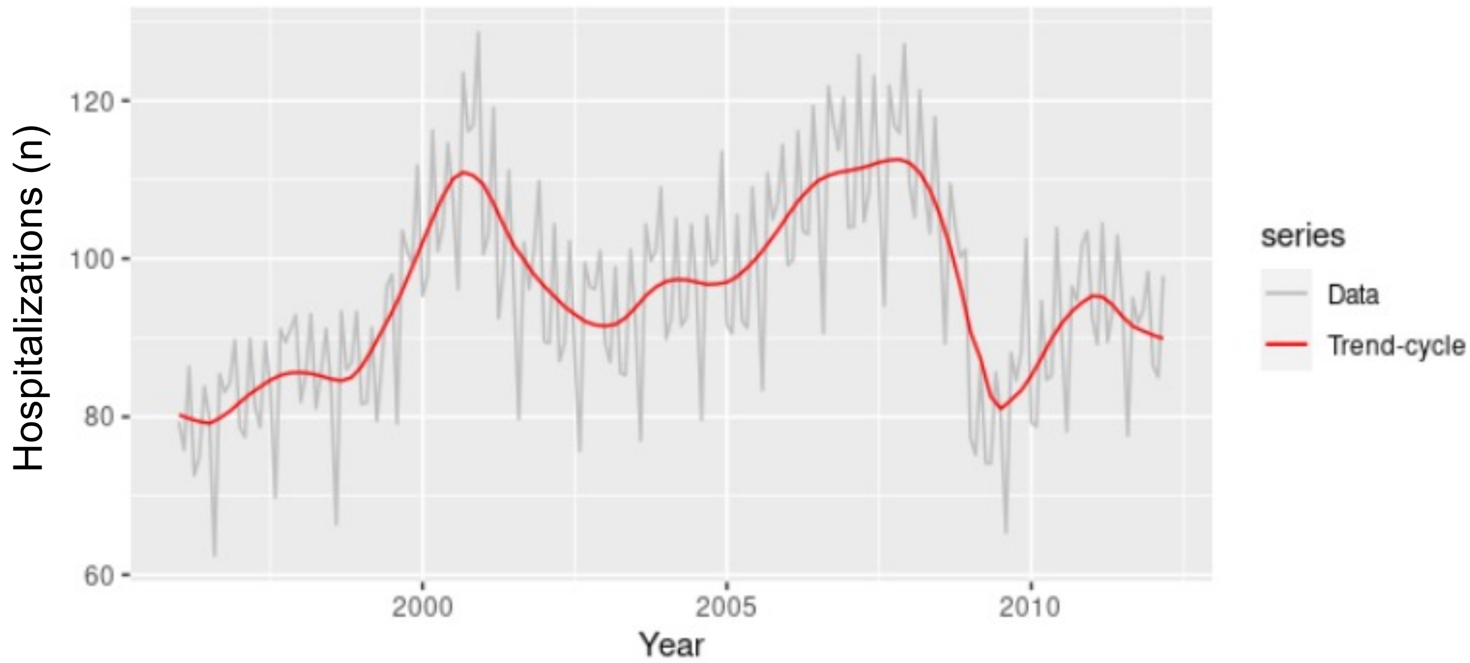
$$g(y_t) = \alpha + f(x_t;\theta) + \sum_{j=1}^{J} s_j(t;\gamma_j) + \sum_{p=1}^{P} h_p(z_{pt};\eta) + \epsilon_t$$

2. Assessment of short-term associations

TEMPORAL DECOMPOSITION

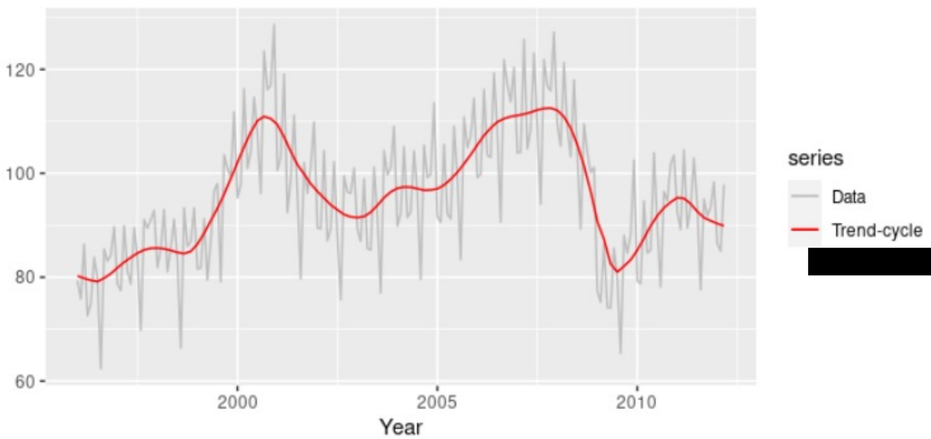
Observed series (almost) always show trends or periodicities → decomposing the variability of the series into components, usually referred to different time scales.

