Methodological considerations for interrupted time series analysis in radiation epidemiology: An overview

Daniel Wollschläger [1]
Anssi Auvinen [2]
Maria Blettner [1]
Hajo Zeeb [3,4]

[1] Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany.
[2] Faculty of Social Sciences (Health Sciences), Tampere University, Tampere, Finland.
[3] Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Deutschland.
[4] Wissenschaftsschwerpunkt Gesundheitswissenschaften, Universität Bremen, Bremen, Deutschland.

Corresponding author:
Daniel Wollschläger
Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI)
University Medical Centre of the Johannes Gutenberg-University Mainz
Langenbeckstraße 1
55131 Mainz
Germany
Phone: +49 6131 17 7029
Fax: +49 6131 17 2968
Email: wollschlaeger@uni-mainz.de
ORCiD: 0000-0002-1009-259X

Keywords:
Epidemiology, Statistical modeling, Ecological study, Confounding

Abstract

Interrupted time series analysis (ITSA) is a method that can be applied to evaluate health outcomes in populations exposed to ionizing radiation following major radiological events. Using aggregated time series data, ITSA evaluates whether the time trend of a health indicator shows a change associated with the radiological event. That is, ITSA checks whether there is a statistically significant discrepancy between the projection of a pre-event trend and the data empirically observed after the event. Conducting ITSA requires one to consider specific methodological issues due to unique threats to internal validity that make ITSA prone to bias. We here discuss the strengths and limitations of ITSA with respect to bias and confounding, data quality, and statistical aspects. We provide recommendations to strengthen the robustness of ITSA studies and reduce their susceptibility to producing spurious results as a consequence of arbitrary modeling decisions.

1 Introduction

In the aftermath of major radiological events like releases from nuclear accidents, it is necessary to assess health consequences for the affected population. Health consequences to be considered include both direct effects from acute radiation exposure or prolonged environmental contamination as well as indirect effects of societal disruption. Epidemiological studies conducted after radiological incidents have improved our knowledge on radiation-induced diseases as well as the understanding of dose-response relationships. Some of these studies have had direct impact on the current system of radiation protection [1,2]. However, conducting epidemiological studies according to gold standard methods with individual-level information on exposure, confounders, and endpoints takes years until results become available.

Sometimes, such studies may not be feasible at all when data collection would involve a prohibitive cost or is otherwise impractical.

In contrast, interrupted time series analysis (ITSA) is an observational quasi-experimental method that can be based on aggregated data on health indicators from publicly available sources. ITSA thus make studies more feasible to carry out and provide faster results. ITSA has a long tradition in evaluating public policy interventions [3,4], health research, and other fields in epidemiology [5–7]. Similar *event study* designs were developed in econometrics [8,9]. ITSA uses a time period prior to an event as a reference with which subsequent development is contrasted. The pre-event trend of the chosen endpoint is modeled and projected onto the post-event period. The projected post-event trend serves as a *counterfactual* baseline comparator in the sense of the Neyman-Rubin model for causal inference [10]. The next step is to determine whether the observed post-event data deviates significantly from this expected baseline. Observed post-event trends can also be contrasted with those in unaffected control regions.

In radiation epidemiology, ITSA has been repeatedly applied to assess the health consequences of major radiological events [11–25]. Examples include ecological studies on perinatal mortality following the nuclear accidents in Three Mile Island [11] or Fukushima [20,21]. Another purpose of analyses similar to ITSA has been to check the plausibility of earlier hypotheses on the magnitude of radiation risk by analogy to historical time series data. Examples include childhood leukemia risk near nuclear installations such as Sellafield and Dounreay in the UK, or nuclear power plants in Germany [26–28].

However, ITSA requires topic-specific methodological considerations, and also involves limitations beyond those of ecological studies in general [29,30]. Previous accounts of ITSA strengths and limitations have mostly been published in relation to specific analyses [31–37]. Frequently, these discussions thus had a restricted scope. Therefore, we here present methodological considerations and provide recommendations for epidemiological ITSA studies in the context of major radiological events.

