Universität St.Gallen

Are Older Nuclear Reactors Less Safe? Evidence from France

Romain Bizet, Petyo Bonev, François Lévêque

November 2022 Discussion Paper no. 2022-16

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor: Mirela Keuschnigg
University of St.Gallen
School of Economics and Political Science
Department of Economics
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen

Electronic Publication: http://www.seps.unisg.ch

Are Older Nuclear Reactors Less Safe? Evidence from France[1]

Romain Bizet, Petyo Bonev, François Lévêque

Author's address: Petyo Bonev
Swiss Institute for Empirical Economic Research
University of St. Gallen
Varnbüelstrasse 14
9008 St. Gallen, Switzerland
Email petyo.bonev@unisg.ch

Romain Bizet - Charles River Associates,
François Lévêque - Mines ParisTech, PSL - Research University

[1] We thank the French Nuclear Safety Authority and the Institute for Radiological Protection and Nuclear Safety who provided the data. We would also like to thank the participants of the Young Energy Economist and Engineers seminar (Edinburgh University) for their valuable comments and ideas.

Abstract

This is the first paper to statistically evaluate the relationship between age of a nuclear power plant and nuclear safety. We use a novel dataset that contains over 13000 small incidents in the French fleet between 1997 and 2015. We find that after an initial period of increase, safety eventually decreases with age. This is consistent with the bathtub effect usually measured in the reliability literature.

Keywords

environmental policy, nuclear power, safety, incidents, age,

JEL Classification

Q58

1 Introduction

Are older nuclear reactors less safe? This question has been at the center of political and scientific debates for more than two decades,[1] since U.S. reactors started to apply for and obtain operating license renewals, extending their lifespans from 40 to 60 years. Opposing views of experts have been reviewed by Tollefson (2016). Advocates of license extensions argue that nuclear reactors do not age, as most of their parts are replaced during their lifespans. Skeptics reply that the uncertainties associated with surveillance, replacement and repair techniques could prevent the observation of some aging patterns. With the European Commission declaring nuclear and gas to be “green” sources of energy, and with repeated Russian military interventions close to Ukrainian nuclear power plants, the debate has been rolled out anew.[2]

Despite its high political and social importance, the question of reactor aging has not been answered in a satisfactory way by the existing literature. The main reason is that major safety accidents are rare - a crucial limitation for statistical inference. The existing literature has therefore relied on datasets composed of accidents from both nuclear reactors and fuel cycle facilities (Hirschberg et al., 2004; Sovacool, 2008; Hofert and Wüthrich, 2011; Wheatley et al., 2017). Although these studies represent important contributions for understanding safety, their internal validity remains questionable.

In this paper, we empirically investigate the relationship between age of a nuclear power plant and nuclear safety. To circumvent the problem of lack of observations, we propose to measure safety indirectly by the number of smaller incidents (such as automatic reactor shutdowns) that occur in a nuclear power plant. Although of minor importance, such incidents are considered as potentially predictive for major accidents.[3] We use a

[1] See e.g. the following wide-audience media article published in 2009 by the Scientific American, https://www.scientificamerican.com/article/nuclear-power-plant-aging-reactor-replacement-/ or in 2015 by the MIT Technology Review, https://www.technologyreview.com/s/544211/how-old-is-too-old-for-a-nuclear-reactor/.

[2] See for example the recent article by the Economist: https://www.economist.com/graphic-detail/2022/07/19/how-safe-is-nuclear-energy.

[3] Considering minor incidents as a measure of safety is in fact common in the current regulatory practice. 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.

novel dataset which contains more than 13000 nuclear safety incidents, referred to as *significant safety events*in the literature, that occurred between 1997 and 2015 in French nuclear reactors. These events represent the most significant deviations from the general standards of operation of nuclear reactors.

To the best of our knowledge, this is the first paper to use significant safety events in order to evaluate the relationship between the age of a nuclear reactor and its safety.[4] Our main empirical finding is that the relationship between age and safety is non-linear. In particular, in the majority of the reactors, we find an initial decrease of occurrences of significant safety events which is followed by a subsequent increase. For the remaining reactors, we find similar although insignificant trends. This indicates that safety evolves with the age of reactors consistently with the *bathtub* trend (see e.g. Aarset (1987); Mudholkar and Srivastava (1993); Xie et al. (2002) or Chen et al. (2011)), usually described as a three-step process composed of an initial increase in reliability due to learning effects, a subsequent phase of constant reliability, and a final phase during which reliability decreases as the system wears out.