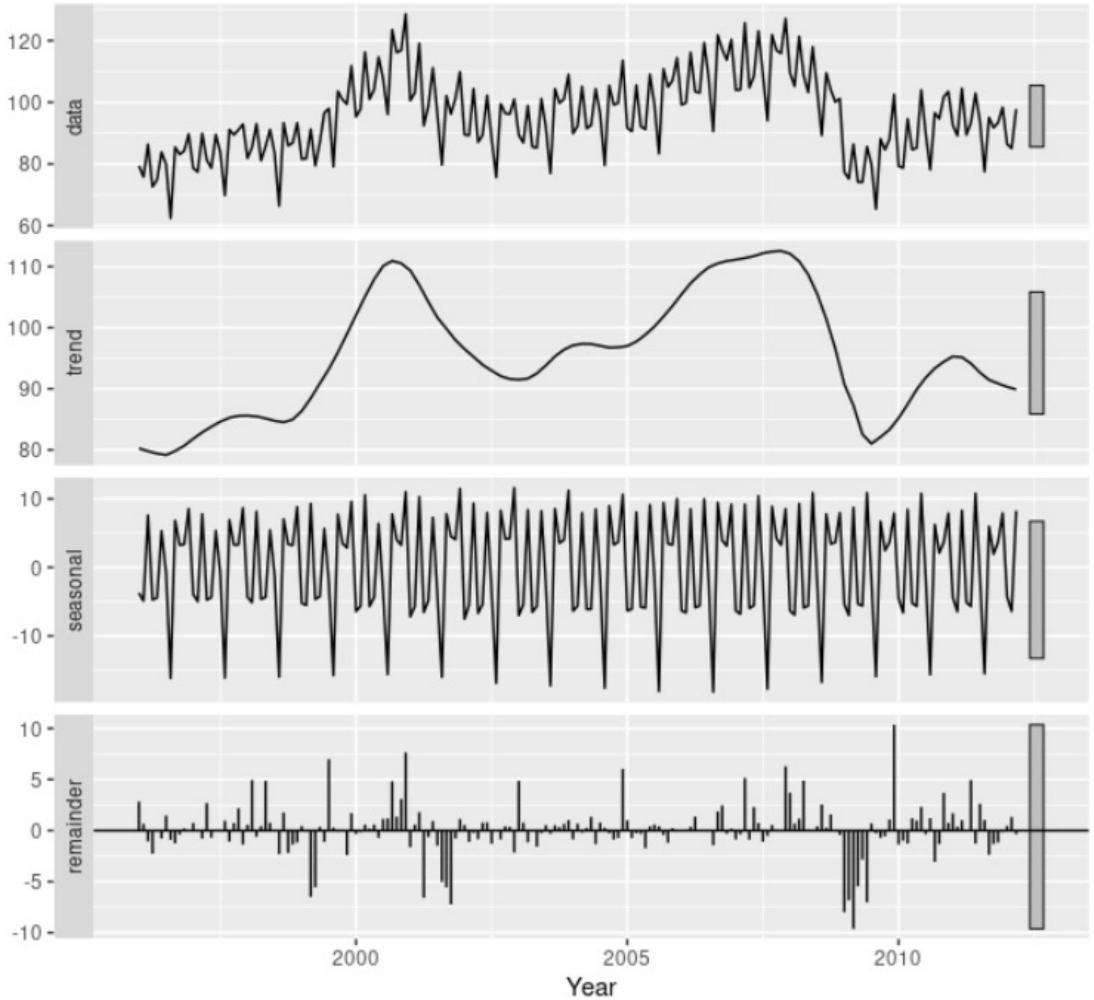
$$y_t = m_t + s_t + w_t$$



2. Assessment of short-term associations

TEMPORAL DECOMPOSITION



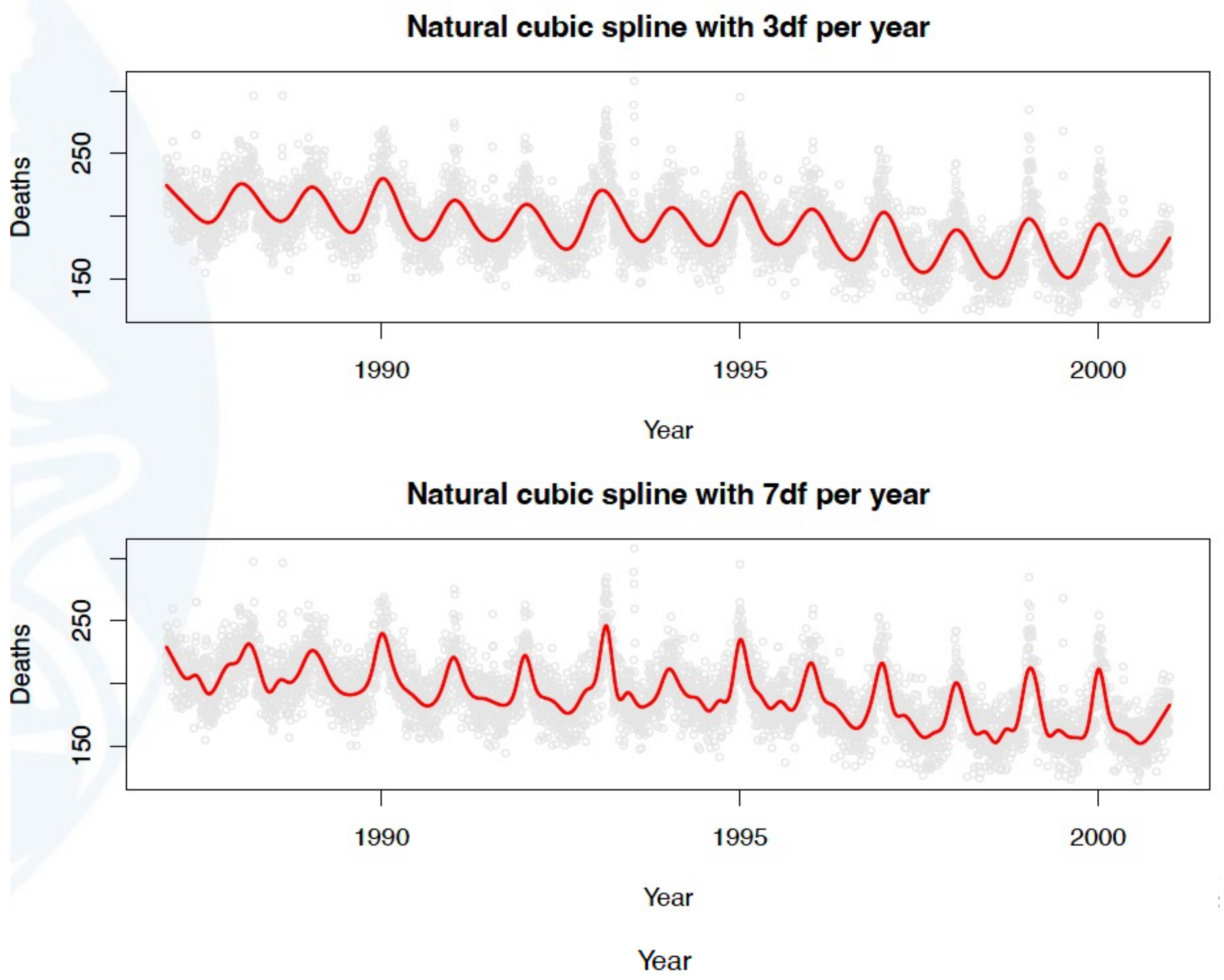


This design exploits decomposition techniques based on regression methods: **the effect of unmeasured confounders varying slowly in time is filtered out.**

2. Assessment of short-term associations

TEMPORAL DECOMPOSITION

- Strata indicators
- Harmonic (Fourier) terms (pairs of sine/cosine functions)
- Splines or other smoothing functions



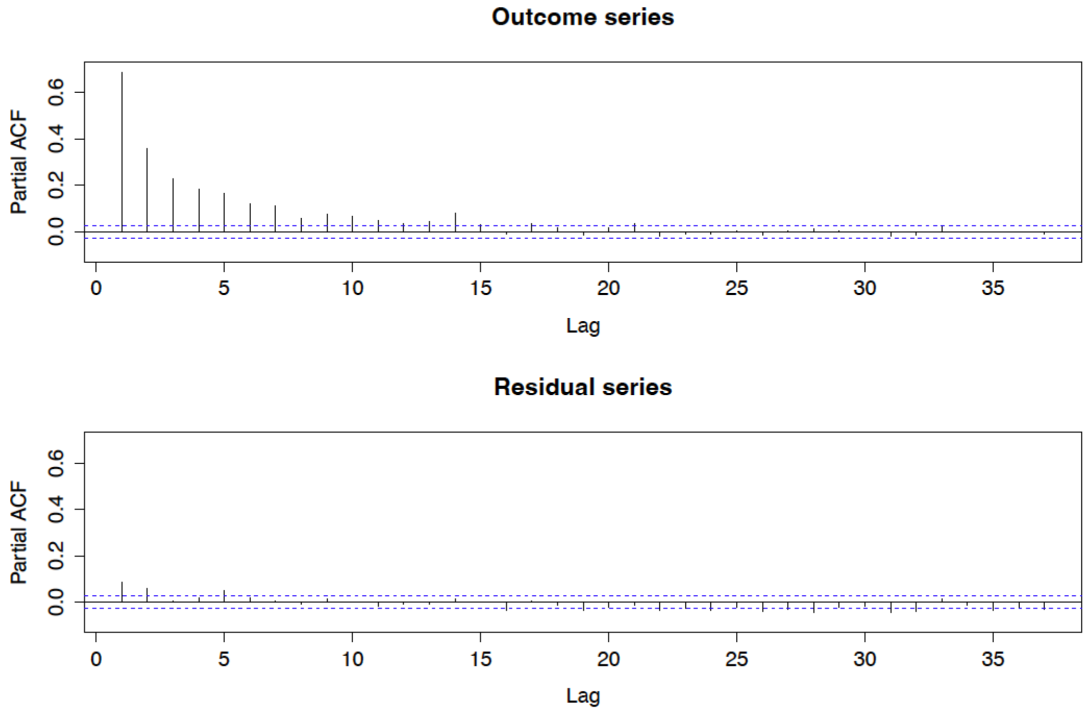
Selection of the appropriate functions
→ Crucial step in the analysis

- Check model fitting (AIC, GCV, etc.)
- Sensitivity analysis
- Check residuals – randomly distributed (PACF)

2. Assessment of short-term associations

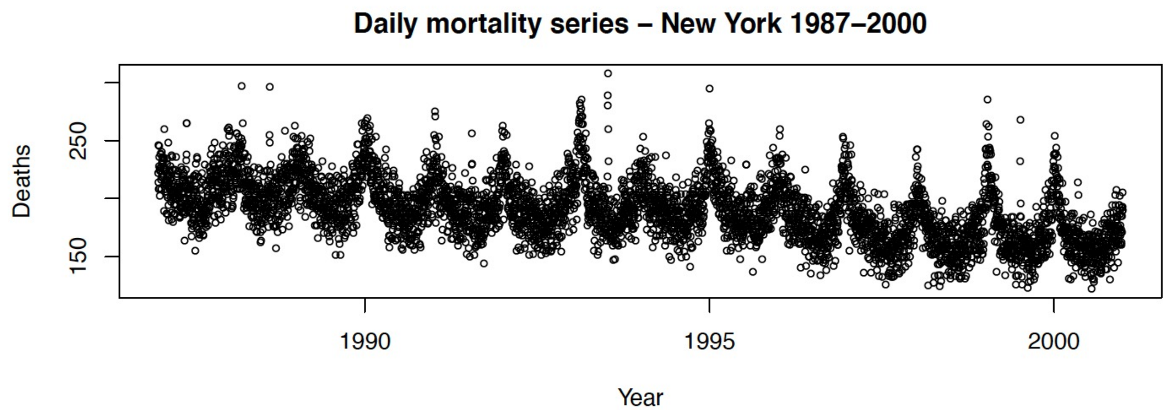
AUTOCORRELATION

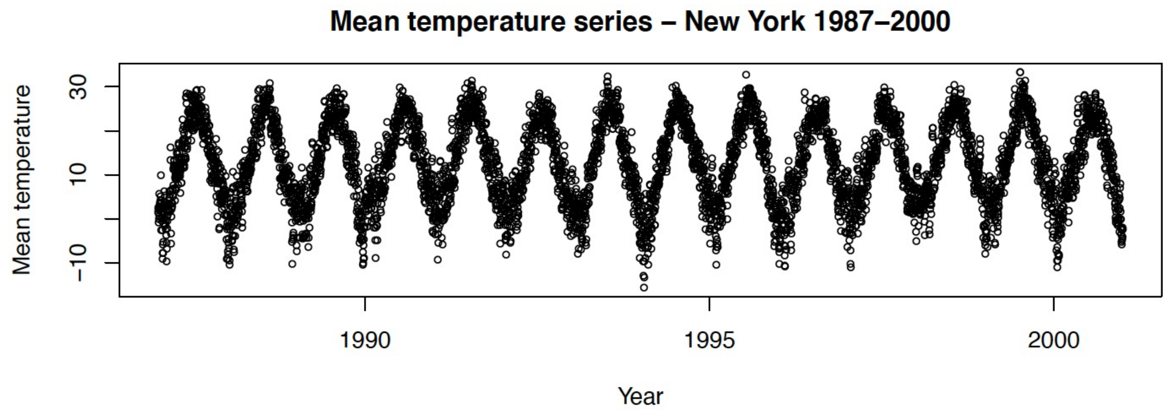
- As in other regression models, the usual assumption is that the residuals are approximately independent and identically distributed.
- Sometimes, a residual correlation may remain between residuals (auto-correlation)
- Exploration of the partial autocorrelation function
- Mostly used in the specification process of a autoregressive model (e.g., ARIMA)