2 Methodological considerations

2.1 Bias and confounding

Ecological studies are based on aggregated data alone and have inherent methodological challenges that severely impede causal inference, including bias due to the ecological fallacy, and limited or absent confounder control [29,30]. Within-group misclassification of exposure can be especially pronounced absent individual dosimetry or knowledge about other sources of exposure, such as for medical purposes. While unaccounted confounding may create spurious associations or hide real ones even in individual-level epidemiological studies, ITSA in the context of major radiological events faces unique additional threats to its internal validity.

Radiological events may involve changes in other risk factors for the chosen health outcome concurrently with radiation exposure [38,39]. Most obviously following the atomic bombings in Hiroshima and Nagasaki, but also after the nuclear accidents in Chernobyl and Fukushima, radiation exposure was accompanied by major physical destruction and ensuing disruptive changes to society. Catastrophic events with an evacuation can traumatize people and create psychological stress as well as exhaustion. Long-term consequences include deteriorated health service conditions after the

complete or partial collapse of medical infrastructure, psychological stress from health concerns and stigmatization, degraded social networks, homelessness, unemployment, and material losses [38,40,41]. These stressors are all linked to adverse health effects.

In Fukushima, profound lifestyle changes were observed after the nuclear disaster [39]. These included changes in dietary habits and sedentary behavior which can affect the risk of diabetes, cardiovascular health, and other health outcomes in the population. Other possible lifestyle changes concern alcohol and tobacco use which both are relevant risk factors for a number of diseases. Furthermore, contaminated areas may experience differential migration patterns as younger age groups may show mobility away from affected areas. Such changes to the population under study can introduce bias analogously to differential drop out in study arms of longitudinal controlled trials. Due to broad effects of maternal age, reproductive outcomes often analyzed in radiation-related ITSA studies [11,16–24] may be particularly sensitive to changes in population structure. Furthermore, intention to conceive is affected by societal disruption, and pregnant mothers are especially prone to psychological distress from health worries [40]. In addition, the maternal health status can vary with healthcare quality [42] and in turn affects children's health endpoints.

ITSA requires selecting a study time period. Since the projected pre-event trend serves as the counterfactual [10] expectation of the post-event baseline, selection of the pre-event time window influences the post-event expectation and may thus introduce bias if not entirely appropriate. This is illustrated in Fig. 1 where the projection of the appropriate fitted model correctly indicates a positive level shift (c) whereas the projection of a more complicated model with slightly better fit to pre-event data incorrectly indicates a negative post-event level shift (d). The projection from an

appropriate model fitted to data from a restricted pre-event time period incorrectly indicates no post-event level shift (e). As official population statistics cover long periods of time, and therefore provide many possible choices, it is especially important to have stringent principles for choosing the pre-event time period for analysis.

One method to enhance the internal validity of ITSA studies is to compare the observed time trend of the region affected by the radiological event with that of control regions. This requires classifying regions according to the assumed exposure of their populations. When reliable information on exposure levels is unavailable, a classification may only be possible based on spatial proximity to the source of exposure as a surrogate measure. The use of an exposure proxy introduces additional uncertainty besides within-group heterogeneity.

In order to avoid distorted comparisons, the selected control regions must satisfy the following requirement: temporal trends of the analyzed health indicators would have been the same in the control regions as in the exposed region if, counterfactually, the radiological event had not occurred [5,6]. Unless explicitly accounted for in the model, all other time-varying risk factors for the health outcome are thus presumed to affect study regions in the same way, independently of the radiological event. However, this assumption may not hold in the context of a large-scale radiological disaster that involves radiation exposure, psychosocial consequences, and societal disruptions.