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 the empirical specifications. Section 4 exposes our results and section 5 concludes. An online appendix contains supplemental material.

2 Institutional setup and data

2.1 Institutional setup

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

[4] More generally, this is one of only four papers to use significant safety events within a statistical analysis, the other three being Bizet et al. (2022), Davis and Wolfram (2012), and Hausman (2014). The paper by Feinstein (1989) uses a related measure of “abnormal occurrences” in the context of monitoring and safety.

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.

utility (EDF). These reactors were built in separate phases from the late 1970s to the late 1990s. The technological design of reactors evolved over the construction of successive reactors. For instance, reactors differ in their nominal capacity, the nature of their fuel, or in their ability to perform load-following. French 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 more minor 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).

Significant safety events. In France, nuclear safety is regulated by the Nuclear Safety Authority (ASN in the following), who defines regulatory standards for the operation of nuclear reactors. In particular, regulatory standards include mandatory reporting criteria that define a detailed list of particular situations, in which there is a deviation from safety standards. Such events are referred as to *significant safety events* (in the following simply "safety events").[5] Managers are required to report safety events to the

[5]The term "safety event" is broadly used by regulators and utilities in the nuclear industry throughout the world. It is also used by EDF and the ASN.

regulator because, as the name suggests, these events are deemed significant for safety. Examples for such events are degradation of safety equipment, a leakage from the primary cooling circuit, and an unanticipated reactor shutdown. The regulator uses the reports on safety events in several ways. First, the reports are used to aggregate information and share experience and best practices among reactors, and to detect generic defaults in the reactors' designs. In particular, even though the typical safety event 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. This justifies why safety events can be used as an indirect description of the safety of a nuclear reactor. Second, the regulator may use a report to impose a safety measure on a given reactor. As an example, in 2017, upon the report of a safety event, ASN shut down all reactors the nuclear power station Tricastin for over a month. The reactors were only opened after safety improvements required by ASN were implemented by the operator. Third, ASN issues an annual report which contains information on the significant safety events. The purpose of this report is to inform the public on the state of the nuclear power fleet.

The self-audit and reporting process. To comply with the reporting criteria set by ASN, plant operators gather information, on a daily basis, on a broad set of situations which depart from normal operation. All events are analyzed by plant managers. The subset of events that match the reporting criteria have to be reported to the safety authority. Upon the detection of a significant safety event, plant managers have two weeks to provide the regulator with a detailed summary of the event and an analysis of its causes and consequences. Plant managers also keep a record of all detected situations, regardless of there significance, in case of an inspection by the regulator.

An important feature of this declaration process is that plant managers analyze a large number of situations, and decide which ones should be reported to the safety authority. A common issue associated with these voluntary reporting processes is the existence of missing observations, in the sense of detection failures (if an event goes undetected by a plant manager) or reporting failures (if an event is not reported despite its significance).[6] In particular, reporting failures could occur for multiple reasons: an operator could misinterpret the significance of a situation and fail to report it, or deliberately decide not to report the event.[7]

The ASN enforces this reporting procedure through periodic and random inspections, during which inspectors get access to the events filed by plant managers as too insignificant to be reported. If inspectors come across a file describing an event which should have been reported as significant for safety, the regulator can engage in punitive actions against the operator, such as lawsuits or temporary shut-downs. Although inspectors get access to all the situations considered by plant managers, situations leading to no declaration are too numerous for inspectors to review them all thoroughly. As a result, inspectors review only a subset of events detected by operators but not reported to the authority.

2.2 Data and descriptive statistics

Our dataset, obtained from the ASN, contains all 13482 significant safety events reported from 1997 to 2015 in the French nuclear fleet.[8] Each event is characterized by a set of variables, describing 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. Details regarding the deletion or multiple count of some particular types of significant safety events are provided in appendix A.