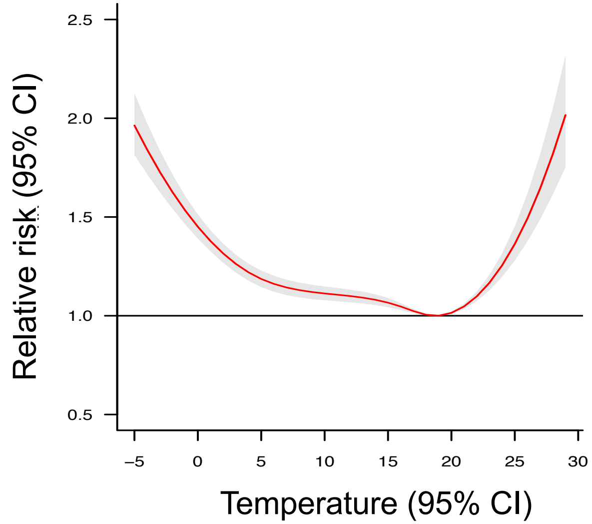
2. Assessment of short-term associations

Environmental exposure – health outcome





Exposure-response association
Temperature-related mortality risk

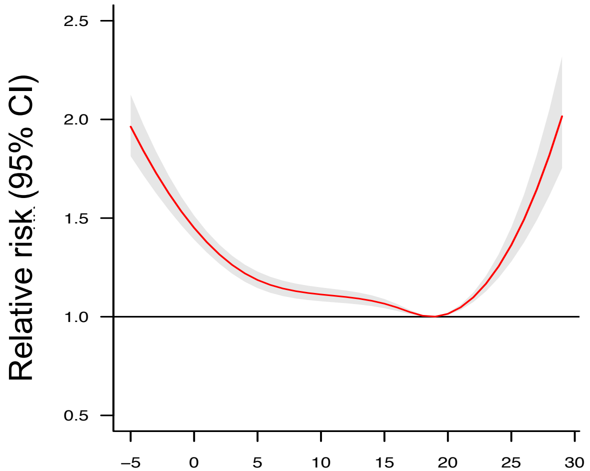


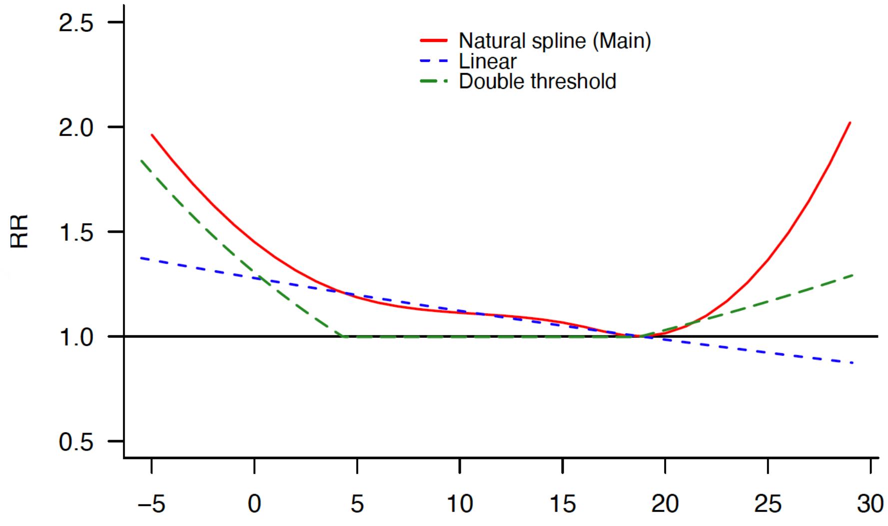
2. Assessment of short-term associations

Environmental exposure – health outcome

Exposure-response association
Temperature-related mortality risk

$$g(y_t) = \alpha + f(x_t; \theta)$$





- Indicator (dummy, categorical)
- Continuous variables – linear or non-linear

\+

Delayed associations?

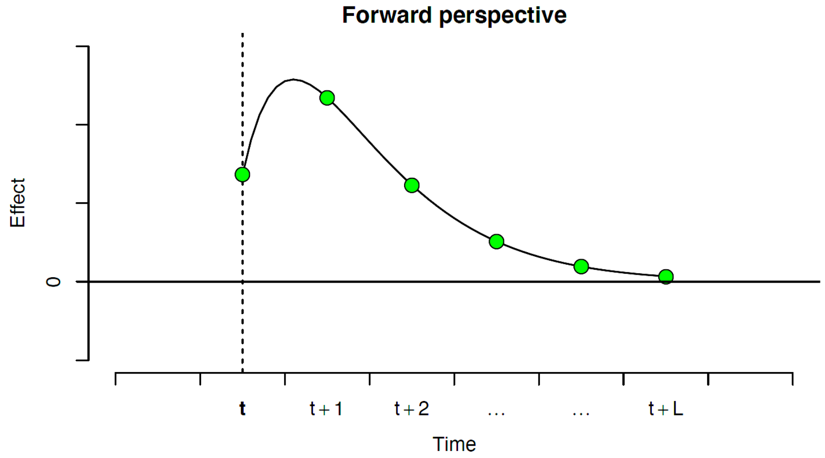
Distributed-lag (non-)linear models

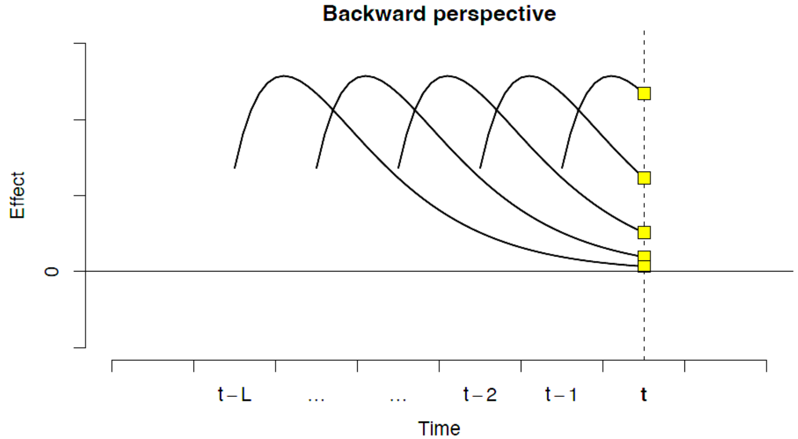
2. Assessment of short-term associations

Environmental exposure – health outcome

Distributed-lag (non-)linear models

- Associations between risk factors and health outcomes are often (always?) characterized by lagged effects
- Challenge: modelling (potentially complex) temporal patterns of risk due to time-varying exposures
- Distributed lag models (DL(N)Ms) represent an elegant methodology to describe such complex relationship
- Originally proposed in econometrics for time series data, then more recently adopted in environmental epidemiology





2. Assessment of short-term associations

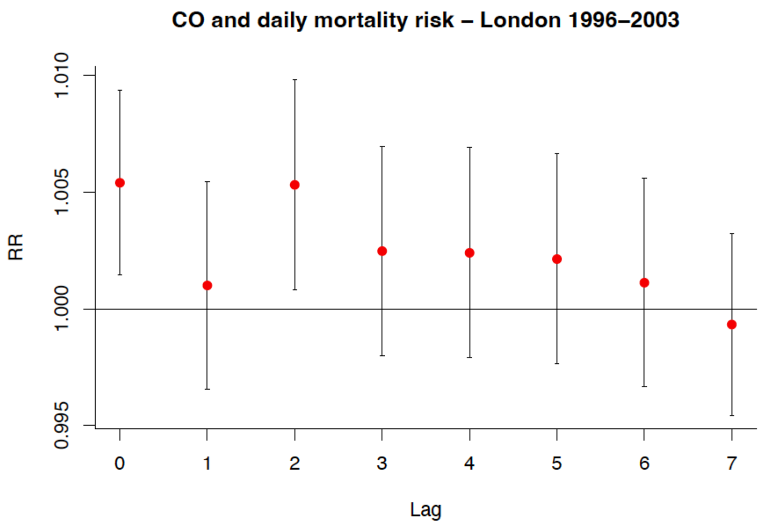
Environmental exposure – health outcome

Distributed-lag (non-)linear models

The main idea is to model the association along lag l in terms of a set of lagged exposures $x_t, \ldots, x_{t-\ell}, \ldots, x_{t-L}$ up to a maximum lag L

The association is parameterized through special cross-basis functions, modelling the exposure-response along x and the lag-response along l

Introduce a function in the lag dimension → any function can be used to model the lag-response (e.g., polynomials, step functions, or splines)