2.2 Data quality

ITSA can require much less resources than individual-level studies because it usually relies on routinely collected data from official statistics or disease registries. Among others, endpoints can include mortality, cancer incidence, perinatal mortality, and sex

ratio at birth. In particular, demographic data are usually collected consistently, with high completeness, and covering long time periods as well as many different geographical regions. Nevertheless, in the case of large-scale catastrophes, ITSA investigators should demonstrate that reporting quality and coverage of the included data did not differ across the compared regions and were not affected by the event.

Access to official aggregated data is typically feasible, such that it is possible to define comparison periods and regions without being limited by data availability. Moreover, routine demographic data collection can typically be assumed to be unbiased even when some parties with suspected or actual vested interest in particular study results exist. The availability of official statistics makes samples accessible that are larger than what may be achieved in individual-level studies, thus improving statistical power and precision.

However, disease-specific endpoints like incidence of cancer or congenital malformations depend on the intensity and on the quality of the diagnostic process. Diagnostic procedures, disease definitions, and coding standards may be highly variable across regions and time periods, impeding valid comparisons. Examples include establishing high-quality population-based cancer registration or the implementation of breast cancer screening programs. In some cases, the radiological event triggers changes in data collection with respect to diagnostic methods, record keeping, or implementation of screening programs, thus increasing the risk of information bias. For example, medical surveillance and special health check-ups for exposed persons increased after the Chernobyl accident [43]. Another example is thyroid cancer screening after the Fukushima nuclear accident. While thyroid cancer screening in general increases diagnoses of papillary carcinoma [44], the program introduced in Fukushima also used

technologically advanced ultrasound devices that were more sensitive than devices already in standard use [39]. Consequently, epidemiological comparisons of data generated with varying methods may provide limited insight into the actual study question. A sensitivity analysis excluding the time period after introduction of a cancer screening program is an example how investigators checked their study's robustness against such an effect [25].

2.3 Statistical modeling

ITSA assumes piecewise smooth trends of the chosen health indicator that are separated by an abrupt structural break at a specific time point related to the radiological event. Different types of trend change can be modeled, among them a discontinuous jump in the average level, or a continuous change in the trajectory or slope over time [3]. There are specific statistical requirements for ITSA studies. Time series data typically exhibit serial correlations which can bias the estimated variance of parameters unless accounted for by using heteroscedasticity- and autocorrelation-consistent estimators of the variance-covariance matrix [5]. In addition, non-stationarity and periodicity must be considered by including appropriate secular trends and seasonality components [5].

ITSA model building involves several choices that can strongly impact the results. These include the functional form assumed for the secular pre-event trend of the endpoint as a result of long-term changes in natural risk factors or population demographics. When data is noisy or the time series is not long enough, methods to select the most appropriate model for the pre-event period may not have sufficient power. The investigator then cannot reliably identify the proper baseline trend among, e.g., linear, log-linear, linear-quadratic, or polynomial, with or without seasonality. Consequently, these baseline trends can result in substantially different post-event projections.

Therefore, the same pre-event data can often be used to equally justify very different post-event projections. Some of these projections may be statistically consistent with the observed post-event data while others are significantly different from it. This can result in spurious findings, especially when the pre-event time window or the employed model do not capture the actual long-term secular trend, but instead only capture a short-term seasonal component (Fig. 1 e, f).

By including many ad hoc statistical parameters to capture patterns of the observed time series, the modeled pre-event trend can closely approximate the observed data. Examples include seasonal trends of different periodicity or higher-order polynomial terms. However, this improvement in nominal model fit may come at the expense of worse generalizability to new data when the model essentially chases noise that is incidental to the observed data. To assess the amount of such *overfitting*, special cross-validation or bootstrap techniques for time series data can serve to estimate the out-of-sample prediction error when the model is applied to independent data.

Depending on the health endpoint, the assumed effect of radiation exposure may be assumed to emerge early or with a delay of months to years. The assumed effect then may be considered as permanent or temporary with convergence back to the projected baseline. When the onset of the assumed effect after the radiological event has an uncertain delay, different choices for the structural breakpoint of the time trend can be justified. For example, the vulnerable phase for in-utero radiation exposure during the gestation period may not be precisely defined for a given endpoint, leading to uncertainties about the calendar date or time period when health effects in live births would be expected. Similar situations may occur when the latency of the assumed radiobiological process is uncertain.