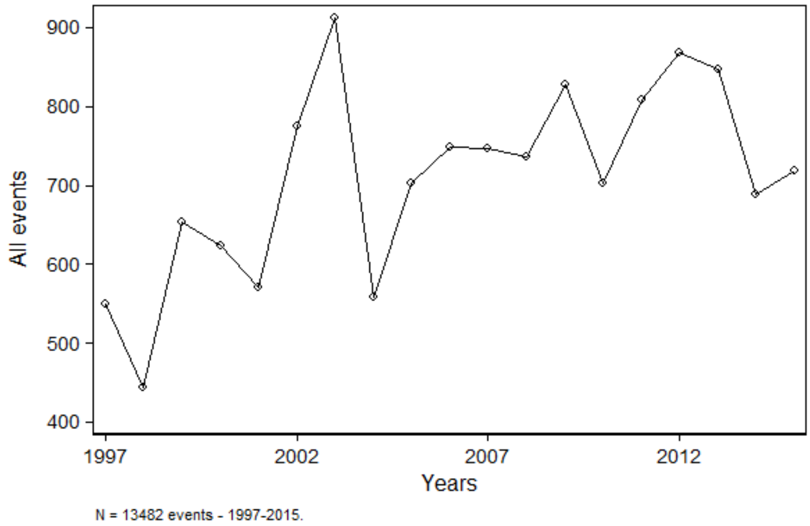
Figure 1 presents the evolution of the total number of reports over time in the French

[6]This issue has been thoroughly documented, from an empirical and theoretical perspective, in other industries. See e.g. Gray and Shimshack (2011) or Shimshack (2014) for recent reviews of this literature.

[7]We stress the fact that we do not necessarily describe detection and reporting failure as *moral hazard*, as the failure to detect or report an event need not be malignant nor intentional. This point is discussed further in the identification section.

[8]In other words, among all the situations detected by plant managers, our dataset is constituted of the events which were reported to the safety authority. Situations reported during the lifetime of

Figure 1: Declarations of significant safety events in the French fleet between 1997 and 2015



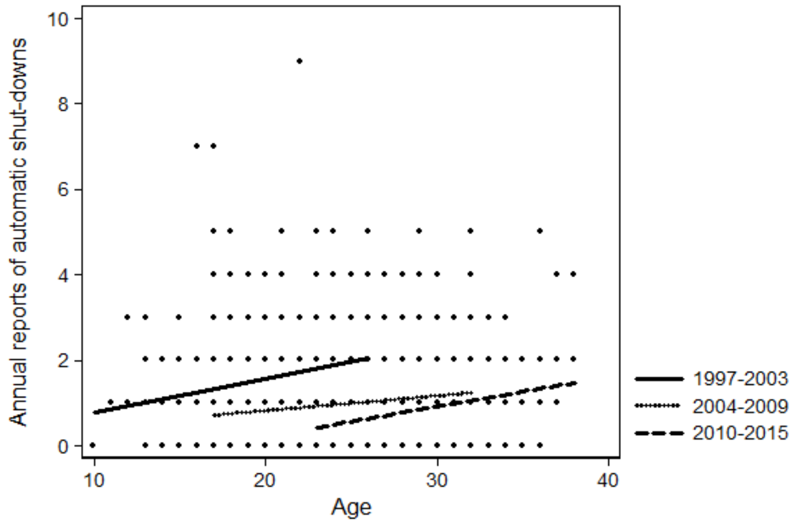
fleet from 1997 to 2015. On average, there are 700 events reported each year in the fleet, or one significant safety event per reactor and per month. Figure 1 shows that reports of significant safety events apparently follow an increasing time trend. This could be due to increases in the stringency of the safety standards, leading the regulator to consider new types of events as significant for safety. An increase in the ability of the operator to detect significant safety events, or a deterioration of nuclear safety with the aging of the fleet, could also explain this increase. The distribution of reports of events across groups of reactors of similar age, technologies, and location are presented and discussed in Section B in the online appendix.

Next, figure 2 displays 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-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,

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.

Figure 2: Reports of automatic shut-downs as a function of reactor age



the solid line corresponds to the reactor.years observed between 1997 and 2003. This graphical representation technique is used for example by Keyes et al. (2010). In our case, the division into three periods amounts to implicitly grouping reactors.years into three aggregated cohorts. The graph reveals two patterns. First, the regression fit of a more recent cohort lies below the regression fit of a more ancient cohort. This could be due, for instance, to learning effects. Second, within each cohort, the older the reactor, the more numerous the reports of automatic shut-downs.

Perfectly detectable and declarable events A subgroup of safety events, called perfectly detectable and declarable (PDD), are of particular importance for our causal analysis. We are now describing this category of events. For any particular type of safety incidents, there are two conditions that guarantee perfect detection and declaration. 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[10] 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 remain unnoticed, which

[10] 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.