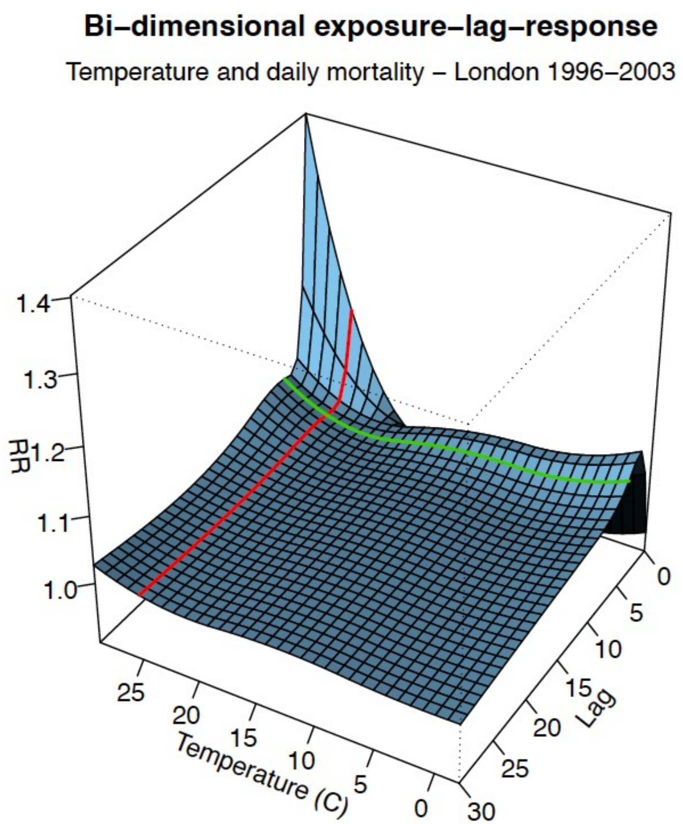


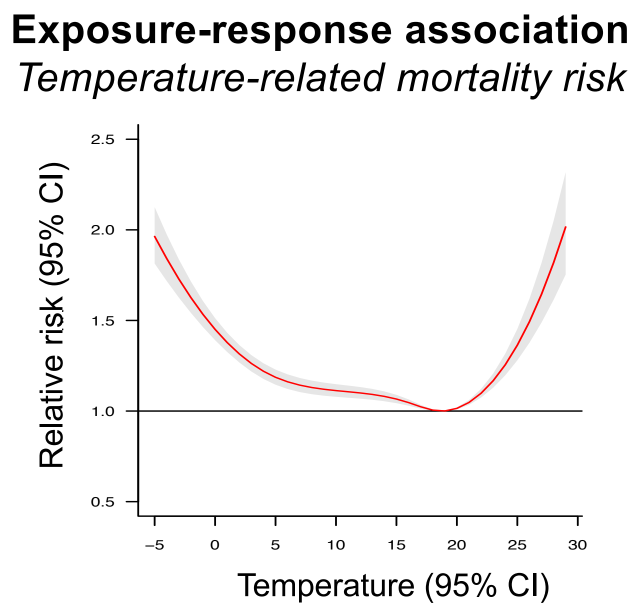
This modelling framework is applied to describe simultaneously (potentially) non-linear and delayed dependencies, termed *exposure-lag-response associations.*

2. Assessment of short-term associations

Environmental exposure – health outcome

Distributed-lag (non-)linear models





- Exposure-response relationships at specific lags
- Lag-response relationships at specific exposure values
- Overall cumulative exposure-response, as the net effect cumulated along lags

2. Assessment of short-term associations

Environmental exposure – health outcome

Distributed-lag (non-)linear models

The framework is fully implemented in the R package dlnm, available from the CRAN since July 2009

Comprehensive documentation through help pages, package vignette, and updated examples from published analyses (see later)

http://www.ag-myresearch.com

Gasparrini et al. 2010 Stats Med

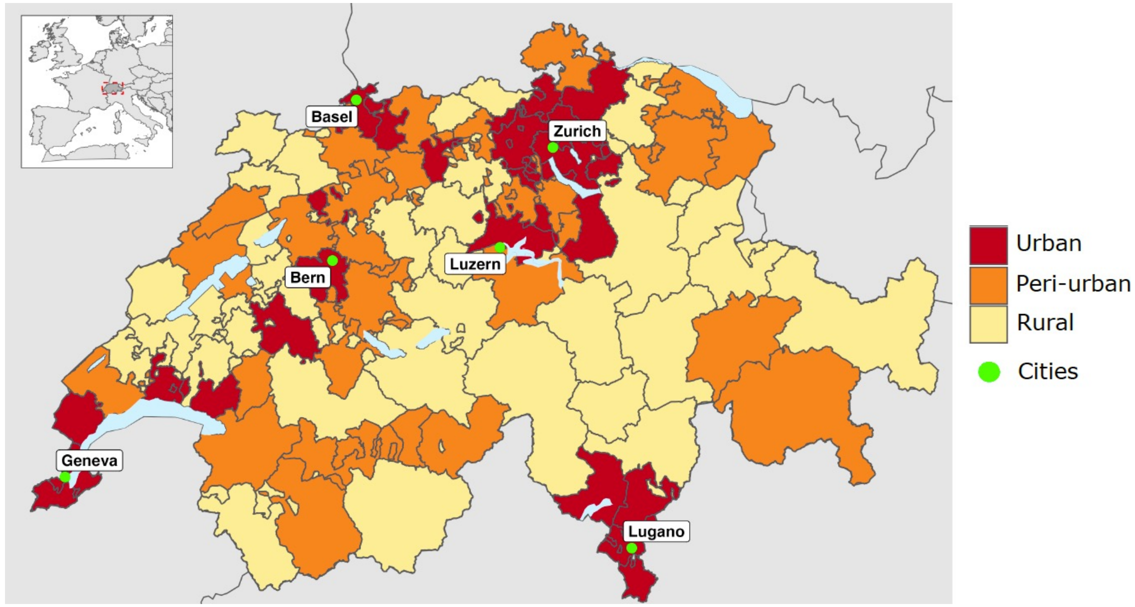
Extensions:
- Multi-location time-series analysis
- Coupled with health impact assessment – estimation of attributable burden.
- DLNMs beyond time-series applications (e.g., cohort analysis).

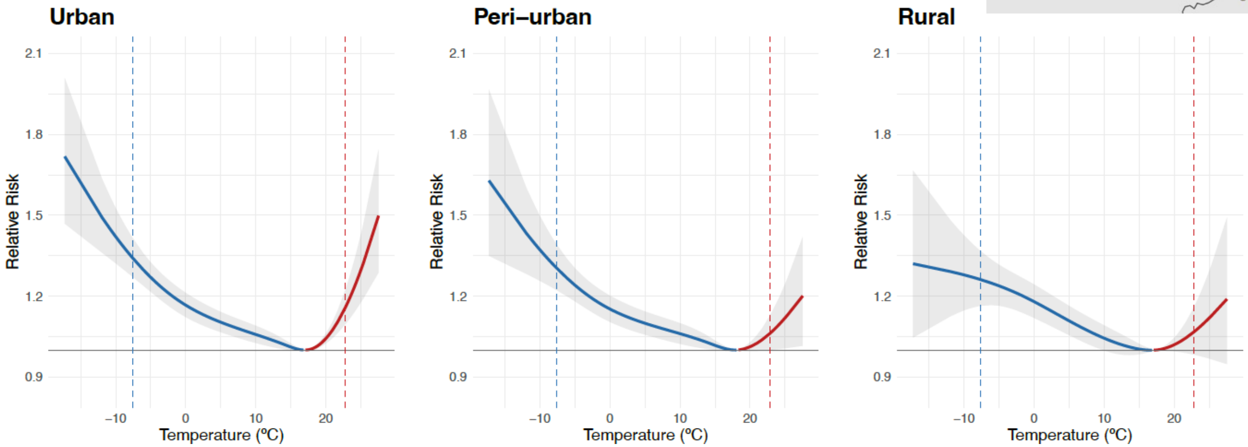
2. Assessment of short-term associations

Environmental exposure – health outcome

Multi-location time-series

- Pool exposure-response curves through multi-variate meta-regression.
- Assessment of potential effect modifiers of the association.





De Schrijver et al. (under review)

2. Assessment of short-term associations

Environmental exposure – health outcome

Projections of health impacts under climate change scenarios

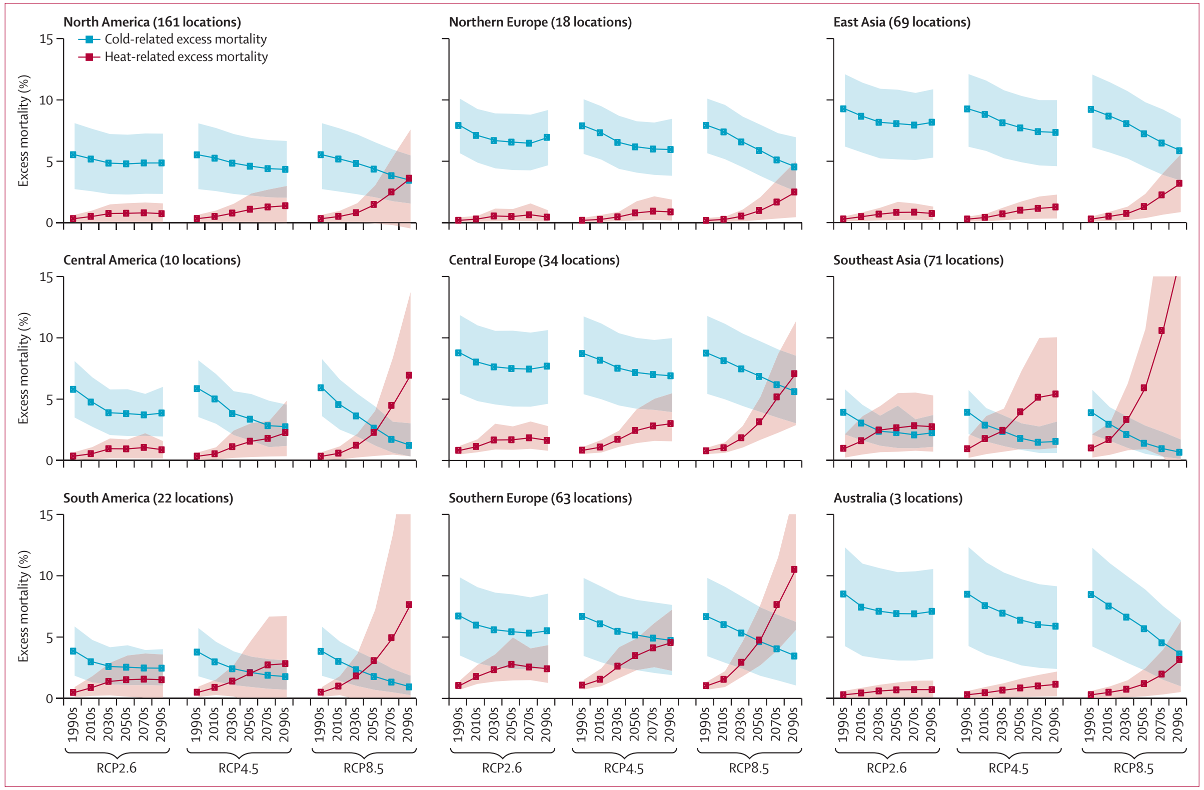


Figure 2: **Trends in heat-related and cold-related excess mortality by region**
The graph shows the excess mortality by decade attributed to heat and cold in nine regions and under three climate change scenarios (RCP2.6, RCP4.5, and RCP8.5). Estimates are reported as GCM-ensemble average decadal fractions. The shaded areas represent 95% empirical CIs. RCP=representative concentration pathway. GCM=general circulation model.