For short-term radiation effects, the behavior of the post-event time trend of the health outcome should be related to trends of post-event exposure levels which may change due to weather-related fallout patterns, decay of radionuclides, migration of radionuclides in soil, or remediation and decontamination efforts.

Whenever there are no compelling a-priori principles to guide necessary modeling choices, decisions on the model implementation may be data-driven or made arbitrarily. However, even small changes to a model that are inconsequential individually can strongly affect the overall conclusion when taken together [45]. An even worse situation arises when modeling choices are made ad hoc depending on the available data. Since all data-driven modeling decisions represent implicit statistical tests, they create issues of multiple testing, and thus increase the risk of false positive results beyond the nominal significance level [46]. Moreover, data-driven modeling creates the potential for *p-hacking*, i.e., selectively making modeling choices that lower the final p-value [47]. This problem may be exacerbated by publication bias. Since there are few financial, organizational, or regulatory barriers to conducting ITSA based on publicly available data, the number of studies that can be carried out quickly is large. As a consequence, there may be statistically significant results that are due to chance, but nevertheless have better publication odds than null results. Studies that replicate the exact methods of an earlier study in a different geographical context may help counter this problem. For example, a study on infant leukemia after the Chernobyl accident conducted in Germany [13] used the methods from an earlier study from Greece [15].

One method to evaluate the risk of detecting spurious associations are sensitivity analyses which use slightly different, yet plausible choices for unconstrained model building decisions. Additional analyses can use sham outcomes or zero controls. These

are features included in the analyses with no a-priori evidence to suggest an association with radiation exposure, and therefore with a null effect as the expected result, as illustrated in a study on cancer incidence after the Chernobyl accident [25].

In diagrams depicting non-parametric smoothed trend estimates, the integration bandwidth or sliding window size affect the smoothness of the curve, and thus its susceptibility to local sudden changes like breaks or jumps. Visual cues such as vertical lines indicating the exposure event can be suggestive, making it difficult to assess whether placing other break-points independent of the exposure event would be equally compatible with the data (Fig. 1 a). Smooth trend-lines added to diagrams of raw data can be suggestive and impede visual evaluation of how consistent the natural variability of the data would be with other kinds of trends or temporal patterns. We therefore recommend parsimonious and thoughtful use of such graphical elements. When estimated trend-lines are shown, uncertainty bands are necessary to provide an impression of the range of consistent trend-lines.

3 Discussion

Even though ITSA has inherent methodological limitations, the approach to use time periods and geographical regions as surrogates for individual exposure may also have advantages. Official statistics and publicly available data enhance the feasibility of studies that may serve as a plausibility check for more elaborate epidemiological research. ITSA may – in principle – generate hypotheses about effects of radiation exposure that remain undetected in conventional studies due to low sample size. Since the current understanding of low-dose radiation effects on different endpoints is still

insufficient, study designs that can potentially contribute to the evidence base should be carefully considered.

However, the potential advantages of ITSA must be weighed against the risk for producing spurious associations caused by bias, confounding, and capitalization on chance. Due to inherent limitations in the ecological approach [29,30], the presence of many potential confounders, and a large number of arbitrary modeling decisions, the risk of producing spurious associations is particularly high in ITSA.

While methods to control for time-dependent confounders are severely limited in ecological studies, it is nevertheless important to carefully evaluate the risk of confounding. Therefore, a discussion of established causal factors acting on the endpoint is necessary, including time trends of these factors during the study period. This facilitates identifying alternative explanations for any observed association between exposure-level surrogates and the endpoint. Some confounders like socioeconomic status (SES) have surrogates for aggregated data such as neighborhood SES whose inclusion in an ecological study may improve its robustness [25].

