Are Older Nuclear Reactors Less Safe? Evidence from Incident Reports in the French Fleet*

Romain Bizet, Petyo Bonev, François Lévêque

October 5, 2017

Abstract

This paper assesses variations of nuclear safety over time, and across reactors of different ages and technologies in the French nuclear fleet between 1997 and 2015. We use a novel dataset describing over 19,000 nuclear safety events reported by France's unique nuclear operator. We face two endogeneity issues when assessing the effect of age on safety. First, reports of safety events are prone to measurement errors as plant managers can fail to detect or report safety events. We deal with this issue by restricting the analysis to a subset of perfectly detected and declared events. Second, we face an Age-Period-Cohort issue as the age, the year of observation and the year of commissioning of a reactor are perfectly collinear. We address this problem by including year and power station fixed-effects in our specifications, and argue that under a mild assumption, this is sufficient for the age effect not to be biased. Our results regarding the evolution of reactor safety are consistent with the existence of a *bathtub* trend in a majority of power stations: the ageing of reactors has a positive, linear effect on safety, as well as a negative quadratic one.

Keywords: nuclear power, safety, age, technology, reactor design, incident data.

1 Introduction

In this paper, we empirically evaluate how nuclear safety varies over time, and across nuclear reactors of different ages, using data on significant safety incidents reported between 1997 and 2015 in the 58 French nuclear reactors. From a policy-oriented perspective, this question comes at a timely moment, as the oldest French power stations are reaching their fortieth anniversary - their initial maximum lifespan - and are now seeking for licence extensions, which led EDF (France's single nuclear utility), safety regulators and policy-makers to argue over the effect of age on the safety of these plants. Despite this high political and social importance, this question has not been answered in a satisfactory way by the existing literature. The available data on nuclear safety is often of limited amount or quality, since major safety accidents are rare. The existing literature has therefore relied on datasets composed of accidents from both nuclear

*We thank the audience of the Young Energy Economist and Engineers seminar in Edinburgh University, as well as four anonymous reviewers for their remarks and comments. All remaining errors are entirely ours.

reactors and fuel cycle facilities (Hirschberg et al., 2004; Sovacool, 2008; Hofert and Wüthrich, 2011; Wheatley et al., 2017). Due to the scarcity of these accidents, these datasets contained at most 216 observations, which limits the possibilities of inferring statistically significant results regarding the variations of safety with the age of nuclear reactors.

The statistical estimation of variations in nuclear safety across reactors and over time faces two methodological challenge. First, when the risks of catastrophic accidents are measured by safety indicators based on information reported by regulated firms,[1] safety may be prone to measurement errors. In this paper, we use safety indicators based on incidents declared by nuclear operators to the safety regulator. If there is a positive probability that these events remain undetected by operators, or if plant managers have incentives not to report some events[2], or if the criteria defining which events ought to be reported vary over time, then a spurious effect in the data could bias the estimates, due to the unobservability of both detection abilities and compliance with declaration criteria.[3] An instance of this spurious effect was described by Rose (1990) within the airline industry, in which safety incidents are also used as a measure of safety. In the following of this paper, we refer to changes in regulatory declaration criteria and measurement errors due to detection failures or non compliant behaviours as *transparency* issues.

The second identification challenge in the estimation of the relation between age and nuclear safety is the classical Age-Period-Cohort issue (APC in the following, see Glenn (2005) or Yang and Land (2013) for reviews of the related literature). The existence of a linear relation between the period of observation, the age of a reactor during this period, and its commissioning cohort makes it logically impossible to disentangle the respective effect of these three factors on without making further assumptions.[4] Partial solutions to this identification issue are still largely debated (see e.g. Bell and Jones (2016); Luo et al. (2016) or Land et al. (2016)), and necessarily require structural assumptions on one effect to identify the other two (O'Brien, 2011; Bell and Jones,

[1]As an example, the World Association of Nuclear Operators defines annual counts of automatic and manual reactor shut-downs as a key nuclear safety indicator.

[2]The fear of regulatory sanctions or of public backlashes embodies these incentives. An anecdotal example of this situation in the nuclear industry is the occupation of the French Fessenheim power plant in 2016 by Greenpeace after a German newspaper claimed that an incident had been understated by the French nuclear safety authority in 2014.

[3]the ability to detect an event might vary with the design of nuclear reactors, while the propensity of firms to comply with declaration criteria might vary over time with the evolution of the stringency of safety regulation.

[4]An illustrative way of framing the ACP problem is provided by Bell and Jones (2013) for the case of suicide frequencies. An age effect is for instance the observation that older people commit suicide more frequently. A cohort effect is for instance the observation that people born during wars commit suicide more often than others, regardless of their age. Finally, a period effect is for instance the observation that people commit suicide more often during economic downturns, regardless of their age or year of birth. Yet, as $Age = Period - Cohort$, it is impossible to measure all three effects in a single experiment without additional constraints on one of the effects.

2014; O'Brien, 2017). We relax this collinearity issue by noting that the multiple reactors within each power station were commissioned during different years, and discuss the hypotheses under which the effect of age on safety can be identified.

This paper contributes to the empirical literature on the relation between age, technology and nuclear safety in several ways. First, we introduce a novel dataset which contains all nuclear safety incidents reported between 1973 and 2015 in French nuclear reactors by France's single utility (EDF). This dataset gathers detailed information on more than 19,000 so-called *significant safety incidents*. These events represent the most significant departures from the general standards of operation of a nuclear reactor.[5] These events are analysed on a case-by-case basis by both plant managers from EDF and by experts from the safety regulator, to identify organizational or technical weaknesses in nuclear stations and foster the exchange of best practices across the nuclear fleet. To the best of our knowledge, no statistical analysis of these events has ever been conducted. We present some descriptive statistics of this dataset which show that different types of events appear to be characterized by various time-trends, while the annual counts of reports of events per reactor seem to be unevenly distributed across reactors of different ages and technological designs.

Second, we use this dataset to estimate how nuclear safety varies over time, and across reactors of different ages. In order to avoid the aforementioned possible measurement errors, we identify a subset of events that exhibit perfect detection and declaration rates (PDD events in the following). This subset contains two types of events: automatic shut-downs and events that require the unplanned use of a reactor's safeguard mechanisms. The former events are perfectly detectable and declared due to their impact on the reactor's electrical output. The latter are considered by the safety authority as particularly severe and easy to detect by both managers and inspectors. The period of observation of these events spans from 1997 to 2015. By considering events characterized by perfect detection and declaration, all variations in the number of declared events must be due to variations in reactor safety. This novel approach for the analysis of nuclear safety is related to the strategy used in Hausman (2014), who uses automatic shut-downs to evaluate the effect of economic incentives on safety in the U.S. nuclear market.

Third, we analyse whether transparency as defined above affects the declaration of significant

[5] The French Nuclear Safety Authority (ASN) defines ten declaration criteria that characterize which events ought to be reported. These events are not deemed significant by the regulator on the basis of their real consequences of damage but rather on the basis of the information they carry regarding the management of plant safety, and on the basis of their causes and potential consequences.

safety events. To do so, we compare the results obtained on the set of perfectly detected and declared events with the results of a similar regression run on the whole set of nuclear incidents, arguing that if transparency has no effect on the declaration process, the results obtained on both sets ought to be similar. This approach is related to the approach in Rose (1990), who analysed the effect of a deregulation of the U.S. commercial airline market on airline safety.[6]

Our results first indicate a general increase in safety over the life of nuclear reactors, potentially due to the major refurbishments carried out during the periodic decade inspections of the reactors. In addition, occurrences of safety events evolve with the age of reactors consistently with the usual *bathtub* trend from the reliability literature (see e.g. Aarset (1987); Mudholkar and Srivastava (1993); Xie et al. (2002) or Chen et al. (2011)). In other words, we measure a significant positive effect of reactor age on safety as well as a significant negative quadratic effect of age on safety. This suggest that initial progresses are made - for instance through learning-by-doing - but that the ageing of reactors can be detrimental to their safety, for instance through the wear-out of materials. We also show that measuring aggregated age effects at the level of large groups of reactors can neglect substantial plant-level heterogeneity, and that failing to disentangle age effects from cohort-specific effects can lead to large biases in the estimates. In particular, the average effect of age on safety can be largely over-estimated if one fails to account for plant-level heterogeneity.

The paper is organized as follows. Section 2 describes the French declaration process and conducts a descriptive analysis of our dataset. Section 3 presents the identification strategy and empirical specifications. Section 4 exposes our results and section 5 concludes.

2 Significant safety events in the French fleet

2.1 Institutional set-up and data

The French nuclear fleet is constituted of 58 pressurized water reactors (PWR), located in 19 nuclear power stations (referred to as sites in the following), and owned by a single utility (EDF). These reactors were built in separate phases from the late 1970s to the late 1990s. The technological design of reactors evolved during construction. For instance, reactors differ in their nominal capacity, the nature of their fuel, or in their ability to perform load-following. French

[6] Feinstein (1989) also carries out a study of reporting behaviours in U.S. power stations based on violation data. The detection-controlled estimator proposed in this study cannot be replicated here, as we do not have data on detected non-compliance. Indeed, Feinstein's model is based on a mixture model that requires "inspector-specific" data that would specifically affect detection of non-compliance situation to determine his coefficient β_2.