Gasparrini et al. 2017 Lancet PH

2. Assessment of short-term associations

Environmental exposure – health outcome

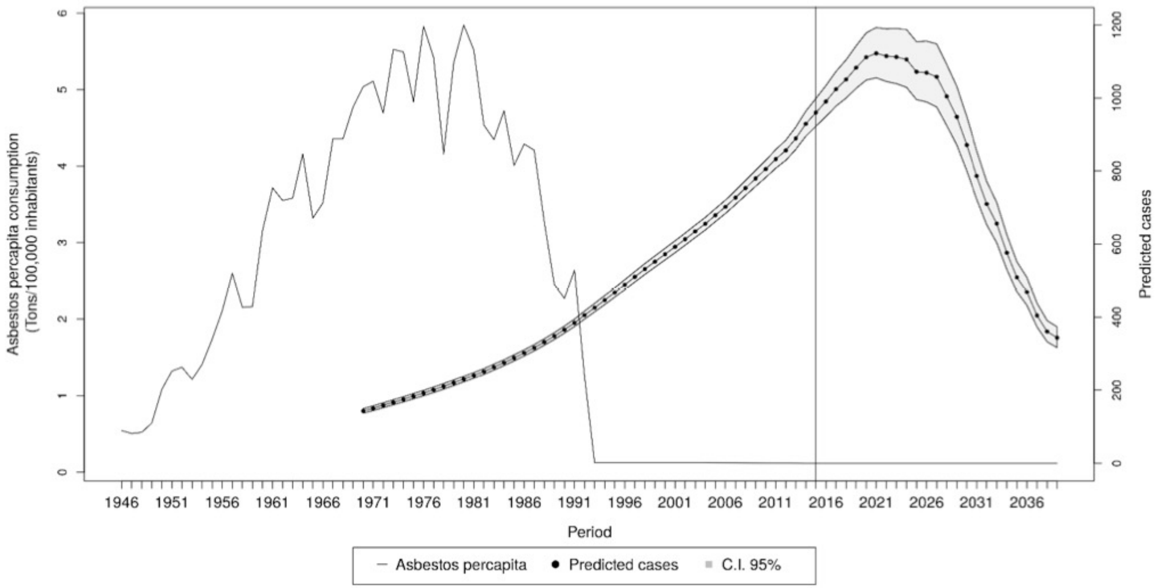


Figure 2. Fitted and predicted (after 2014) MPM cases with related 95% CI. To the left, asbestos per capita consumption in the period 1946–1992 in Italy.

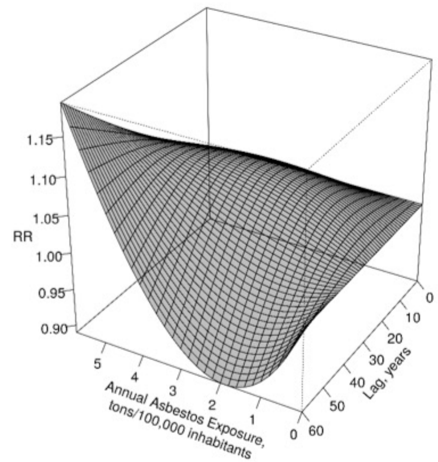


Figure 3. Exposure–lag–response association between the number of MPM deaths and the related past asbestos exposure (tons/100,000 inhabitants). Tridimensional exposure–lag surface.

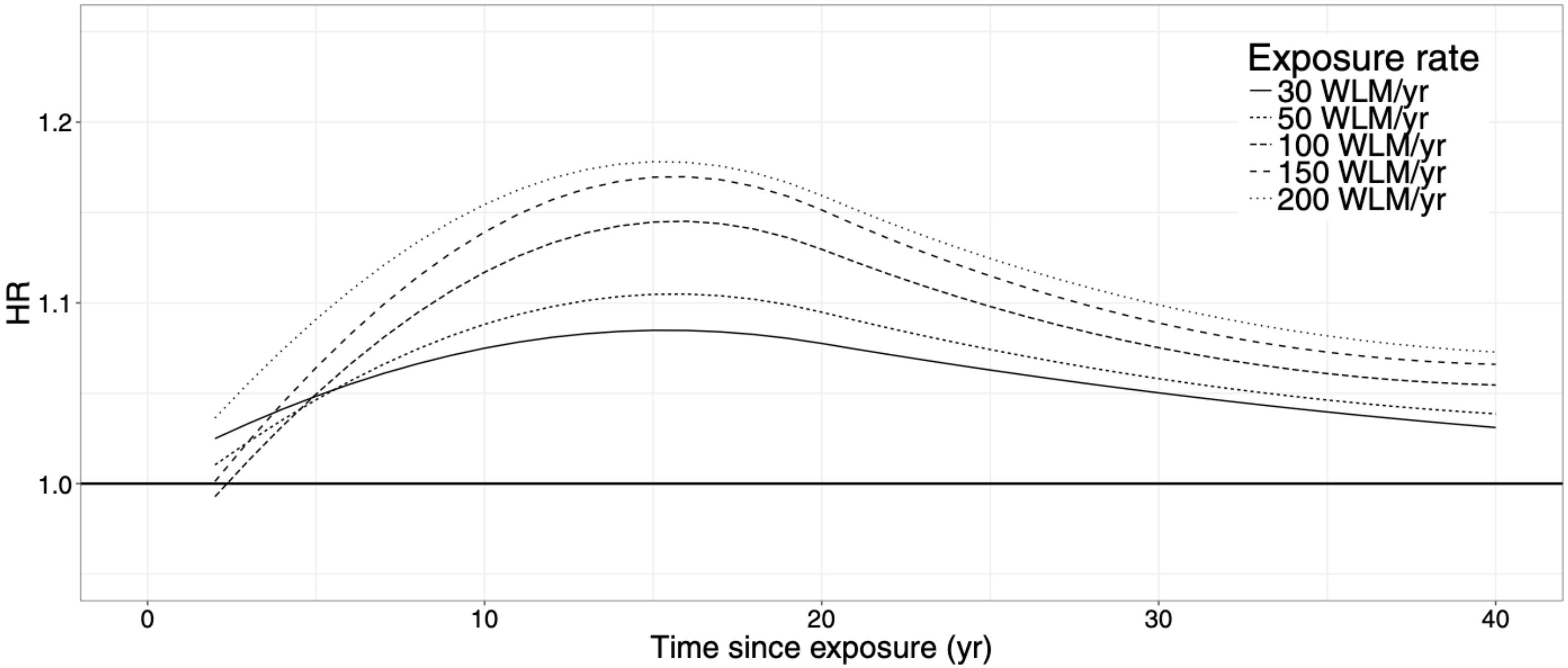
Oddone et al. 2021 Cancers

To predict the future burden of pleural mesothelioma deaths in Italy until 2039 and to describe the exposure–response curve considering time lasted from the exposure start to asbestos

2. Assessment of short-term associations

Environmental exposure – health outcome

Exposure–lag–response associations between lung cancer mortality and radon exposure in German uranium miners

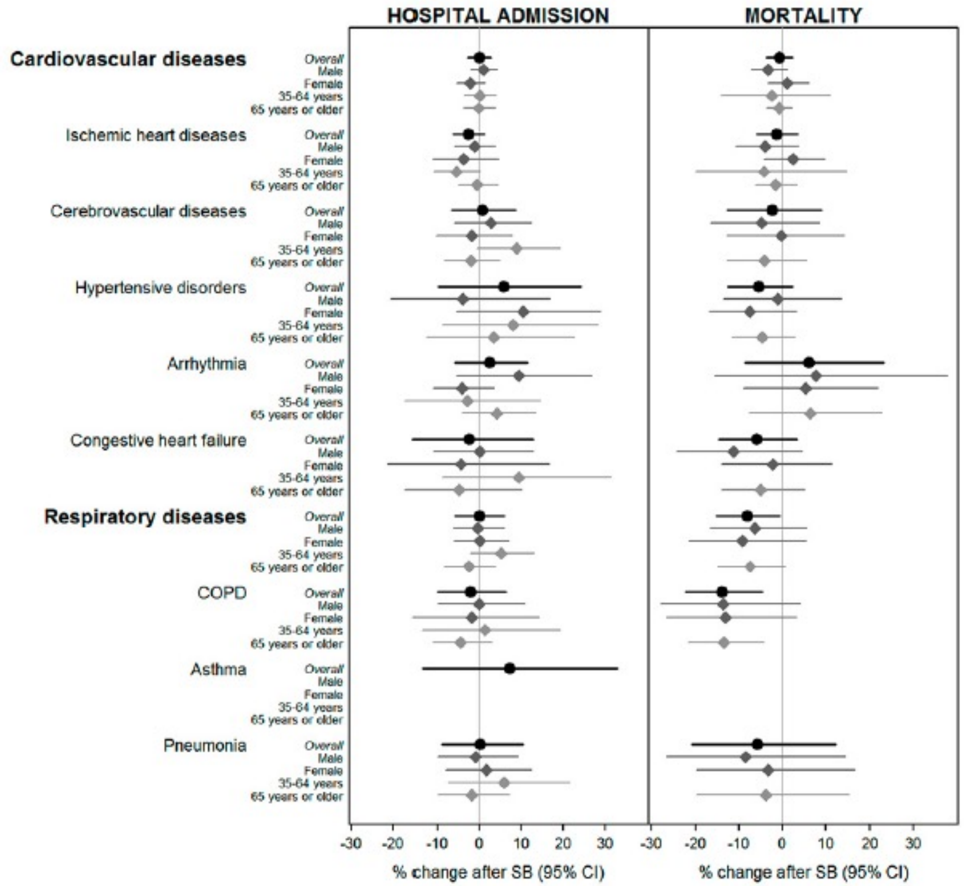


Abenmacher et al. 2019 Radiation and Environmental Biophysics

3. Assessment impact of interventions

Research question: *Did the implementation of smoking bans in Switzerland reduce the rates of cardiovascular and respiratory outcomes?*