ITSA results should be checked for plausibility in light of previous epidemiological evidence, in particular, when some data on environmental or population radiation exposure levels are available. Authors should then assess whether the effect size suggested by the identified temporal or regional patterns in health indicators are compatible with existing dose-response models or with results from landmark studies [1]. When ITSA results are at odds with these studies, it should be noted what changes to exposure estimates, dosimetry methods, or to dose-response models would be required to accommodate the ITSA results. Consideration should also be given to the possibility

that discrepancies that seem to call for re-evaluation of scientific knowledge underlying current radiation protection guidance may instead reflect shortcomings of the ITSA.

4 Conclusion

Anticipating future exposure incidents that will trigger new time series modeling studies investigating health effects, it is necessary to pre-emptively limit the number of arbitrary choices that can lead to selection bias and capitalization on chance. Following best practices across scientific disciplines, we therefore recommend basing ITSA on a study protocol and statistical analysis plan. This helps to pre-constrain the study implementation and statistical modeling as much as possible. Analysis principles that can be specified ahead of time can include criteria for classifying exposure events as *major* or *relevant*, criteria for choosing control regions and time periods, as well as criteria for selecting statistical models, the functional form of time trends, and sensitivity analyses. Aside from following relevant reporting guidelines [48,49] publications should transparently describe choices made in model selection. They should also include scientifically motivated sensitivity analyses to examine the impact of such choices on study findings.

5 Acknowledgements

The authors declare that there is no conflict of interest.

6 References

1. Ozasa K, Grant EJ, Kodama K. Japanese legacy cohorts: The Life Span Study atomic bomb survivor cohort and survivors' offspring. J Epidemiol. 2018;28:162–9.

2. International Commission on Radiological Protection. The 2007 recommendations of the International Commission on Radiological Protection. ICRP publication 103. Ann ICRP. 2007;37:1–332.

3. Shadish R, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin; 2001.

4. McDowall D, McCleary R, Bartos BJ. Interrupted time series analysis. Oxford, UK: Oxford University Press; 2019.

5. Jandoc R, Burden AM, Mamdani M, Lévesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: Systematic review and recommendations. J Clin Epidemiol. 2015;68:950–6.

6. Cochrane Effective Practice and Organisation of Care Review Group. Suggested risk of bias criteria for EPOC reviews: Risk of bias for interrupted time series (ITS) studies [Internet]. 2017 [cited 2021 Apr 19]. Available from: https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/suggested_risk_of_bias_criteria_for_epoc_reviews.pdf

7. Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JPT, et al. Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews. J Clin Epidemiol. 2017;91:56–69.

8. MacKinlay AC. Event studies in economics and finance. Journal of Economic Literature. 1997;35:13–39.

9. Kothari SP, Warner JB. Econometrics of event studies. In: Eckbo BE, editor. Handbook of empirical corporate finance. San Diego: Elsevier; 2007. p. 3–6.

10. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiol. 1999;10:37-48.

11. Goldhaber MK, Staub SL, Tokuhata GK. Spontaneous abortions after the Three Mile Island nuclear accident: A life table analysis. Am J Public Health. 1983;73:752–9.

12. Ivanov EP, Tolochko GV, Shuvaeva LP, Ivanov VE, Iaroshevich RF, Becker S, et al. Infant leukemia in Belarus after the Chernobyl accident. Radiat Environ Biophys. 1998;37:53–5.

13. Steiner M, Burkart W, Grosche B, Kaletsch U, Michaelis J. Trends in infant leukaemia in West Germany in relation to in utero exposure due to Chernobyl accident. Radiat Environ Biophys. 1998;37:87–93.

14. Török S, Borgulya G, Lobmayer P, Jakab Z, Schuler D, Fekete G. Childhood leukaemia incidence in Hungary, 1973-2002. Interpolation model for analysing the possible effects of the Chernobyl accident. Eur J Epidemiol. 2005;20:899–906.