Table 2: Descriptive statistics for three dependent count variables

Variable	Definition	Mean	Std. Dev.
ASD_{it}	Automatic shut-downs	1.122	1.233
SFG_{it}	Events requiring the use of safeguard systems	0.377	0.693
ALL_{it}	All events	12.256	5.105

ASN data. This table provides summary statistics of three count variables. Variable ALL presents the total number of significant safety events reported in year t in a reactor i. Variable ASD and SFG focus on specific types of significant safety events. ASD provides the specific number of automatic shut-downs. Variable SFG provides the counts per reactor and per year of safeguard events. These variables are defined over a panel of 58 French reactors and 19 years (1997-2015). As one reactor enters the panel in 1999, the panel contains 1100 observations.

eliminates the incentives for plant managers not to report them.

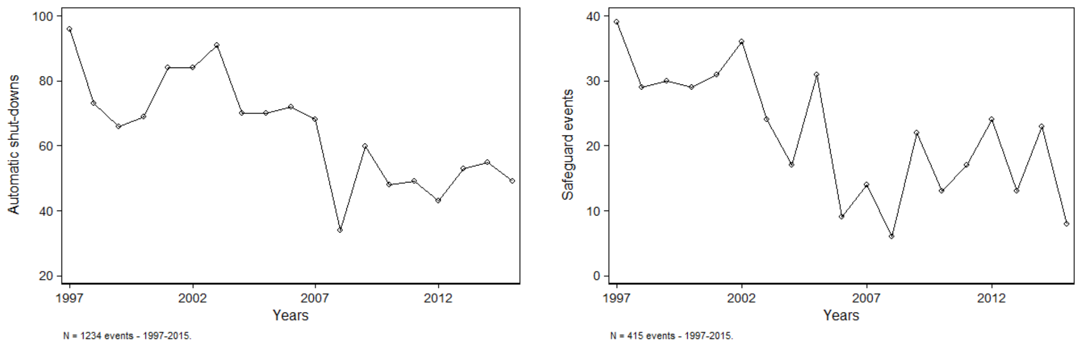
Two categories of safety events satisfy at least one of these conditions. First, automatic reactor shut-downs (or scrams in U.S. terminology, ASD in the following) stop the electrical production of nuclear reactors.[11] These events have also been used by Hausman (2014) as a proxy for nuclear safety. Second, events that require the unplanned use of safeguard systems (safeguard events or simply SFG in the following) are subject to specific auditing efforts during the inspection of the power stations by the regulator. The definition of these events and their relative severity makes them easy and natural targets for the ASN inspectors during the routine inspections of nuclear stations.[12]

Table 2 and figure 3 provide descriptive statistics associated with the annual counts of automatic shut-downs and safeguard events. In particular, figure 3 shows that both automatic shut-downs and safeguard events are characterized, at the fleet level, by decreasing time trends.

[11] 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: http://www.wano.info/en-gb/library/performanceindicators

[12] 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 3: Historical occurrences of automatic shut-downs (left) and safeguard events (right) in the French fleet between 1997 and 2015.



3 Empirical framework

3.1 Treatment effect of interest

Let the random variable AGE_{it} denote the age of nuclear reactor i, $i = 1 \ldots, n$ in a given calendar period t. Furthermore, let Y_{it} be the number of safety related events (incidents) i in period t. We are interested in the *relationship* between AGE_{it} and Y_{it}, which we model as the conditional expectation $\mathbb{E}[Y_{it}|AGE_{it} = a]$. The derivative of this conditional expectation w.r.t. a, $m(a) = \partial\mathbb{E}[Y_{it}|AGE_{it} = a]/\partial a$, describes how the relationship of interest changes when the age changes. We will loosely refer to m as to the "effect" of age on safety. However, m does not describe a true causal effect. The main reason is that the age variable is not manipulatable. To be specific, a major requirement of causal inference is that that an experiment could in principle be constructed, in which the experimenter can control the treatment (i.e. she can assign different values of the treatment to different subjects/units of observation).[13] This manipulability principle does not apply for the age of a reactor: there is no conceivable way to assign different ages to different reactors within a randomize experiment. Thus, age is not a causal factor of safety.