Table 1: The French fleet, by conception and nominal capacity levels.

Capacity	Conception	Power stations	Reactors	Construction
900 MW	CP0	2	6	1971-1979
	CP1	4	18	1974-1985
	CP2	3	10	1976-1988
1300 MW	P4	3	8	1977-1986
	P'4	5	12	1980-1992
1450 MW	N4	2	4	1984-2000

Note: The 900 MW (MegaWatt) cohort contains reactors of three different conceptions (CP0, CP1 and CP2 reactors). The 1300 MW cohort contains reactors of two different conceptions (P4 and P'4 reactors). The 1450 MW cohort contains only one conception cohort (N4). Construction phases span from the beginning of the construction of the first reactor and until the connection of the last one.

reactors can be split within three cohorts of reactors according to their electrical production capacity. Each of these cohorts is constituted of one or more sub-cohorts that capture less important design features. These groups of reactors are summarized in table 1. The first column of this table lists the three capacity cohorts of reactors. The second column lists their respective sub-cohorts. The following columns describe the number of nuclear power stations and reactors that belong to each cohort, as well as their construction period. An important remark is that all reactors within a given power station share the same technological design (i.e. they share the same capacity and belong to the same sub-cohort).

The safety of French nuclear power stations is regulated by the Nuclear Safety Authority (ASN in the following), which sets regulatory standards for the operation of nuclear reactors. In particular, regulatory standards include mandatory reporting criteria that define a list of particular situations, or events, that have to be reported to the safety authority, as they are deemed significant for safety. These events are referred to in the following of the paper as *significant safety events* . This terminology is also the one used by both EDF and the French nuclear safety authority. The aim of these reports is to aggregate information and share experience and best practices among reactors, and to detect generic defaults in the reactors' designs.

To comply with these reporting criteria, plant operators gather information, on a daily basis, on a much broader set of situations which depart from normal operation. All these events are analysed and filed on site. The most severe ones, i.e. the subset of events that match the reporting criteria, have to be reported to the safety authority. Upon the detection of an event requiring a report, plant managers have two weeks to provide the regulator with a detailed summary of the event and an analysis of its causes and consequences.

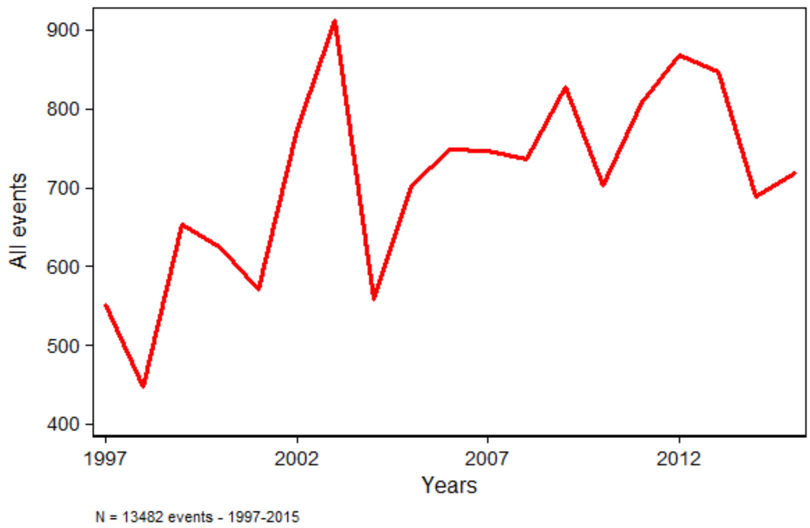
The ASN enforces this reporting procedure through periodic and random inspections, during which inspectors get access to the events filed by operators as too insignificant to be reported.[7] Although they get access to all this information, the quantity of situations assessed by operators but leading to no declaration is too important for inspectors to review them all. Despite this possibility for inspectors not to detect all reporting violations, a detection would allow the regulator to engage in repressive actions against the plant manager, such as lawsuits or temporary shut-downs.

Our dataset, obtained from the ASN, contains all 19,575 significant safety events reported from 1973 to 2015 in the French nuclear fleet.[8] Each event is characterized by a set of variables, which contain information on the location and date of declaration of the event, the nature of the components, materials and systems of the reactor affected by the event, its level on the International Nuclear Event Scale[9] (INES), the reporting criterion associated with the event, the state of the reactor at the time of detection (e.g. production, refuelling or maintenance) and a description of its causes and consequences.

Not all of these events are used for our analysis. First, generic events characterizing the whole fleet or particular cohorts of reactors are excluded.[10] Second, we discarded the events declared during construction, as the safety authority considers the reporting criteria to be tailored for the operation and maintenance of reactors rather than for their construction periods. Third, we exclude the period of observation 1973-1996 due to incompleteness of the description of the events. In 1996, a reform of the reporting criteria led to a more stringent and complete reporting process.

Conversely, when aggregating safety events into counts of reports per reactor and per year, events that affected systems common to multiple reactors within a given power station are counted in each affected reactor.

Figure 1: Annual declarations of four categories of events in the French fleet



2.2 Descriptive evidence

Figure 1 presents the evolution of the total number of reports over time in the French fleet from 1997 to 2015. Approximately 700 events are reported each year, which amounts to approximately one significant safety event per reactor and per month. This figure shows an increasing annual number of declarations over time. This could be due to increases in the stringency of the safety standards, leading the regulator to consider new types of events as significant for safety. Competing rationales would be an increase in the ability of the operator to detect significant safety events, or a deterioration of nuclear safety with the ageing of the fleet, leading to more numerous significant safety events.

Next, we compare the frequencies of safety events across groups of reactors. [11] Figure 2 presents four box plots showing how the annual declarations of events are distributed across four definitions of groups of reactors. In each box plot, multiple comparisons of group means are performed. We present the results of these comparisons by indicating a reference group in green,

[7]These regulatory inspections occur several times a year. In addition to these frequent audits, each reactor is thoroughly reviewed every ten years, and mandatory safety investments are defined by the ASN in order to pursue operation. These investments consist, for instance, in the replacement of materials which are assumed to have been deteriorated in a way that precludes their further safe use.

[8]Events declared during the lifetime of permanently shut-down reactors are not included in our dataset.

[9]The International Nuclear Event Scale is a severity scale for nuclear events, defined by the International Atomic Energy Agency.

[10]Generic events consist in detections of conception failures which are specific to either a group of reactors, or to the whole fleet. According to the regulator, these events capture specific efforts made by EDF to increase its knowledge of the conception of his reactors, as well as their reliability.

[11]We refer to statistics describing counts of events occurring in a given reactor during a given year as reactor.year statistics. For instance, the Gravelines-5-2004 reactor.year is constituted of all the events declared in the fifth reactor of the Gravelines power station in 2004.

and the groups that differ significantly (at the 5%-level of significance) from this reference in red. The statistics on which these comparisons are based account for multiple pairwise comparisons[12].

For the upper two sub-figures in figure 2, the x-axes represent the capacity and conception cohorts, ordered by date of construction. It appears that the reactors that belong to the most recent capacity cohort (1450 MW) declare a significantly larger average number of events when compared to the 900 MW cohort.[13] Some heterogeneity appears within capacity cohorts when split into the six underlying conception cohorts. Within the 900 MW group, reactors of the oldest design ($CP0$) declare a significantly higher number of events than the reactors of the other two designs. The same conclusion can be drawn for the reactors of the $P4$ conception cohort with respect to their younger siblings of the $P'4$ conception cohort. Note that in this second box plot, only the results of pairwise comparisons within each capacity cohorts are represented.

The bottom-left graph on figure 2 is dedicated to France's regional regulatory subdivisions.[14] Regional distributions of reports of significant safety events are rather equally distributed, as only two pairwise comparisons of the mean annual number of reports are significant. It appears that the reactors overseen by the Bordeaux division (e.g. the 8 reactors located in the Blayais, Golfech and Civaux power stations) declare significantly lower numbers of events than the reactors overseen by the Strasbourg division and the Caen division.

On the bottom-right graph, four age groups are defined by the decade of operation of a reactor at the time of observation.[15] Reactors in their third and fourth decade of operation declare significantly larger numbers of events than reactors in their first decade of operation. It is interesting to notice the relative under-dispersion of the fourth-decade group with respect to the three younger groups. This could be due to a relative lack of observation of reactors in their fourth decade. Indeed, as of 2014, only 45 reactor.years in the fourth decade have been observed, whereas respectively 743, 563 and 414 reactor.years have been observed in the first, second and third decade of operation.