15. Petridou E, Trichopoulos D, Dessypris N, Flytzani V, Haidas S, Kalmanti M, et al. Infant leukaemia after in utero exposure to radiation from Chernobyl. Nature. 1996;382:352–3.

16. Scherb H, Voigt K. Trends in the human sex odds at birth in Europe and the Chernobyl nuclear power plant accident. Reprod Toxicol. 2007;23:593–9.

17. Scherb H, Voigt K. The human sex odds at birth after the atmospheric atomic bomb tests, after Chernobyl, and in the vicinity of nuclear facilities. Environ Sci Pollut Res Int. 2011;18:697–707.

18. Scherb H, Mori K, Hayashi K. Comment on 'Perinatal mortality after the Fukushima accident'. J Radiol Protect. 2019;39:647.

19. Scherb H, Hayashi K. Spatiotemporal association of low birth weight with Cs-137 deposition at the prefecture level in Japan after the Fukushima nuclear power plant accidents: An analytical-ecologic epidemiological study. Environ Health. 2020;19:82.

20. Körblein A, Küchenhoff H. Perinatal mortality after the Fukushima accident. J Radiol Protect. 2017;37:800.

21. Körblein A, Küchenhoff H. Perinatal mortality after the Fukushima accident: A spatiotemporal analysis. J Radiol Protect. 2019;39:1021–30.

22. Körblein A. Short term increase in low birth weight babies after Fukushima. Environ Health. 2020;19:120.

23. Körblein A, Küchenhoff H. Reduction in live births in Japan nine months after the Fukushima nuclear accident: An observational study. PLOS ONE. 2021;

24. Peterka M, Peterková R, Likovský Z. Chernobyl: Relationship between the number of missing newborn boys and the level of radiation in the Czech regions. Environ Health Perspect. 2007;115:1801–6.

25. Auvinen A, Seppä K, Pasanen K, Kurttio P, Patama T, Pukkala E, et al. Chernobyl fallout and cancer incidence in Finland 1988–2007. Int J Cancer. 2014;134:2253-63.

26. Darby SC, Doll R. Fallout, radiation doses near Dounreay, and childhood leukaemia. BMJ. 1987;294:603–7.

27. Wakeford R, Darby SC, Murphy MFG. Temporal trends in childhood leukaemia incidence following exposure to radioactive fallout from atmospheric nuclear weapons testing. Radiat Environ Biophys. 2010;49:213–27.

28. Wakeford R. The risk of leukaemia in young children from exposure to tritium and carbon-14 in the discharges of German nuclear power stations and in the fallout from atmospheric nuclear weapons testing. Radiat Environ Biophys. 2014;53:365–79.

29. Morgenstern H. Ecologic studies in epidemiology: Concepts, principles, and methods. Ann Rev Public Health. 1995;16:61–81.

30. Greenland S, Robins J. Invited commentary: Ecologic studies – biases, misconceptions, and counter-examples. Am J Epidemiol. 1994;139:747–60.

31. Rossi HH. A response to 'Perinatal mortality in Germany following the Chernobyl accident'. Radiat Environ Biophys. 1997;36:137.

32. Körblein A, Küchenhoff H. Response to the letter to the editor by H. H. Rossi. Radiat Environ Biophys. 1998;36:301–3.

33. Soleman S, Fujitani T, Harada K. Letter to the editor regarding 'Spatiotemporal association of low birth weight with Cs-137 deposition at the prefecture level in Japan after the Fukushima nuclear power plant accidents'. Environ Health. 2020;19:121.

34. Scherb H, Hayashi K. Response to the 'Letter to the editor' by Sani Rachman Soleman et al., 'Spatiotemporal association of low birth weight with Cs-137 deposition at the

prefecture level in Japan after the Fukushima nuclear power plant accidents'. Environ Health. 2020;19:123.