- Traditional study designs (cohort or case-control) are hardly applicable, given the lack of a suitable comparison.
- Simple comparison before vs after the ban can be affected by underlying trends in outcome rates
- Ideally, we would need a study design that relies on a pure within-population comparison (no external controls) and able to control for time trends



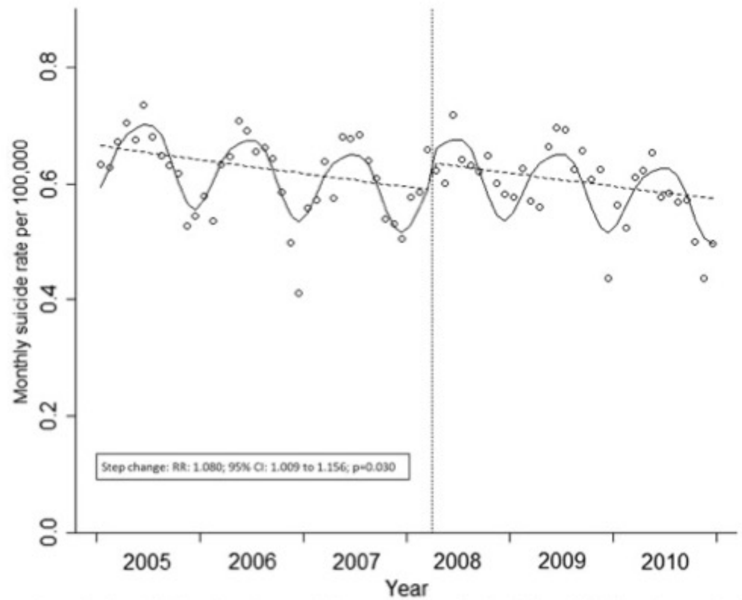
Interrupted time-series analysis

Vicedo-Cabrera et al. 2016 SMW

3. Assessment impact of interventions

Interrupted time-series analysis

- powerful and increasingly popular design for <u>evaluating public health and health service interventions</u>
- a **quasi-experimental design**: it often uses a before-after comparison accounting for underlying trends, in which a population acts as its own control
- <u>Advantages</u>: time-invariant factors are controlled by design
- But… the design is based on a series of strict assumptions that need to be accounted for in the assessment of causal hypotheses.
- Not only to interventions, but also to any kind of events.



Seasonally adjusted. Circles = observed rates; solid line = modelled rates fitted to the data; dashed line = deseasonalised trend. Vertical dotted line = onset of the financial crisis

Figure 1 Trend in monthly suicide rates for all of Spain before and since the financial crisis

Lopez-Bernal et al. 2013

3. Assessment impact of interventions

Interrupted time-series analysis

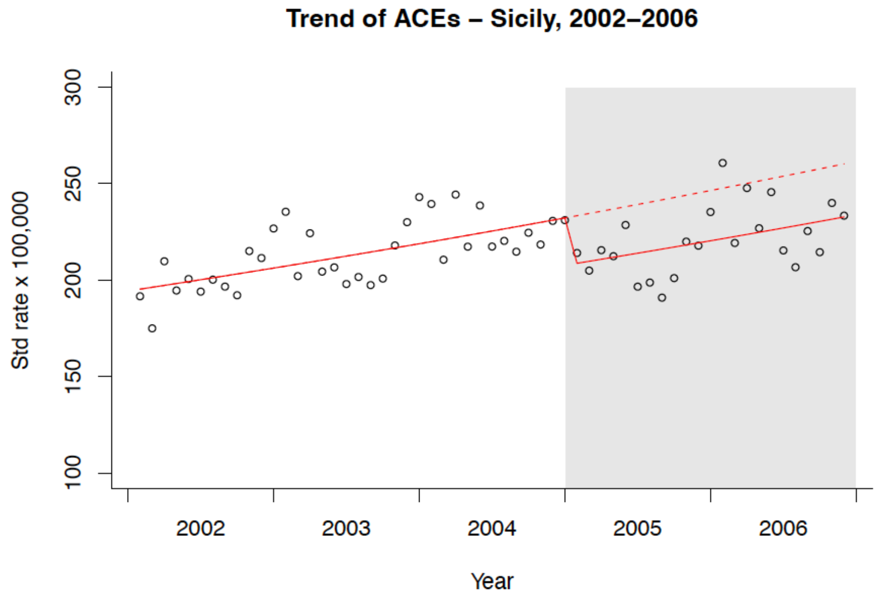
→ analyzing trends in the outcome of interest and estimating the change in trend following an intervention relative to the counterfactual (the expected ongoing trend if the intervention had not occurred).

1) **defining the counterfactual** (by extrapolating the underlying trends observed before the intervention to the postintervention period)

    preintervention period that will be included, any time-varying confounders, whether trends may vary within different subgroups of the population and whether trends are linear or nonlinear.

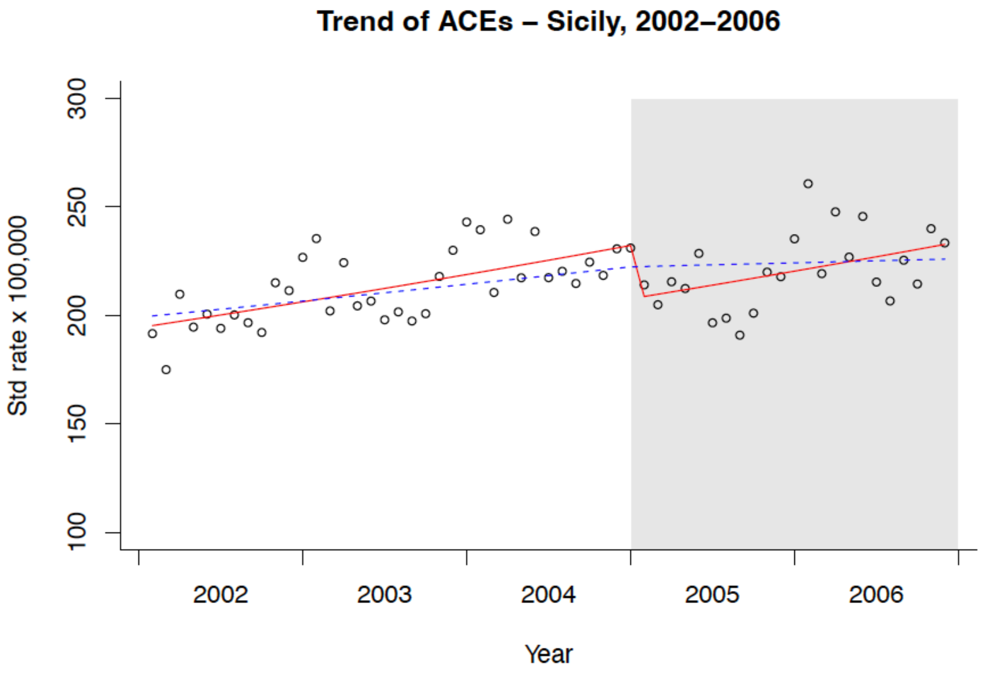
2) **defining the type of effect** that the intervention is expected to have on the outcome (impact model).