Another piece of descriptive evidence that safety may vary with the age of nuclear reactors, we provide in figure 3 multiple linear fits of the reports of automatic shut-downs on the age of the reactors at the time of reporting. Each point on the graph corresponds to a given reactor-

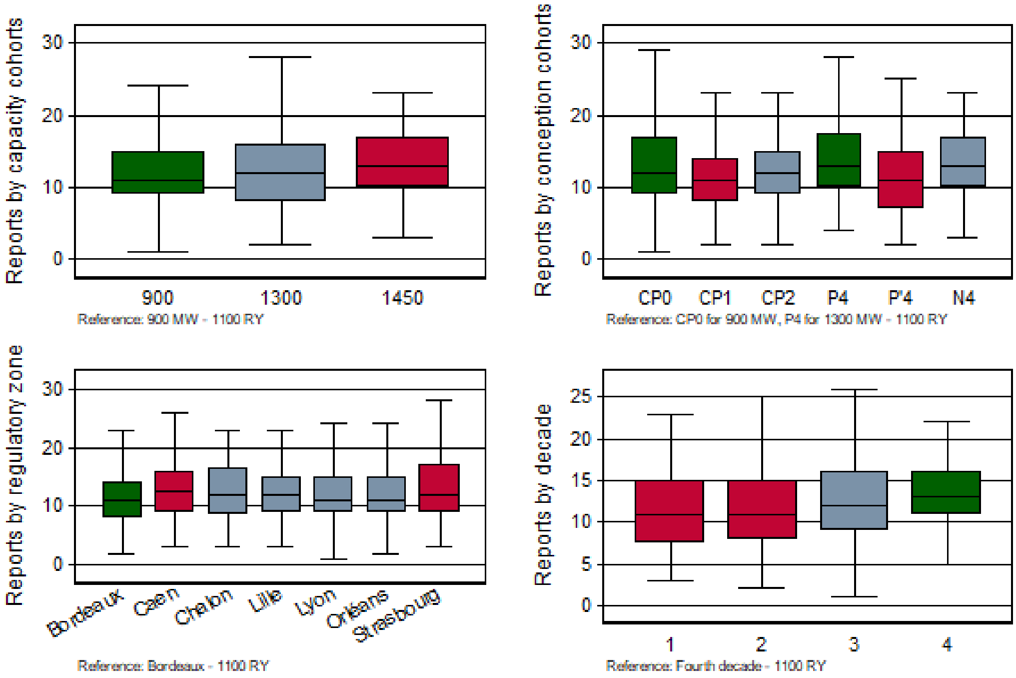
[12] The significance of the difference in means among the pairs identified in figure 2 is robust to several methods of adjustment for multiple comparisons (i.e. Tukey, Sidak and Bonferonni's methods).

[13] According to the safety authority, this is be due to the comparatively more complex design of these recent plants, which led to large quantities of events in their early years of operation.

[14] The ASN delegates the duty of inspection to its 7 territorial subdivisions, who have some level of discretion in their interaction with the plant managers.

[15] For instance, an event observed in a 26 year-old reactor is associated with the third decade of operation of this reactor.

Figure 2: Annual declarations by reactor for different groups of reactors



year, and each regression line is obtained by performing a simple linear regression of the annual counts of reports on the age of the reactor during the year observed. Three linear fits are presented, based on the reactor-years observed during three periods of roughly equal length: 1997-2003, 2004-2009 and 2010-2015. For instance, the solid line corresponds to the reactor-years observed between 1997 and 2003. In all three linear fits included in the figure, it appears that more automatic shut-downs are reported among older plants. In addition, the positions of the regression lines seem to suggest that the average number of automatic shut-downs is decreasing over time.

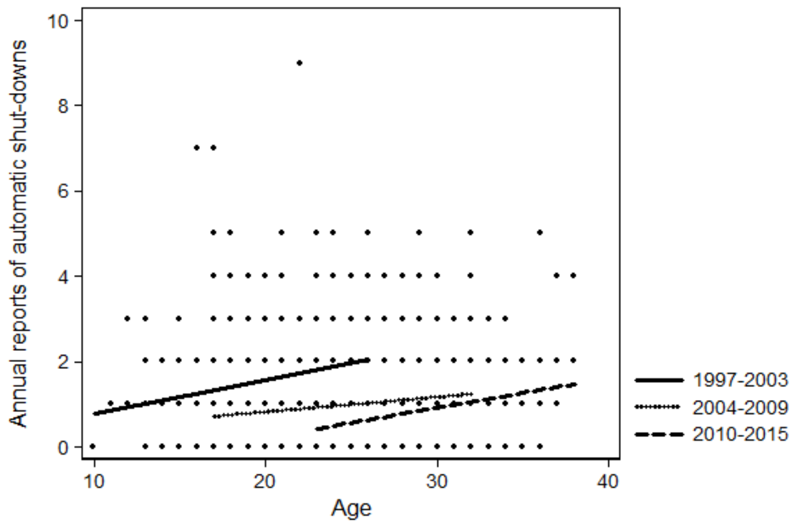
3 Identifying safety variations across reactors

3.1 Identification strategy

3.1.1 Endogeneity of age and technology

Assume Y is a binary random variable such that $Y = 1$ when a safety event occurs. Our objective is to measure the evolutions of *safety*, which we define here as $\mathbb{P}\{Y = 1\}$. The rationale for this definition is twofold. First, it is consistent with the definition of safety as a probability of occurrence of events which may harm people or goods, which has become usual in the economics

Figure 3: Reports of automatic shut-downs as a function of reactor age



literature.[16] Second, even though the event $\{Y = 1\}$ is minor in terms of real consequences, reducing its probability of occurrence has a direct impact on the likelihood of major accidents, as major accidents are often composed of a combination of individually minor events.

Consider the model

$$\mathbb{P}\{Y = 1|W, C\} = \mathbb{E}\{Y|W, C\}, \tag{1}$$

where the random vector W contains observed factors of safety such as age and technology, and the random variable C is unobserved. Examples for factors in C are the unobserved (reactor-specific) ability to detect an event, or the propensity of an operator to report a significant event when detected. g is the regression function in the population. We are interested in the marginal effect of a covariate W_j (e.g. age) on the likelihood that an event occurs: $\frac{\partial \mathbb{E}\{Y|W,C\}}{\partial W_j}$. If, however, C is related to W, then the observed regression function $\mathbb{E}\{Y|W\}$ will capture this effect, which will lead to a bias, $\mathbb{E}\{Y|W\} - \mathbb{E}\{Y|W, C\} \neq 0$.

We are faced with three major endogeneity concerns. First, there might be a secular time trend in the overall nuclear safety. Potential drivers of such a trend are unobserved changes in regulatory standards, learning effects, as well as changes in the safety care exerted by plant managers. These changes might be correlated with both age and technological change. An instance of regulatory change is the evolution of the reporting criteria defining significant safety events. Whereas only three criteria existed before 1996, in later years the number of these criteria was increased to ten, directly influencing the quantity of information reported by managers

[16] See for instance Shavell (1984); Hansson and Skogh (1987); Faure and Skogh (1992) or Laffont (1995).

to the safety authority. If all observations were pooled together and such regulatory changes ignored, different reactors could reveal different observed frequencies of events simply because the measurements were taken at different points in time. Thus, ignoring secular changes in the regulatory framework could lead to a spurious effect of age in the data.

Second, there might be other unobserved factors that are related to the age and technology of a nuclear reactor. An important category of such factors are the so-called cohort effects, which reflect the existence of specific common conditions characterizing the time of construction or commissioning of a set of reactors, such as a power station or a group of reactors sharing a common design.[17] Cohort effects may include internal conditions such as infrastructure or management culture within the particular set of reactors, as well as external conditions, such as geographical aspects or common exposure to regulatory inspections and norms at the time of the construction (see e.g. Glenn (2005) and Suzuki (2012) for a more in-depth definition of cohort effects). The same logic applies more broadly to any time-fixed reactor-specific effects that might be related to age and/or technology. Omitting cohort (or more generally, age-related individual) effects from the list of controls would bias the estimators of the independent variables.

Third and most importantly, the statistical analysis of safety is potentially hampered by measurement errors due to missing observations. There are two channels through which missing observations might occur: plant managers might fail to detect a safety event, or they might not report an observed event to the safety authority.[18] To formalize this, assume O is a binary random variable such that $O = 1$ if a safety event is observed, or detected, by the plant manager. $\mathbb{P}\{O = 1|Y = 1\}$ represents the plant manager's *ability to detect* events. Assume D is a binary random variable such that $D = 1$ if a safety event is reported - or declared - by the plant manager to the regulator. $\mathbb{P}\{D = 1|O = 1\}$ captures the plant manager's *propensity to declare* events. With this notation, the only data observable to the econometrician are the reports, which occur only when the realized state is $\{Y = 1, O = 1, D = 1\}$. Detection of events by plant managers might be related to the technology. Furthermore, incentives not to report events may vary across reactors. Thus, both reasons for missing observations are potentially related to age and technology. Whereas the compliance channel is common to most industries subject to environmental regulation and self-reporting rules(the main example being the CO_2-emitting industries, such as the pulp and paper industry or the coal industry), the detection channel is

[17]We are thankful to an anonymous referee for pointing this out.

[18]Several reasons may lead to failures to report a significant event. If a manager's pay-off entails performance objectives indexed on occurrences of these events, non-declaration may result from voluntary non-compliance. Subjective misconceptions of the reporting criteria are a second rationale for non-reporting of an event.

mainly endemic to the complexity of the nuclear energy industry.

3.1.2 Identification strategy

A standard approach to solve endogeneity problems 1 (secular time trend) and 2 (cohort/individual effects) is to add year dummy variables and individual fixed effects, respectively. Time dummies capture period-specific factors that are common for all reactors, such as unobserved changes in regulatory stringency and in safety efforts. Individual fixed effects, on the other hand, account for time-constant factors that are specific to a particular reactor or to a cohort of reactors. One major problem in our set-up, however, is that we are mainly interested in the effect of age on safety. Including time dummies, individual fixed effects and an age variable leads to perfect multicollinearity. The easiest way to see this is to think about the case in which individual-effects are constituted solely of cohort effects, where a cohort of reactor is a group of reactors commissioned during the same year. In this case, since age = year - cohort, fixing two of the three variables uniquely determines the third one. As a result, there is no variation left in the data to identify the effect of the third, uniquely determined variable.