35. Bochud F, Jung T. Comment on 'The human sex odds at birth after the atmospheric atomic bomb tests, after Chernobyl, and in the vicinity of nuclear facilities', Hagen Scherb & Kristina Voigt environ, Sci Pollut Res (2011) 18:697-707. Environ Sci Pollut Res Int. 2012;19:2456–9.

36. Scherb H, Voigt K. Response to F. Bochud and T. Jung: Comment on 'The human sex odds at birth after the atmospheric atomic bomb tests, after Chernobyl, and in the vicinity of nuclear facilities', Hagen Scherb & Kristina Voigt, Environ Sci Pollut Res (2011) 18:697–707. Environ Sci Pollut Res Int. 2012;23:4234–41.

37. Blettner M. European stillbirth proportion and Chernobyl. Int J Epidemiol. 2000;29:596–9.

38. Abend M, Nisbet A, Gering F, Averin V, Andersson K, Schneider T, et al. 'Living in contaminated areas' – consideration of different perspectives. Health Phys. 2020;119:2–11.

39. United Nations Scientific Committee on the Effects of Atomic Radiation. UNSCEAR 2020 Report – Scientific Annex B: Levels and effects of radiation exposure due to the accident at the Fukushima Daiichi Nuclear Power Station: implications of information published since the UNSCEAR 2013 Report. 2020.

40. Houts PS, Tokuhata GK, Bratz J, Bartholomew MJ, Sheffer KW. Effect of pregnancy during TMI crisis on mothers' mental health and their child's development. Am J Public Health. 1991;81:384–6.

41. Rahu K, Auvinen A, Hakulinen T, Tekkel M, Inskip PD, Bromet EJ, et al. Chernobyl cleanup workers from Estonia: Follow-up for cancer incidence and mortality. J Radiol Protect. 2013;33:395–411.

42. Parkhurst JO, Danischevski K, Balabanova D. International maternal health indicators and middle-income countries: Russia. BMJ. 2005;331:510–3.

43. United Nations Scientific Committee on the Effects of Atomic Radiation. UNSCEAR 2008 Report – Scientific Annex D: Health effects due to radiation from the Chernobyl accident. 2008.

44. United Nations Scientific Committee on the Effects of Atomic Radiation. Evaluation of data on thyroid cancer in regions affected by the Chernobyl accident. A white paper to guide the scientific committee's future programme of work. 2017.

45. Blettner M, Sauerbrei W. Influence of model-building strategies on the results of a case-control study. Stat Med. 1993;12:1325–38.

46. Gelman A, Loken E. The statistical crisis in science: Data dependent analysis – a 'garden of forking paths' explains why many statistically significant results don't hold up. Am Sci. 2014;102:460–5.