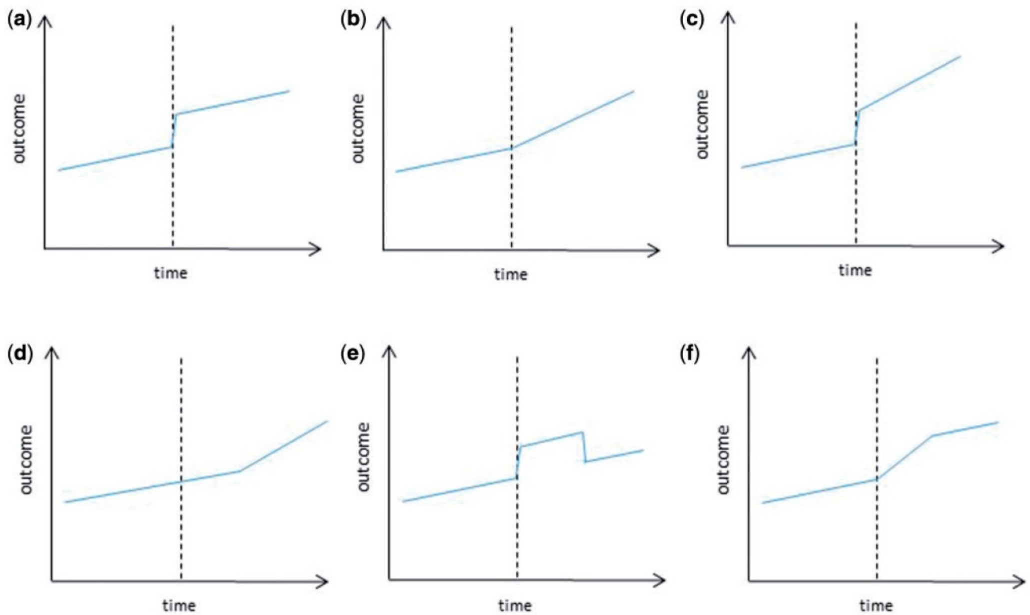
    allow for an abrupt level change, gradual slope change, a lag before any effect on the outcome, transition period during which the intervention is being implemented, and whether a ceiling or floor effect might be expected



3. Assessment impact of interventions

Interrupted time-series analysis





3. Assessment impact of interventions

Interrupted time-series analysis

$$g(y_t) = \alpha + f(x_t, t) + \beta t + s(d_t) + \sum_{p=1}^{P} h_p(z_{pt}) + \epsilon_t$$

Impact model (different parametrizations)

Linear trend

Seasonality

Contributions of other time-varying factors

- If step unction (dummy) log-relative risk of the outcome in the presence vs. absence of the intervention

3. Assessment impact of interventions

Interrupted time-series analysis

CAUTION: Inappropriate model specification can bias the results of an ITS analysis (Lopez-Bernal et al. 2018)

- model not closely tailored to the intervention or testing multiple models increases the risk of false positives being detected.
- use substantive knowledge to customize their ITS model a priori to the intervention and outcome under study.

→ susceptible to bias: (time-varying confounding) since it is unable to disentangle true effects of the intervention from some other simultaneously occurring event or co-intervention → "history bias" threatens the validity of causal inference in ITS.

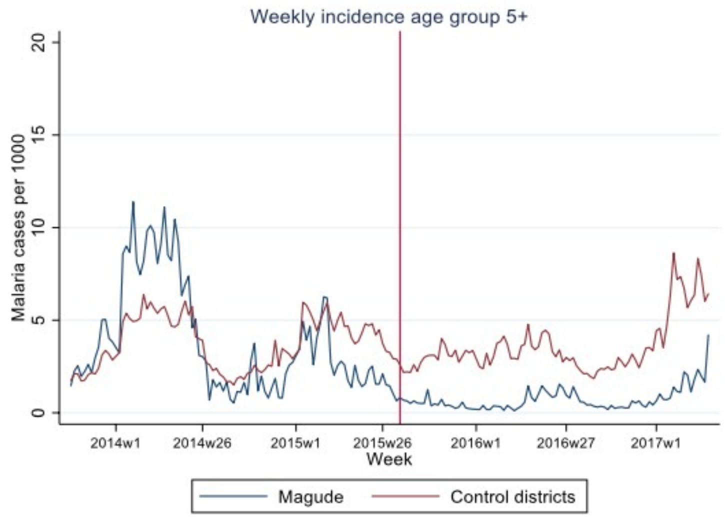
Solution: use of controls (to account for confounding from other co-occurring events by adding an external control series that has been exposed to such events but not to the intervention). (Degli Esposti et al. 2020)
Can be "true" (control population) or synthetic.

Synthetic control: data-driven technique that derives a comparison unit from a weighted average of eligible comparison units (the 'donor pool') that minimizes the differences between pre-intervention trends in the treated and synthetic control series (Abadie et al.

3. Assessment impact of interventions

Interrupted time-series analysis

The short-term impact of a malaria elimination initiative in Southern Mozambique



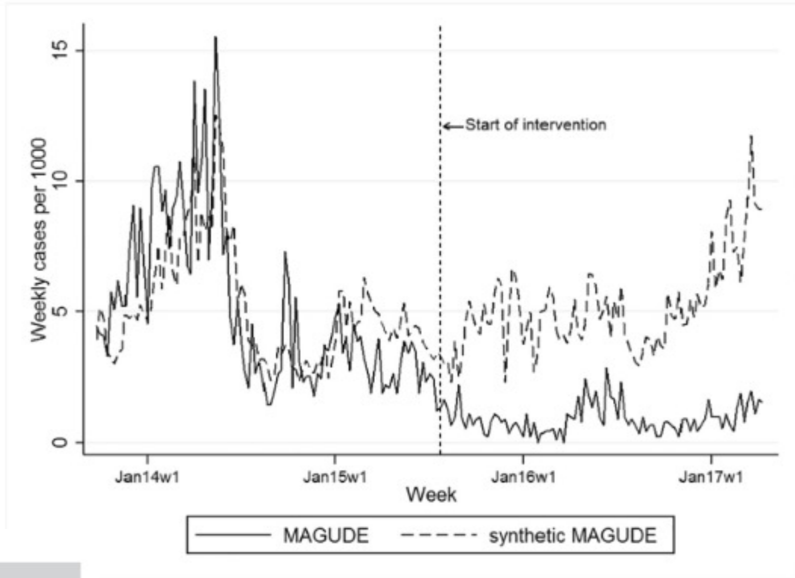


TABLE 2 Average treatment effect and number of cases averted

	Weekly reduction in malaria incidence per 1000		Number of cases averted (% reduction) All ages
	Age 0–4 years model	Age 5+ years model	
Aug 2015–Jul 2016 (intervention year 1)	−3.64	−1.94	6261 (77%)[a]
Aug 2016–Apr 2017 (intervention year 2)	−5.07	−3.21	7061 (80%)[a]
Total (Aug 2015–Apr 2017)			13,322 (78%)[b]
Peak malaria season (Dec 2015–Apr 2016)			3214 (87%)[b]
Peak malaria season (Dec 2016–Apr 2017)			5251 (79%)[b]

Abbreviation: SCM, synthetic control method.

[a]Number of cases averted calculated from the age group specific SCM model. [b]Totals for the entire analysis period and peak malaria season are calculated by summing age group specific SCM model estimates.

Thomas et al. 2021 Health Economics

3. Assessment impact of interventions

Interrupted time-series analysis

Table 2. Effect of the introduction of the NEMS on hospital admission rates in Sierra Leone. The results are presented as immediate and gradual (monthly) change in the rates of hospital admissions after the introduction of the intervention.

Group	Effect of the Intervention	Rate Ratio (RR)	95% CI	*p*-Value
All	Immediate	1.43	1.26 to 1.61	<0.001
	Gradual	1.02	0.99 to 1.05	0.103
Pregnant	Immediate	1.54	1.33 to 1.77	<0.001
	Gradual	1.01	0.98 to 1.04	0.505
Under 5	Immediate	0.90	0.72 to 1.13	0.362
	Gradual	1.05	0.99 to 1.10	0.062
Non-FHCI [1]	Immediate	2.95	2.47 to 3.53	<0.001
	Gradual	1.01	0.97 to 1.05	0.603

[1] Non-FHCI = not included in the Free Health Care Initiative.

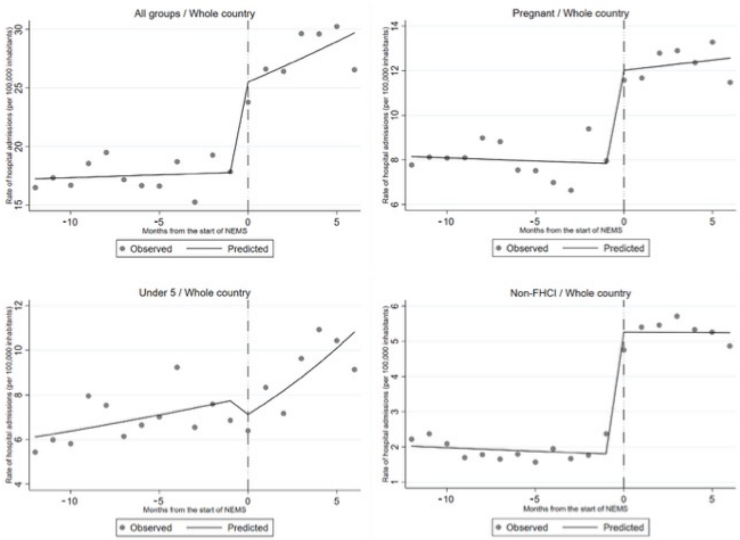


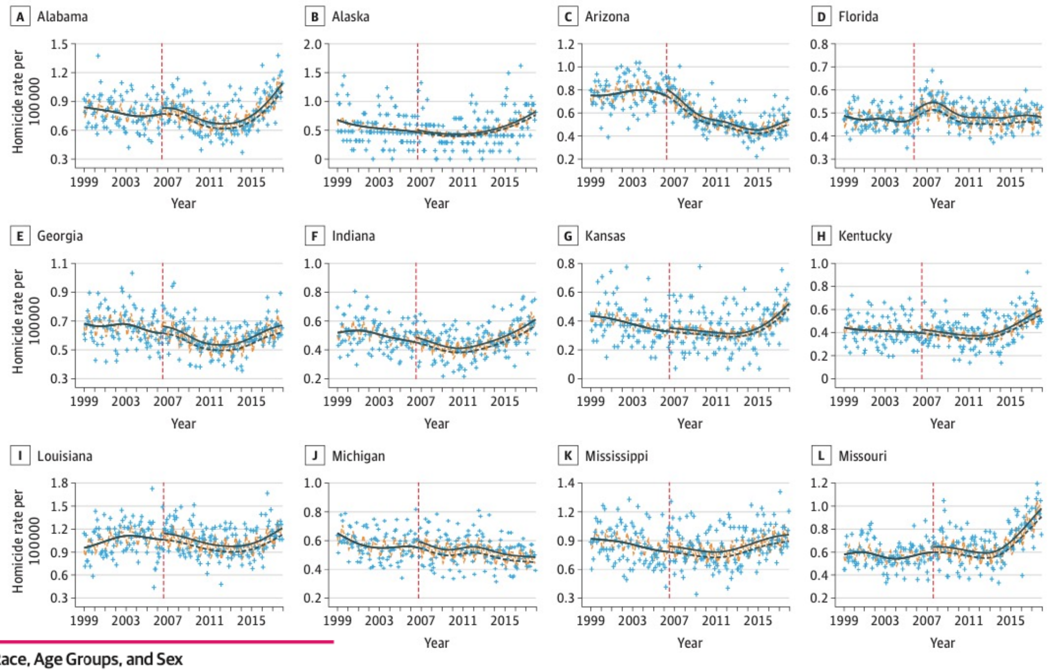
Figure 2. Observed and predicted hospital admission rates in Sierra Leone before and after the introduction of the NEMS. The results are also stratified by categories of the Free Health Care Initiative.