This is a fundamental identification problem referred to as the Age-Period-Cohort (APC) identification problem by the literature, see Bell and Jones (2013, 2014, 2016) or Keyes et al. (2008, 2010).[19] In our case, using reactor fixed effects instead of a cohort variable leads to an equivalent problem, since the additional degree of freedom is sufficient to trigger a perfect multicollinearity in the data. There are several methods to tackle the APC problem developed in the literature, the most important ones being the constraint-based method by Mason et al. (1973), the Holford approach by Holford (1983, 1991), the median polish approach (Tukey, 1977; Selvin, 2004), and the hierarchical APC model by Yang and Land (2006).[20] Our identification approach explores age variation of reactors *within* site cohorts of reactors. A site (or power station) consists of up to six reactors, which are built in a pre-specified order within the common geographic area defined by the perimeter of the power station.

Our strategy amounts to assuming that all individual-specific time-fixed endogeneity (including cohort effects) can be captured with a site fixed-effect. This assumption is based on the observation that all reactors within a site share a very large number of characteristics. First and foremost, they share a common technological design. In addition, in most of the cases,

[19] Again, we thank to an anonymous referee for pointing out this literature to our knowledge.

[20] Exploring all of these methods is well beyond the scope of the paper. Moreover, as the paper by Bell and Jones (2013) beautifully points out, the validity of the results produced by these methods hinges crucially on the theoretical foundation of their assumptions. Black box application of the methods may lead to arbitrary results with no causal interpretation.

reactors within a site were even built by the same set of subcontractors. Second, due to their geographical proximity, reactors within a site are exposed to the same climatic and seismic (and other geography related) conditions. Furthermore, for the same reasons, these reactors share common infrastructures, such as their cooling source. Third, reactors within a site share operational management and staff. At the same time, reactors within a site have some (even if not very large) variation in age, as they were typically not built and commissioned simultaneously. Thus, we can treat reactors within a site as a cohort and explore the age variation as a source of identification. Based on these considerations, we spell our first identification assumption:

Once controlled for observed site characteristics, time and site fixed effects, all endogenous variation is driven by transparency, i.e. by the ability to detect an event (O) and the manager's propensity to report safety events (D). Formally, we assume that

$$C = (O, D, \varepsilon) \quad \text{with}\, W \perp\!\!\!\perp \varepsilon, \tag{A1}$$

where ϵ is an idiosyncratic error term, and W contains now all observed factors, as well as time dummies and individual fixed effects.

Assumption A1 is closely related to the identification approach of Yang and Land (2006): the endogenous parts (i.e. cohort effects) are treated as common to a whole cohort, whereas the age is an individual variable. This allows to break the perfect multicollinearity between Age, Period and Cohort effects. As with most exogeneity assumptions, assumption A1 cannot be tested directly without a valid instrument. There are two main deviations from A1 that one has to worry about. First, the order of building the reactors within a site might have a long-term impact on the performance of the reactors. This would be the case if learning effects during building the first reactor significantly contributed to the safety of the reactors that were built later. Such an effect would be related to age and therefore violate assumption A1. The violation would lead to a negative bias in the estimate: older reactors would appear less safe solely due to age (instead of due to learning effects). We deal with this possibility by introducing a binary variables indicating whether a reactor was the first one to be built in a site, or the first of its particular design (see below the empirical specifications section). A second possibility of a violation is that there are reactor-specific time-varying factors which are related to age but not captured by the time dummies. [21]

We now turn to the crucial problem of imperfect observability of events (problem 3). First,

[21] Although it is difficult to think of such factors, the simulation of Bell and Jones (2014) reveals the necessity to consider such a possibility.

given the description of the declaration process and its regulatory oversight, we assume that type I errors cannot occur:

$$\mathbb{P}\{O = 1|Y = 0\} = \mathbb{P}\{D = 1|O = 0\} = 0. \tag{A2}$$

Intuitively, assumption (A2) means that the plant manager cannot detect an event which did not occur, nor can he declare an event which has not been previously detected.

Our identification strategy consists in finding a subset of the sample, for which detection and reporting failures are precluded. We define an event to be subject to *perfect detection and declaration* (PDD events in the following) if the following assumption is fulfilled:

$$\mathbb{P}\{O = 1|Y = 1\} = \mathbb{P}\{D = 1|O = 1\} = 1. \tag{A3}$$

According to the definition presented in (A3), an event is said to be perfectly detectable and declared if it is certain that an operator will observe the event conditionally on its occurrence, and declare it conditionally on its observation. We denote the subset of all PDD events as Θ_{PDD}. We obtain the following result:

Lemma 3.1. *Assume that the regulatory standards are fixed and that assumptions* (A1)*,* (A2) *and* (A3) *hold. Further, assume that the idiosyncratic noise* ε *has an expectation equal to* 0*. Then, for the model*

$$Y = g(W, C) = g(W, O, D) + \varepsilon \tag{2}$$

it holds

$$\mathbb{E}\{Y|W, C\} = \mathbb{E}\{Y|W\}. \tag{3}$$

The proof can be found in the appendix. The assumption $\mathbb{E}\{\epsilon\} = 0$ is a trivial assumption and can be achieved through a corresponding normalization of the model. Additivity of ε is implicitly assumed in the model presented in equation (2). This assumption is common for most econometric specifications used in empirical work.

Thus, according to lemma (3.1), the model $g(W, C)$ can be uncovered from the observed data $\mathbb{E}\{Y|W\}$ for the subset of PDD events. On this subset, variations in the frequencies of occurrence of safety events are necessarily caused by variations in safety. We describe the set Θ_{PDD} in the following section. A discussion of external validity is provided in section 4.2.

Table 2: Descriptive statistics: three dependent variables

Variable	Definition	Mean	Std. Dev.
ASD_{it}	Automatic shut-downs declared during year t in reactor i	1.122	1.233
SFG_{it}	Events requiring the use of safeguard mechanisms declared during year t in reactor i	0.377	0.693
ALL_{it}	All events declared during year t in reactor i	12.256	5.105

1100 reactor.years - 1997-2015.

3.2 Perfectly detected and declared events

For any particular type of safety incidents, there are two conditions that guarantee perfect detection and declaration (i.e. A3). First, events that have a direct effect on the electrical output of a power station cannot be undetected nor hidden as the Transportation System Operator[22] monitors the electric production of each power station. Second, events which are subject to particular auditing efforts during inspections led by the safety authority ought to be declared truthfully, as it can be argued that such events are (nearly) impossible to be hidden, which eliminates the incentives of plant managers not to report them.

Two categories of safety events satisfy one of these conditions. First, automatic reactor shut-downs stop the electrical production of nuclear reactors.[23] These events have also been used by Hausman (2014) as a proxy for nuclear safety. Second, events that require the unplanned use of safeguard systems (safeguard events in the following) are subject to specific auditing efforts during the inspection of the power stations by the regulator. The relative severity of these events makes them rather easy and natural targets for the ASN inspectors during the routine inspections of nuclear stations.[24]

Table 2 and figure 4 provide descriptive statistics associated with the annual counts of automatic shut-downs and safeguard events.

3.3 Empirical specifications

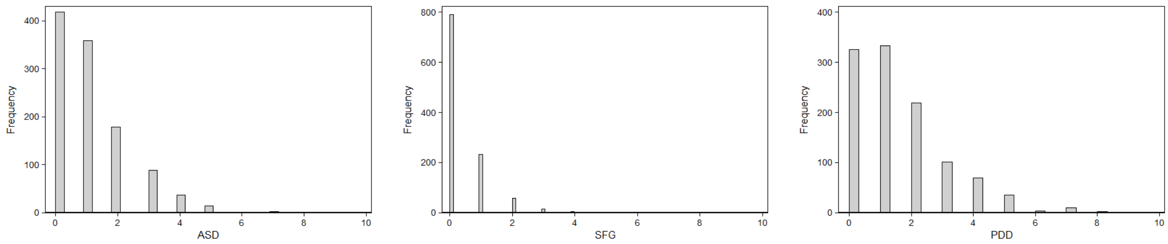
Let Y_{it} denote the counts of safety events declared during year t in reactor i, $t \in \{1997, \ldots, 2015\}$, $i \in \{1, \ldots, 58\}$. Let AGE_{it} be the age of reactor i in year of observation t, and let X_{it} denote a set of reactor and year specific control variables, such as whether the reactor is a first-of-a-kind

[22] In France, until 2000, the electricity transportation network was managed by EDF. Since 2000, transmission and production have been unbundled, and the transmission network has been handled by a single operator (RTE), which remains a subsidiary of EDF.

[23] Using automatic shut-downs as a proxy for nuclear safety is also supported by the fact that the annual number of automatic shut-downs is retained by the World Association of Nuclear Operators as one of their safety performance indicators. See for instance WANO's yearly performance reports on their website.

[24] Interviews conducted with both the ASN and EDF seem to suggest that making the assumption that safeguard events are perfectly detected and declared is reasonable.