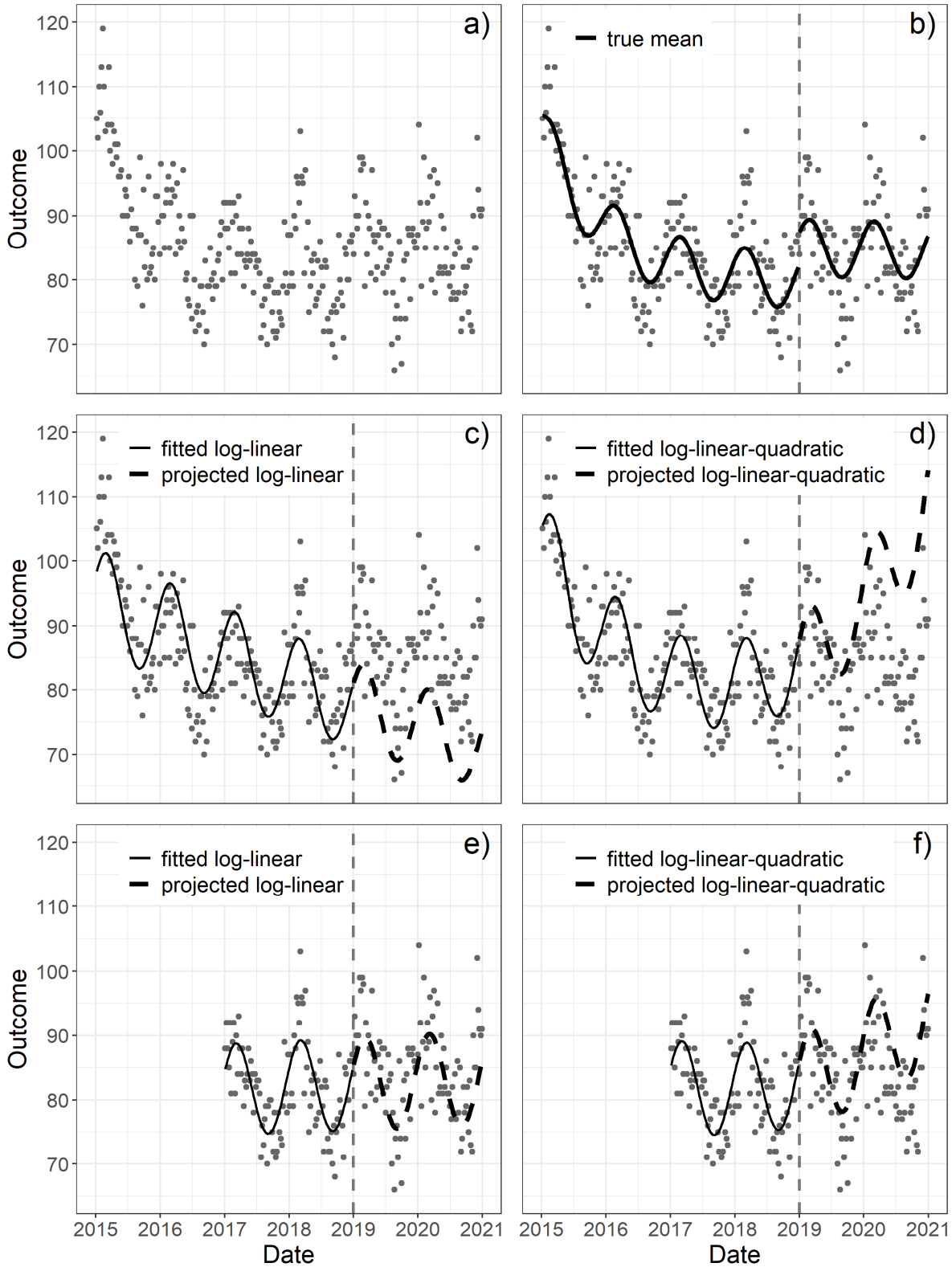
47. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLOS Biology. 2015;13:e1002106.

48. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344–9.

49. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLOS Medicine. 2015;12:e1001885.

7 Figures

Figure 1: Simulated noisy time series data of an arbitrary outcome that follows a decreasing log-linear secular trend with a periodic cycle and an abrupt level shift at 2019-01-01. a) Simulated data. b) Data with true mean (grey) and indicator for the time of the incident. c) Data with fitted (solid black) pre-event mean and projected (dashed black) post-event mean from log-linear model. d) Data with fitted (solid black) pre-event mean and projected (dashed black) post-event mean from log-linear-quadratic model. e) Data with fitted (solid black) pre-event mean and projected (dashed black) post-event mean from log-linear model restricted to a subset of pre-event data. f) Data with fitted (solid black) pre-event mean and projected (dashed black) post-event mean from log-linear-quadratic model restricted to a subset of pre-event data.


a)
b)
true mean
c)
fitted log-linear
projected log-linear
d)
fitted log-linear-quadratic
projected log-linear-quadratic
e)
fitted log-linear
projected log-linear
f)
fitted log-linear-quadratic
projected log-linear-quadratic
Outcome
Date
120
110
100
90
80
70
2015
2016
2017
2018
2019
2020
2021