→ increase in access to healthcare services following the implementation of an ambulance-based medical service in a low-income country

Caviglia et al. 2021 Int J Env Res Pub Health

Time series analysis – applications in PH

Figure 1. Associations of "Stand Your Ground" Laws With Changes in Monthly Homicide Rates Across the United States

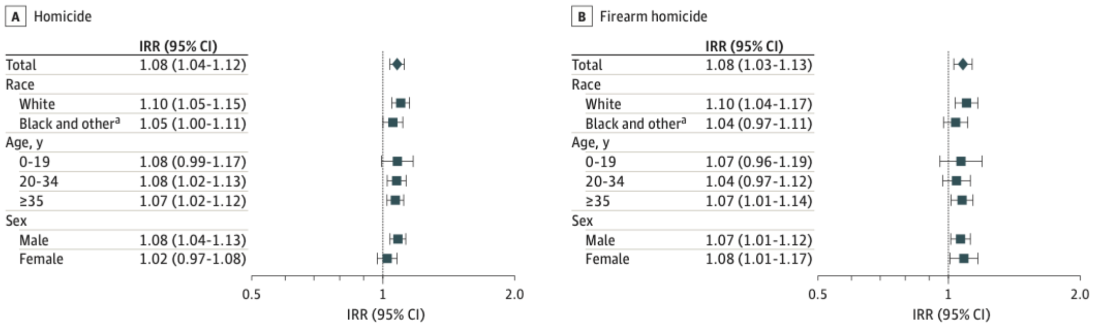


3. Assessment impact of interver

Interrupted time-series analysis

To evaluate the association of "stand your ground" laws with homicide and firearm homicide, nationally and by state, while considering variation by the race, age, and sex of individuals who died by homicide.

Figure 2. Associations of Stand Your Ground Laws With Changes in Monthly Homicide and Firearm Homicide Rates by Race, Age Groups, and Sex



IRR indicates incidence rate ratio, given as rate per 100 000 population.

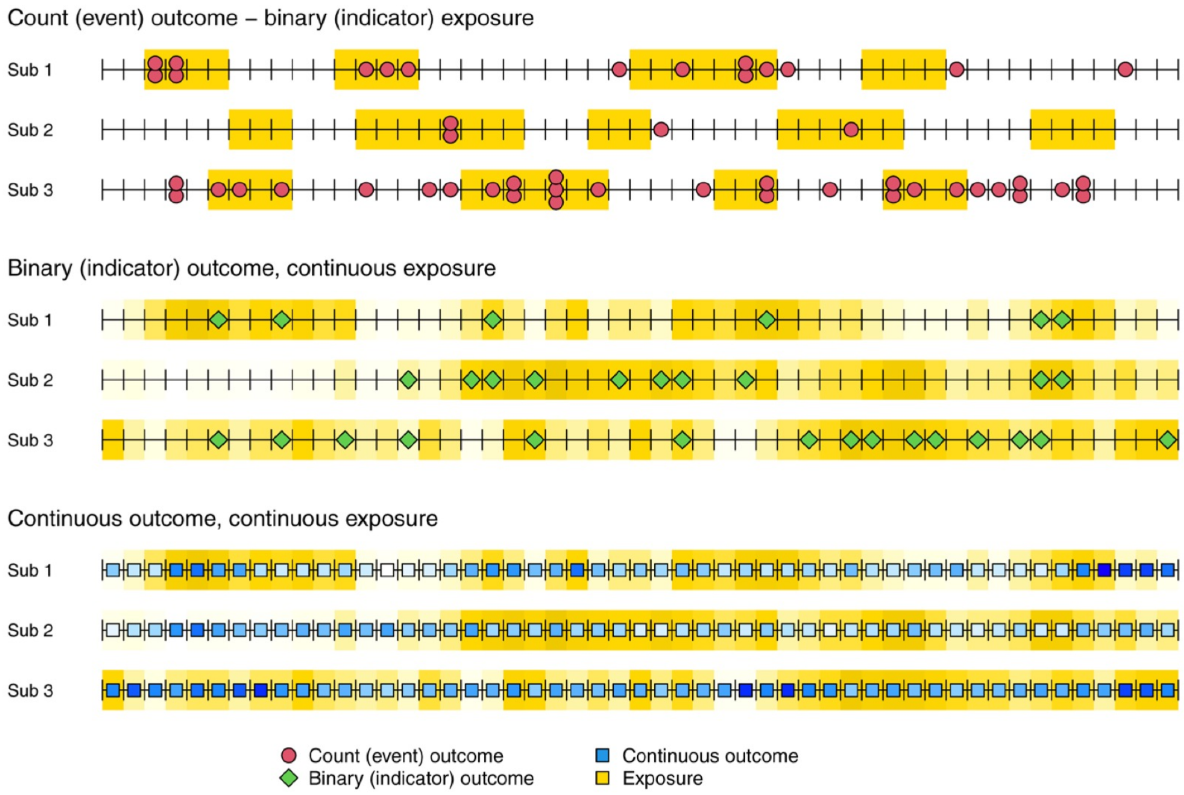
Degli Sposti et al. 2020 JAMA Network

Additional resources

ORIGINAL ARTICLE

The Case Time Series Design

Antonio Gasparrini[a,b]



Webinar https://www.youtube.com/watch?v=y-rhnss7hv4

Gasparrini 2021. The case time series design. Epidemiology

Additional resources

E E P E

European Educational Programme in Epidemiology

Residential Summer Course in Epidemiology

Modern time series methods for public health and epidemiology (5 days course) (4 – 8 July 2022)
Dr. Antonio Gasparrini, London School of Hygiene & Tropical Medicine, London, UK, Dr. Ana Maria Vicedo-Cabrera, University of Bern, Bern, Switzerland, and Dr. Francesco Sera, London School of Hygiene & Tropical Medicine, London, UK

https://eepe.org/courses-2/

Additional resources

General literature on time series analysis in epidemiology

Dominici F, Sheppard L and Clyde M. Health effects of air pollution: a statistical review. International Statistical Review. 2003;71:243-276.

Peng RD and Dominici F. Statistical Methods for Environmental Epidemiology with R - A Case Study in Air Pollution and Health. Springer, 2008.

Wagner AK et al. Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics. 2002;27:299-309.

Wood SN. Generalized Additive Models: an Introduction with R. Chapman & Hall/CRC, 2006.

Ruppert D, Wand MP, and Carroll RJ. Semiparametric Regression. Cambridge University Press, 2003.

Peng RD, Dominici F and Louis TA. Model choice in time series studies of air pollution and mortality. Journal of the Royal Statistical Society: Series A. 2006;169,179-203.

Bhaskaran K, Gasparrini A, Hajat S, et al. Time series regression studies in environmental epidemiology. International Journal of Epidemiology. 2013;42(4):1187-1195.

Distributed lag (non-)linear models

Gasparrini A, et al. Distributed lag non-linear models. Statistics in Medicine. 2010;29:2224-2234.

Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. Statistics in Medicine. 2014;33(5):881-899.

Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011;43(8):1-20.

Gasparrini A. Modelling lagged associations in environmental time series data: a simulation study. Epidemiology 2016;27(6):835-84

Schwartz J. The distributed lag between air pollution and daily deaths. Epidemiology. 2000;11:320-6.

Armstrong B. Models for the relationship between ambient temperature and daily mortality. Epidemiology. 2006;17(6):624-31.

Additional resources

Interrupted TS

Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. International Journal of Epidemiology. 2017;46(1):348-355.

Biglan A, et al. The value of interrupted time-series experiments for community intervention research. Prevention Science. 2000;1(1):31-49.

Wagner AK, et al. Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics. 2002;27:299-309.

Kontopantelis E, et al. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ. 2015;350:h2750.

Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol 2018;47:2082–93.

Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's Tobacco Control Program. J Am Stat Assoc 2010;105:493–505

Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. J Epidemiol Community Health 2018;72: 673–78

Degli Esposti M, Spreckelsen T, Gasparrini A, Wiebe DJ, Bonander C, Yakubovich AR, Humphreys DK.
Can synthetic controls improve causal inference in interrupted time series evaluations of public health interventions?.Int J Epidemiol. 2021 Jan 23;49(6):2010-2020. doi: 10.1093/ije/dyaa152.PMID: 33005920