Figure 4: Occurrences of perfectly detected and declared events per reactor.year



($FOAK$ in the following) or a first-of-a-site ($FOAS$ in the following). Our main results are based on an exponential specification of the conditional mean:

$$\mathbb{E}(Y_{it}|W_{it}) = \exp\Bigg(\beta \cdot X_{it} + \sum_{Year} \beta_{Year} \cdot \mathbb{1}_{Year} + \sum_{Site} \beta_{Site} \cdot \mathbb{1}_{Site} + \sum_{Site} \gamma_{Age,site} \cdot \mathbb{1}_{Site} \times Age_{it}\Bigg), \quad (4)$$

where exp denotes the exponential function and W_{it} denotes the regressors included in the model. Time dummies $\mathbb{1}_{Year}$ take the value of 1 when $t = Year$, and capture possible time trends or shocks associated with particular years, such as post-Fukushima-Daiichi safety upgrades. Site dummies $\mathbb{1}_{Site}$ take the value of 1 when reactor r belongs to $Site$. These fixed effects capture time-constant, site-specific unobserved sources of heterogeneity. Hence, coefficients β_{Year} and β_{Site} capture time and site-specific unobserved heterogeneity.

In addition, we include in equation (4) an AGE variable, which is interacted with site dummies $\mathbb{1}_{Site}$ in order to measure the effect of age within each nuclear power station. In the results section, we present two tables that correspond to two different definitions of the AGE variable. First, we define age as the decade of operation of a reactor during the year of observation. This coarse definition aims to measure a general trend associated with the ageing of reactors and their periodic decade inspections, which are characterized by major refurbishments.

We also estimate a specification where the age of a reactor is defined as the time elapsed since the first nuclear activity of its core[25]. In this specification, we also add another interacted variable obtained by multiplying site dummies with age squared (AGE^2). The second specification can

[25] Age could alternatively be measured with respect to other reference dates, such as the beginning of construction, the first connection to the grid, or the beginning of commercial operation. The date of the first divergence is chosen as it best captures the amount of radiations received by the reactor's different systems.

be described by the following equation:

$$\mathbb{E}(Y_{it}|W_{it}) = \exp\Bigg(\beta \cdot X_{it} + \sum_{Year} \beta_{Year} \cdot \mathbb{1}_{Year} + \sum_{Site} \beta_{Site} \cdot \mathbb{1}_{Site} \\ + \sum_{Site} \gamma_{Age,site} \cdot \mathbb{1}_{Site} \times Age_{it} + \sum_{Site} \gamma_{Age^2,site} \cdot \mathbb{1}_{Site} \times Age_{it}^2\Bigg), \quad (5)$$

By measuring simultaneously the effects of age and age squared on the occurrences of safety events, we test the existence of a *bathtub* trend. The *bathtub* trend can be presented as a three-state process (Chen et al., 2011). The first state is characterized by increases in reliability due to learning effects. It is followed by a steady state in which reliability remains constant. In the final state, reliability decreases as the system wears out.[26] Using equation (5), tentative evidence of the bathtub trend is obtained if the coefficients associated with the age variables are negative and significant, while the coefficients associated with the quadratic age variable are positive and significant.

The count specification presented in equations (4) and (5) has several advantages over the standard linear model, see e.g. Wooldridge (2002) or Cameron and Trivedi (2013). Most importantly, the linear model might produce negative predictions for feasible values of the observed covariates, something we would like to preclude. The quantitative interpretation of the coefficients is discussed in the results section. We estimate this model using a Negative Binomial estimator with quadratic over-dispersion and site-clustered standard errors. This estimator is preferred to the Poisson-QMLE estimator and to the Negative-Binomial estimator with linear over-dispersion. Model-selection is performed using the Akaike and Bayesian information criterion, whose values are reported in appendix B. One advantage of the Negative Binomial estimator is that it better fits over-dispersed data, consistently with the descriptive evidence in figure 4 on page 16. A second advantage of the Negative Binomial estimator is that it is more efficient than the Poisson-QMLE estimator in some cases, see Cameron and Trivedi (2013) for a discussion.[27] As a robustness check, however, we also estimate a linear specification.

Equations (4) and (5) are estimated using three baseline specifications, allowing us to answer the questions that motivate this paper. In specification (1), the dependent variable Y_{it} is defined as the annual count of automatic shut-downs per reactor (ASD). Then, we test whether the trends observed on automatic shut-downs can also be observed when using safeguard events

[26] The reliability literature identifies bathtub trends by estimating hazard rates and their variations across time (Aarset, 1987; Mudholkar and Srivastava, 1993; Xie et al., 2002).

[27] An advantage of the Poisson-QMLE estimator is that it is robust to functional form misspecification. Therefore, we check the robustness of our results using the Poisson estimator.

as a dependent variable. As automatic shut-downs interrupt the production of electricity of a reactor, they provide monetary incentives for plant managers to exert particular efforts to reduce their occurrences. Hence, in specification (2), the dependent variable Y_{it} is defined as the annual counts of safeguard events declared per reactor (SFG). All explanatory variables from specification (1) are left unchanged.

Specification (3) aims to measure the importance of transparency in the declaration process. For this purpose, we change the definition of the dependent variable in order to relax our restriction to the set of PDD events. Y_{it} is then defined as the unrestricted annual count of safety significant events per reactor (ALL). Following Rose (1990), we argue that transparency can be neglected if the results obtained under specification (3) match those obtained under the first two specifications. This claim will be discussed further in the results section.

4 Results and interpretation

4.1 Reactor age and nuclear safety

For a simple interpretation of the coefficients reported in tables 3 and 4, the results of an OLS regression using specification (1) is provided in the first column of both tables. For the other specifications, the coefficients reported are obtained through a Negative-Binomial regression model, and can be interpreted using incidence rate ratios: given any explanatory variable X and its coefficient β_X, e^{β_X} represents the ratio of the expected counts of events obtained after and before a unit increase of X.[28] For instance, using the notations defined in equation 4, $e^{\beta_{Age}}$ represents the (multiplicative) average effect of a unit increase in the age of a reactor on the expected number of occurrences of events.

In table 3, we provide the results of the estimations of specifications (1), (2) and (3) when the age variable is defined as the decade of operation of a reactor during the year of observation. Age is shown to have a negative and significant effect on occurrences of both automatic shut-downs and safeguard events. Quantitatively, we can interpret the coefficient of the decade variable in specification (1) as an average 42% decrease in automatic shut-down frequency with each decade of operation.

Though, the coefficients associated with the interaction of the decade variable with site dummies show that this effect is heterogeneous across power stations. However, in appendix

[28]When β_i is small for all i, β represents the vector of semi-elasticities of the dependent variable Y in the explanatory variables X. In addition, when coefficients are small and explanatory variables are included in logarithmic form, then β can be interpreted as a regular elasticity.

Table 3: Age and safety (age as decade of operation)

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ALL
FOAS	0.011	0.017	0.058	0.089***
FOAK	0.036	0.0093	-0.14	-0.12***
Decade	-0.31**	-0.53***	0.036	0.18***
*Decade*xBlayais	-0.14***	0.17***	-0.30***	-0.38***
*Decade*xBugey	-0.26***	0.24***	-1.04***	-0.18***
*Decade*xCattenom	0.68***	0.94***	-0.43***	-0.29***
*Decade*xChinon	0.30***	0.45***	-0.43***	-0.21***
*Decade*xChooz	0.0047	0.074	-0.80***	-0.46***
*Decade*xCivaux	-0.58***	-0.33***	-1.21***	-0.23***
*Decade*xCruas	0.72***	0.90***	-0.36***	-0.15***
*Decade*xDampierre	0.25***	0.38***	-0.88***	-0.47***
*Decade*xFessenheim	0.78***	0.86***	0.28***	-0.15***
*Decade*xFlamanville	0.16***	0.23***	-0.44***	-0.16***
*Decade*xGolfech	0.30***	0.41***	-0.98***	-0.41***
*Decade*xGravelines	0.29***	0.45***	-0.94***	-0.32***
*Decade*xNogent	0.088***	0.27***	0.089***	-0.38***
*Decade*xPaluel	-0.24***	0.27***	-0.72***	-0.39***
*Decade*xPenly	0.65***	0.84***	-0.57***	-0.40***
*Decade*xSt-Alban	0.36***	0.56***	-0.53***	-0.15***
*Decade*xSt-Laurent	0.11**	0.47***	-1.23***	-0.43***
*Decade*xTricastin	0.44***	0.65***	-0.48***	-0.12***
Site FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Site-clustered standard errors
*** p<0.01, ** p<0.05, * p<0.1
1100 observations. Omitted intercepts.

B.1, table 6 shows that in a majority of power stations, the effect of age remains negative and statistically significant. Notable exceptions are the Cattenom, Cruas, Fessenheim and Penly stations, in which occurrences of automatic shut-downs are characterized by a significant and positive effect of age.

This first result suggests that safety increases with the age of reactors, even when time and site specific effects are accounted for. Given the definition of the age variable used in table 3, a possible explanation of this effect is that the periodic decade inspections of nuclear power stations lead to safety upgrades that increase their safety.

Next, we turn to the analysis of the results presented in table 4. In this table, the age of a reactor is defined as the time elapsed since the first activity of its nuclear core. It appears that the coefficient associated to the *Age* variable is negative and significant while the coefficient of the squared age variable is positive and significant. This is the case for both automatic shut-downs and safeguard events. This result is consistent with the bathtub trend: at an early stage of the life of nuclear reactors, the effect of age is predominant and the occurrences of these two categories of events decrease with the age of reactors. During a subsequent period, the positive coefficient associated to the quadratic age term implies that this decreasing effect weakens as the age of the reactor increases.

Quantitatively, the decrease observed at the beginning of the life of a reactor is obtained by interpreting the coefficient of the AGE variable in regression (1). This coefficient is equivalent to an average 30% decrease of the frequency of automatic shut-downs with each additional year of age. The following example describes how the effect of quadratic age is assessed. When a reactor goes from age 20 to age 21, the squared age variable varies from 400 to 441. The average increase in occurrences of automatic shut-downs is thus equal to $\exp(41 \times \beta_{AGE^2})$. Using the results from specification (1), this is equivalent to a 50% increase in annual frequencies. However, this large effect is mitigated by the linear and negative effect of age.

Again, the coefficients of the interaction of the age or age-squared variables with site dummies show that the effect of age across power stations is heterogeneous. In appendix B.1, the results of table 7 show that these effects of age and squared-age are consistent with the bathtub trend in a majority of power stations. A notable exception is the Paluel power station, whose occurrences of automatic shut-downs exhibit a *reverse bathtub* trend. Automatic shut-downs in this station are characterized by a significant positive age trend and a significant negative quadratic age trend. Safeguard events do not exhibit this particular trend in this station.

To illustrate the heterogeneity, in appendix B.1, the first regression of table 9 presents the results of a regression in which site fixed effects are included, but the age variable is only interacted with capacity cohort dummies, corresponding to the three groups of reactors described in table 1 in section 2. This estimation shows no significant effect of age on the dependent variable in any capacity cohort. This result confirms that measuring the effect of age on the safety of a group of reactors (e.g. capacity cohorts) can neglect large heterogeneities across power stations.

The coefficients associated with the $FOAS$ and $FOAK$ variables in table 3 and 4 show that reactors which were built first in their site - or first in their capacity cohort - do not exhibit significantly different reporting behaviours. This suggests that the order of construction of reactors within a site is not a significant driver of their future safety level. This result is in contrast with the literature on the costs of construction of the French nuclear reactors, which showed evidence of learning-by-doing in terms of lead-time and construction costs within sites and technological designs (see e.g. Berthélemy and Rangel (2015); Rangel and Lévêque (2015)).

Table 4: Empirical evidence of the bathtub trend

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ASD
First of a Site	-0.046	-0.051	-0.044	0.028
First of a Kind	0.12	0.090	0.017	-0.069**
AGE	-0.35***	-0.37***	-0.36**	-0.045
AGE_reg_2	0.0096***	0.010***	0.012**	0.0029**
*Age*xBlayais	0.47***	0.48***	-0.60***	0.094***
*Age*xBugey	0.23***	0.31***	-0.43***	-0.18***
*Age*xCattenom	0.39***	0.43***	0.35***	0.15***
*Age*xChinon	0.23***	0.23***	0.14**	0.071***
*Age*xChooz	0.063	0.16***	-0.034	0.066***
*Age*xCivaux	0.067	0.24***	0.55***	0.15***
*Age*xCruas	0.27***	0.28***	-0.35***	-0.10***
*Age*xDampierre	-0.11***	-0.13***	0.013	-0.15***
*Age*xFessenheim	-0.81***	-0.32***	0.12	0.49***
*Age*xFlamanville	0.34***	0.42***	-0.22***	0.17***
*Age*xGolfech	0.24***	0.23***	-0.34***	0.0085
*Age*xGravelines	0.17***	0.18***	0.49***	-0.071***
*Age*xNogent	0.15***	0.22***	0.064***	-0.00084
*Age*xPaluel	0.68***	0.60***	-0.51***	0.035***
*Age*xPenly	0.41***	0.42***	0.68***	0.12***
*Age*xSt-Alban	0.47***	0.46***	-0.58***	0.00080
*Age*xSt-Laurent	0.41***	0.38***	-1.30***	-0.034*
*Age*xTricastin	0.086**	0.13***	-0.42***	0.018
Age^2xBlayais	-0.012***	-0.012***	0.0092***	-0.0037***
Age^2xBugey	-0.0082***	-0.0089***	0.0013	0.0018***
Age^2xCattenom	-0.0089***	-0.0097***	-0.012***	-0.0053***
Age^2xChinon	-0.0060***	-0.0058***	-0.0062***	-0.0028***
Age^2xChooz	0.0042***	0.00079***	0.0079***	-0.0035***
Age^2xCivaux	0.0013*	-0.0052***	-0.017***	-0.0060***
Age^2xCruas	-0.0063***	-0.0064***	0.0039***	0.0012***
Age^2xDampierre	0.000046**	0.00032***	-0.0055***	0.00094***
Age^2xFessenheim	0.012***	0.0033***	-0.0049***	-0.0097***
Age^2xFlamanville	-0.0092***	-0.011***	0.0034***	-0.0050***
Age^2xGolfech	-0.0044***	-0.0038***	0.013***	-0.0016***
Age^2xGravelines	-0.0051***	-0.0051***	-0.017***	-0.00024
Age^2xNogent	-0.0044***	-0.0061***	-0.0010***	-0.0012***
Age^2xPaluel	-0.018***	-0.016***	0.0088***	-0.0024***
Age^2xPenly	-0.0093***	-0.0093***	-0.024***	-0.0050***
Age^2xSt-Alban	-0.012***	-0.011***	0.012***	-0.00093***
Age^2xSt-Laurent	-0.011***	-0.0095***	0.020***	-0.0014***
Age^2xTricastin	-0.0038***	-0.0046***	0.0043***	-0.0017***

Site-clustered standard errors

*** p<0.01, ** p<0.05, * p<0.1

1100 observations. Omitted intercepts.

Although initial differences in safety across reactors within a power station may have existed shortly after the beginning of their operation, knowledge spillovers in management probably unified safety across reactors within sites. In this case, these spillovers would be captured by site fixed effects, which in turn would explain why the coefficient associated with the $FOAS$ variable is not statistically significant. As a robustness check, when no site fixed effects are included in the regression, the coefficient associated with the $FOAS$ variable becomes negative and statistically significant. This result provides additional empirical evidence in favour of this interpretation, and is provided in appendix B.1 in table 9.

A second interesting result is provided in the second regression presented in table 9 in the appendices. Here, age is shown to have a significant effect on occurrences of automatic shut-downs in both the 900 MW cohort and the 1300 MW cohort when only capacity cohort fixed effects are included. These results suggest that automatic shut-down frequencies increase when these reactors get older. In addition, all year fixed effects become statistically significant and suggest increases in safety over time, whereas year fixed effects were not significant when site-fixed effects were included. This is an illustration of the ACP problem raised in section 3.1: omitting the cohort variable C (i.e. site fixed effects) can bias the results of the estimations of the effects of age A and period P on the dependent variable. In this case, the effect of age is biased upwards, while the coefficients associated with the time dummies are biased downwards.

Finally, the results of the Negative Binomial regressions using specification (1) presented in tables 4 and 3 can be compared to the results obtained using the OLS estimator and using the same empirical specification. The results are fairly similar in signs and statistical significance, but the OLS estimator provides coefficients whose absolute values are (in nearly all cases) smaller. This can be seen as a test of the robustness of our results to model misspecification. An additional robustness check is provided in appendix B.1, where table 8 presents the results of two regressions in which site fixed effects are replaced by reactor fixed effects. To avoid the multicollinearity issue, time fixed effects are replaced by three period dummies, which respectively take the value of 1 when the year of observation belongs to the periods 1997-2003, 2004-2009, 2010-2015. Our results regarding automatic shut-downs are robust to this additional specification.

4.2 Transparency

We now turn to the interpretation of specification (3), in which the dependent count variable is defined as the total number of events declared by nuclear reactors per year, without any restriction to specific types of events. Results of the regressions using this specification are reported in table 3 and 4. The coefficients estimated in these two regressions largely differ from the coefficients obtained under specifications (1) and (2). We now propose possible interpretations of this result.

Let the random variable T denote the type of an event, with $T = PDD$ being one of the possible types. Assume first that the effect of age on the probability of occurrence of an event is the same for the full sample and the restricted set Θ_{PDD}:

$$\frac{\partial \mathbb{E}\{Y|W,C\}}{\partial Age} = \frac{\partial \mathbb{E}\{Y|W,C,T = PDD\}}{\partial Age}. \tag{A4}$$

Assumption (A4) is an external validity assumption. Then, if transparency plays no role in the declaration process, we should observe similar results for specifications (1), (2) and (3). We however fail to find similar results in specification (3). Thus, either transparency does indeed bias the results, or assumption (A4) fails to hold.

However, assumption (A4) can be defended. First, comparison of regression (1) and (2) revealed no particular differences on the subsets of automatic shut-downs and safeguard events. Second, it can be argued that numerous investments in safety will have positive spillovers for all types of safety events. For instance, hiring skilled employees, investing in the training of safety engineers, or enhancing organizational practices are safety investments which will decrease the probabilities of occurrence of safety events, regardless of their types.

Provided assumption (A4) holds, our finding provides evidence that transparency can bias the analysis of the reports of nuclear safety significant events. Thus, neglecting the unobserved changes in declaration criteria, detection abilities, and rate of non-compliance with declaration criteria could bias the estimation of safety variations with age and technology.[29]

5 Implications

The takeaways from this paper are the following. First, we describe a novel dataset obtained from the French Nuclear Safety Authority, encompassing over 19.000 significant safety events declared in the French fleet between 1973 and 2015. This dataset contains more events than previous datasets used for the assessments of the evolutions of nuclear safety. It also contains information of a better quality, as the declarations it gathers are verified by both nuclear plant

[29] This result contrasts the results of Rose (1990), who discarded the importance of pilots' subjectivity in airline incidents reports.

managers and experts from the French safety regulator, whereas previous studies used press articles and academic publications to build their datasets (Sovacool, 2008; Wheatley et al., 2017).

Second, restricting this dataset to a subset of events characterized by perfect detection and declaration, we disentangle the effects of safety and transparency on the occurrences of safety significant incidents. Our results are consistent over both types of perfectly detected and declared events considered. This finding supports the hypothesis that our results are not driven by specific efforts exerted by plant managers and dedicated to one particular type of events. On the other hand, our results are not robust to considering the unrestricted counts of safety events declared in each nuclear power reactor, which suggests that the level of transparency of plant managers during the declaration process introduces some bias which precludes the observation of safety variations when studying the complete set of safety events.

Third, from a quantitative perspective, we observe that reactor safety is positively affected by the ageing of reactors, but suffers from negative quadratic effect of ageing. This finding is consistent with the *bathtub* trend from the reliability literature, which expects the reliability of technical systems to improve with the early age of system due to learning, and to subsequently deteriorate as the system wears out. It also appears that measuring aggregated age effects over capacity cohorts fails to capture subsequent heterogeneity across reactors. Finally, we show that omitting cohort effects when measuring the effect of the ageing of nuclear reactors leads to substantial biases in the estimation. This has important consequences regarding the on-going debate regarding the early closure of nuclear plants, as omitting cohort effects can lead to substantial overestimations of the effect of age on safety.

Finally, we can derive some implications of these results for future nuclear policy. First, even though our study stands on French data, our results could have implications beyond the French fleet, as the pressurized water technology is the most widely used nuclear reactor technology, with 277 reactors operated on the planet according to the World Nuclear Association. Second, our results suggest that focusing on the age of nuclear reactors when studying their safety, for instance by having large public debates regarding the maximum lifespan of nuclear reactors, may be ill-advised. Instead, this paper suggests to focus on the heterogeneities that characterize the safety of each nuclear site. In other words, setting a maximum lifespan irrespective of technological idiosyncrasies could be an inefficient policy, entailing premature shut-downs of safe reactors or prolonged operation of unsafe ones.

Table 5: Test-statistics associated with the Poisson and NB regression models

Regression model	Over-dispersion	S-E clusters	Ln L	Pearson	AIC	BIC
Poisson	-	Reactor	-1473	1198	3060	3345
Poisson	-	Site	-1473	1198	2982	3072
Neg. Bin.	NB1	Reactor	-1471	-	3056	3341
Neg. Bin.	NB1	Site	-1471	-	2978	3068
Neg. Bin.	NB2	Reactor	-1471	-	3056	3341
Neg. Bin.	NB2	Site	-1471	-	2978	3068

Test-statistics based on specification (1).

A Appendix 1: Proof of lemma 3.1

Proof of lemma 3.1. Note that

$$\mathbb{E}\{Y|W=w\} = \mathbb{E}\{\mathbb{E}\{Y|W=w,D,O\}\} = \mathbb{E}\{f(D,O)\}, \tag{6}$$

with $f(D,O) = \mathbb{E}\{Y|W=w,D,O\}$. In the following, we suppress the dependency on W for the sake of representational simplicity. It holds

$$\mathbb{E}\{f(D,O)\} = \sum_{d,o} f(d,o)\mathbb{P}\{D=d,O=o\}. \tag{7}$$

We observe that due to assumption A2, $\mathbb{E}\{Y|O=0\} = \mathbb{P}\{Y=1|O=0\} = 0$, and thus, under A2 and A3, equation (7) gives

$$\mathbb{E}\{Y|W=w\} = f(1,1)\mathbb{P}\{D=1,O=1\} = \mathbb{E}\{Y|W=w,D=1,O=1\}. \tag{8}$$

□

B Appendix 2: Model selection

Calculations of the Pearson statistics for the Poisson regressions (using specification 1 and 2) allow to strongly reject the Poisson distribution for both automatic shut-downs and safeguard events. Moreover, the standard Akaike and Bayesian information criteria (AIC and BIC), computed using specification (1) under linear and quadratic over-dispersion support the use of the negative binomial models with site-clustered standard errors and quadratic over-dispersion specification. These statistics are gathered in table 5.

Table 6: Robustness checks: reference group under age as decade

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ASD
FOAS	0.011	0.017	0.058	0.089***
FOAK	0.036	0.0093	-0.14	-0.12***
*Decade*xBelleville	-0.31**	-0.53***	0.036	0.18***
*Decade*xBlayais	-0.44***	-0.36***	-0.26	-0.20***
*Decade*xBugey	-0.57***	-0.29***	-1.01***	0.00011
*Decade*xCattenom	0.38***	0.41***	-0.40**	-0.11***
*Decade*xChinon	-0.0033	-0.075	-0.39**	-0.033
*Decade*xChooz	-0.30*	-0.45***	-0.76***	-0.28***
*Decade*xCivaux	-0.89***	-0.86***	-1.17***	-0.054
*Decade*xCruas	0.41***	0.37***	-0.32	0.021
*Decade*xDampierre	-0.055	-0.14	-0.84***	-0.30***
*Decade*xFessenheim	0.48***	0.34***	0.31*	0.027
*Decade*xFlamanville	-0.15	-0.29***	-0.40**	0.020
*Decade*xGolfech	-0.0096	-0.12	-0.94***	-0.23***
*Decade*xGravelines	-0.019	-0.071	-0.90***	-0.14***
*Decade*xNogent	-0.22*	-0.26***	0.12	-0.20***
*Decade*xPaluel	-0.55***	-0.26**	-0.68***	-0.22***
*Decade*xPenly	0.35***	0.32***	-0.54***	-0.22***
*Decade*xSt-Alban	0.056	0.039	-0.49***	0.022
*Decade*xSt-Laurent	-0.19	-0.053	-1.20***	-0.25***
*Decade*xTricastin	0.13	0.12	-0.45**	0.058
Site FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Site-clustered standard-errors
*** p<0.01, ** p<0.05, * p<0.1
1,100 Observations. Omitted intercepts.

B.1 Appendix 3: Robustness checks

References

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1):106–108.

Bell, A. and Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social science & medicine*, 93:163–165.

Bell, A. and Jones, K. (2014). Another'futile quest'? a simulation study of yang and land's hierarchical age-period-cohort model. *Demographic Research*, 30:333.

Bell, A. and Jones, K. (2016). The hierarchical age-period-cohort model: Why does it find the results that it finds? *Quality & Quantity*, pages 1–17.

Berthélemy, M. and Rangel, L. E. (2015). Nuclear reactors' construction costs: The role of lead-time, standardization and technological progress. *Energy Policy*, 82:118–130.

Table 7: Robustness check: reference group in the study of the bathtub trend

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ASD
First of a Site	-0.046	-0.051	-0.044	0.028
First of a Kind	0.12	0.090	0.017	-0.069**
AGExBelleville	-0.35***	-0.37***	-0.36**	-0.045
AGExBlayais	0.11	0.10	-0.96***	0.048
AGExBugey	-0.12	-0.065	-0.79***	-0.23***
AGExCattenom	0.036	0.060	-0.0063	0.10***
AGExChinon	-0.12*	-0.14***	-0.22*	0.025
AGExChooz	-0.29***	-0.22***	-0.39***	0.021
AGExCivaux	-0.29***	-0.13***	0.19*	0.10***
AGExCruas	-0.086	-0.093	-0.71***	-0.15***
AGExDampierre	-0.47***	-0.50***	-0.35	-0.20***
AGExFessenheim	-1.16***	-0.70***	-0.24	0.45***
AGExFlamanville	-0.013	0.046	-0.58***	0.13**
AGExGolfech	-0.11**	-0.14***	-0.70***	-0.037
AGExGravelines	-0.18***	-0.20***	0.13	-0.12***
AGExNogent	-0.20**	-0.15*	-0.29	-0.046
AGExPaluel	0.33***	0.23***	-0.87***	-0.011
AGExPenly	0.052	0.051	0.32**	0.071**
AGExSt-Alban	0.12	0.086	-0.94***	-0.045
AGExSt-Laurent	0.054	0.0011	-1.66***	-0.080
AGExTricastin	-0.27**	-0.24**	-0.78***	-0.028
AGE^2xBelleville	0.0096***	0.010***	0.012**	0.0029**
AGE^2xBlayais	-0.0027	-0.0021	0.021***	-0.00081
AGE^2xBugey	0.0013	0.0012	0.013***	0.0047***
AGE^2xCattenom	0.00067	0.00048	0.000032	-0.0024***
AGE^2xChinon	0.0035***	0.0043***	0.0059**	0.000077
AGE^2xChooz	0.014***	0.011***	0.020***	-0.00063
AGE^2xCivaux	0.011***	0.0049**	-0.0045	-0.0031**
AGE^2xCruas	0.0033	0.0038	0.016***	0.0041***
AGE^2xDampierre	0.0096***	0.010***	0.0066	0.0038***
AGE^2xFessenheim	0.022***	0.013***	0.0072	-0.0068***
AGE^2xFlamanville	0.00037	-0.0012	0.016***	-0.0021
AGE^2xGolfech	0.0051***	0.0063***	0.025***	0.0013
AGE^2xGravelines	0.0044***	0.0050***	-0.0049*	0.0027***
AGE^2xNogent	0.0052**	0.0040	0.011**	0.0017
AGE^2xPaluel	-0.0087***	-0.0057**	0.021***	0.00045
AGE^2xPenly	0.00030	0.00081	-0.012**	-0.0021*
AGE^2xSt-Alban	-0.0023	-0.0011	0.024***	0.0020
AGE^2xSt-Laurent	-0.00098	0.00064	0.032***	0.0015
AGE^2xTricastin	0.0057**	0.0056**	0.016***	0.0012
Site FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Site-clustered standard errors
*** p<0.01, ** p<0.05, * p<0.1
1100 observations. Omitted intercepts.

Table 8: Robustness checks: long periods and reactor fixed-effects

VARIABLES	(1) ASD	(2) SFG
AGE	-0.38***	0.0017
AGE^2	0.0081***	0.00040
AgexBlayais	0.56***	-0.52***
AgexBugey	0.41***	-0.097*
AgexCattenom	0.38***	0.23***
AgexChinon	0.20***	0.067**
AgexChooz	0.13***	-0.21***
AgexCivaux	0.24***	0.26***
AgexCruas	0.32***	-0.19***
AgexDampierre	-0.054**	0.25***
AgexFessenheim	-0.25***	0.28***
AgexFlamanville	0.48***	-0.16***
AgexGolfech	0.24***	-0.55***
AgexGravelines	0.21***	0.50***
AgexNogent	0.25***	0.043***
AgexPaluel	0.69***	-0.45***
AgexPenly	0.43***	0.62***
AgexSt-Alban	0.51***	-0.56***
AgexSt-Laurent	0.43***	-1.01***
AgexTricastin	0.20***	-0.21***
Age^2xBlayais	-0.013***	0.010***
Age^2xBugey	-0.0100***	-0.0011
Age^2xCattenom	-0.0087***	-0.0091***
Age^2xChinon	-0.0049***	-0.0024**
Age^2xChooz	0.00062***	0.0064***
Age^2xCivaux	-0.0075***	-0.014***
Age^2xCruas	-0.0068***	0.0022***
Age^2xDampierre	-0.00040***	-0.0072***
Age^2xFessenheim	0.0028***	-0.0035***
Age^2xFlamanville	-0.013***	0.0029***
Age^2xGolfech	-0.0038***	0.016***
Age^2xGravelines	-0.0051***	-0.014***
Age^2xNogent	-0.0066***	-0.00050***
Age^2xPaluel	-0.017***	0.0091***
Age^2xPenly	-0.0095***	-0.025***
Age^2xSt-Alban	-0.012***	0.013***
Age^2xSt-Laurent	-0.0099***	0.018***
Age^2xTricastin	-0.0052***	0.0035***
$\mathbb{1}_{1997-2002}$	-0.19	1.00
$\mathbb{1}_{2003-2008}$	-0.019	0.69
$\mathbb{1}_{2009-2014}$	-0.021	0.94*
Reactor FE	Yes	Yes

Site-clustered standard errors
*** p<0.01, ** p<0.05, * p<0.1
1100 observations. Omitted intercepts.

Table 9: Analysis of the ACP bias

VARIABLES	(1) ASD	(1) ASD
AGEx900MW	0.044	0.087***
AGEx1300MW	0.035	0.079***
AGEx1450MW	-0.031	0.0087
FOAS	-0.060	-0.19**
FOAK	0.076	0.15
class of reactor = 1300		0.53*
class of reactor = 1450		1.96***
Blayais	0.12	
Bugey	0.26	
Cattenom	0.0098	
Chinon B	-0.20	
Chooz B	1.02*	
Civaux	1.47**	
Cruas	-0.0085	
Dampierre	-0.35	
Fessenheim	0.16	
Flamanville	0.0014	
Golfech	0.11	
Gravelines	-0.25	
Nogent	0.25***	
Paluel	0.65**	
Penly	0.41**	
Saint-Alban	0.34***	
Saint-Laurent	0.30	
Tricastin	0.064	
1998	-0.32	-0.35*
1999	-0.49**	-0.56***
2000	-0.48*	-0.59***
2001	-0.32	-0.47**
2002	-0.35	-0.55**
2003	-0.31	-0.54***
2004	-0.60	-0.88***
2005	-0.64	-0.96***
2006	-0.64	-1.02***
2007	-0.74	-1.15***
2008	-1.47**	-1.93***
2009	-0.94	-1.43***
2010	-1.20	-1.74***
2011	-1.22	-1.80***
2012	-1.37	-2.01***
2013	-1.21	-1.89***
2014	-1.21	-1.93***
2015	-1.37	-2.13***

Site-clustered standard errors
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
1,100 observations. Omitted intercepts.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chen, Y., Wang, Z., Qiu, J., Zheng, B., and Huang, H.-Z. (2011). Adaptive bathtub hazard rate curve modelling via transformed radial basis functions. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2011 International Conference on*, pages 110–114. IEEE.

Faure, M. G. and Skogh, G. (1992). Compensation for damages caused by nuclear accidents: A convention as insurance. *The Geneva Papers on Risk and Insurance*, 17:499–513.

Feinstein, J. S. (1989). The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections and Abnormal Occurences. *Journal of Political Economy*, 97:115–154.

Glenn, N. D. (2005). *Cohort analysis*, volume 5. Sage.

Hansson, I. and Skogh, G. (1987). Moral hazard and safety regulation. *The Geneva Papers on Risk and Insurance*, 12:132–144.

Hausman, C. (2014). Corporate incentives and nuclear safety. *American Economic Journal: Economic Policy*, 6(3):178–206.

Hirschberg, S., Burgherr, P., Spiekerman, G., and Dones, R. (2004). Severe accidents in the energy sector: comparative perspective. *Journal of Hazardous Materials*, 111(1):57–65.

Hofert, M. and Wüthrich, M. V. (2011). Statistical review of nuclear power accidents. *Asia-Pacific Journal of Risk and Insurance*, 7:1–13.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, pages 311–324.

Holford, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual review of public health*, 12(1):425–457.

Keyes, K. M., Grant, B. F., and Hasin, D. S. (2008). Evidence for a closing gender gap in alcohol use, abuse, and dependence in the united states population. *Drug and alcohol dependence*, 93(1):21–29.

Keyes, K. M., Utz, R. L., Robinson, W., and Li, G. (2010). What is a cohort effect? comparison of three statistical methods for modeling cohort effects in obesity prevalence in the united states, 1971–2006. *Social science & medicine*, 70(7):1100–1108.

Laffont, J.-J. (1995). Regulation, moral hazard and insurance of environmental risks. *Journal of Public Economics*, pages 319–336.

Land, K. C., Zang, E., Fu, Q., Guo, X., Jeon, S. Y., and Reither, E. N. (2016). Playing with the rules and making misleading statements: a response to Luo, Hodges, Winship, and Powers. *American journal of sociology*, 122(3):962–973.

Luo, L., Hodges, J., Winship, C., and Powers, D. (2016). The sensitivity of the intrinsic estimator to coding schemes: comment on yang, schulhofer-wohl, fu, and land. *American Journal of Sociology*, 122(3):930–961.

Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American sociological review*, pages 242–258.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2):299–302.

O'Brien, R. M. (2011). The age–period–cohort conundrum as two fundamental problems. *Quality & Quantity*, 45(6):1429–1444.

O'Brien, R. M. (2017). Mixed models, linear dependency, and identification in age-period-cohort models. *Statistics in Medicine*, 36(16):2590–2600.

Rangel, L. E. and Lévêque, F. (2015). Revisiting the cost escalation curse of nuclear power: New lessons from the french experience. econ. *Economics of Energy and Environmental Policy*, 4(2).

Rose, N. L. (1990). Profitability and product quality: Economic determinants of airline safety performance. *Journal of Political Economy*, 98(5):944–964.

Selvin, S. (2004). *Statistical analysis of epidemiologic data.* Oxford University Press.

Shavell, S. (1984). A model of the optimal use of liability and safety regulation. *Rand Journal of Economics*, 15:271–280.

Sovacool, B. K. (2008). The costs of failure: A preliminary assessment of major energy accidents,1907-2007. *Energy Policy*, 36(5):1802–1820.

Suzuki, E. (2012). Time changes, so do people. *Social Science & Medicine*, 75(3):452–456.

Tukey, J. W. (1977). Exploratory data analysis.

Wheatley, S., Sovacool, B., and Sornette, D. (2017). Of disasters and dragon kings: a statistical analysis of nuclear power incidents and accidents. *Risk analysis*, 37(1):99–115.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data. *Cambridge, MA: MIT Press.*

Xie, M., Tang, Y., and Goh, T. N. (2002). A modified weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76(3):279–285.

Yang, Y. and Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological methodology*, 36(1):75–97.

Yang, Y. and Land, K. C. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications.* CRC Press.