The age of a nuclear reactor is rather a compound approximation for all factors of safety (i) that arise as a consequence of using the facility and (ii) for which it takes time to realize and which are therefore strongly correlated with age. One example is the level

[13]This principle was postulated e.g. in Rubin (1974) and later referred to as "No causation without manipulation" by Holland (1986).

of amortization of the different components (if not exchanged) of a reactor. Were all such factors observable and controlled for, there would be no need to estimate m. However, the majority of these factors are not directly observed and the only way to learn about their impact on safety is indirectly through estimating m. In this sense, we treat m as a quasi-causal object and interpret the estimates with caution.

The interpretation of m as a (quasi-) causal effect of aging on safety is hampered by three major problems. First, there might be a secular time trend in the overall nuclear safety that is unrelated to how much a given facility is used. A main potential driver of such a trend is a change in regulatory standards over time. 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 to the safety authority. If such regulatory changes are ignored, different reactors could reveal different observed frequencies of events simply because the measurements were taken at different points in time, leading to a spurious effect of age in the data.

Second, there might be other unobserved factors that are related to the age 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.[14] An example for cohort effects is common exposure to regulatory inspections and norms at the time of the construction. Not controlling for cohort effects potentially distorts the meaning of m as it would capture the effect of characteristics unrelated to the intensity of the usage of the reactors. More generally, the age of a reactor might be correlated with unobserved factors of safety unrelated to the usage of that particular reactor. Examples are improvements in technology, evolution of the management culture as well as accumulation of knowledge *across reactors*.

Third, the statistical analysis of safety is potentially hampered by measurement errors due to missing observations. There are two channels through which missing observations

[14] See e.g. Glenn (2005) and Suzuki (2012) for an in-depth definition of cohort effects.

might occur: plant managers might fail to detect a safety event (henceforth "detection channel"), or they might not report an observed event to the safety authority ("compliance channel").[15] Detection of events by plant managers might be related to management practices and technology, and incentives not to report events may vary across reactors. Thus, both reasons for missing observations are potentially related to age and technology, and may hence bias the estimates. 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 mainly endemic to the complexity of the nuclear energy industry.

3.2 Empirical strategy

Solving the time trends and cohort effects problems. A standard approach to solve endogeneity problems 1 (secular time trend) and 2 (cohort or individual effects) is to add year dummy variables and individual fixed effects (FE), respectively, and to assume that all endogeneity is captured by these effects. However, in our setup, this would lead to perfect multicollinearity because the age variable is a linear combination of year and reactor dummies. 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).

Instead, 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

[15] Several reasons may lead to failures to report a significant event. First, there exist sanctions associated with the reporting of significant safety events. For instance, the ASN requested the shut-down of the Tricastin nuclear station on the 28th of september 2017 due to the report of a significant safety event. Subjective misconceptions of the reporting criteria are a second rationale for the non-reporting of an event. For instance, the following administrative subtlety of the reporting process is consistent with the existence of reporting failures. Managers actually report two types of safety events to the ASN: *significant* safety events and *interesting* safety events. The former constitute our source of data, and correspond to the reporting criteria presented above. The latter are much more numerous, and much less significant. Missing observation could occur if a *significant* safety event was reported (intentionally or not) as only *interesting* for safety. Such misreports can occur unintentionally, for instance if a plant manager underestimates the potential consequences of a particular situation.

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, 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 variation in age as they were 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. This approach 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.

There are two main pitfalls of this approach. 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 lead to a negative bias in the estimate: older reactors would appear less safe solely due to age. 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). Second, there could reactor-specific time-varying factors which are related to age but not captured by the time dummies. [16] We deal with this possibility in a comprehensive way in section 4 and in

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

appendix D.1.

Solving the imperfect observability problem. One way to deal with the imperfect observability problem is to restrict the sample on the subset of PDD events. For these events, the observability problem does not hold (per definition), which would allow to estimate a relationship of age and safety free of the measurement error related to missing observations.[17] However, this approach comes with a cost. In particular, the perfectly detectable major pitfall of this approach is that the PDD events might not capture all aspects of safety. Alternatively, they might capture different aspects of safety than a general INES 0 or 1 event. Therefore, estimation results using only PDD events have to be interpreted with caution.

4 Empirical investigation

4.1 Age and safety

We estimate the following model of the conditional expectation of Y given AGE:

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

where W_{it} denotes the full set of regressors included in the model and X_{it} denote a set of reactor and year specific control variables, such as whether the reactor is a first-of-a-kind ($FOAK$) or a first-of-a-site ($FOAS$) and an intercept. 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 i belongs to $Site$. These fixed effects capture time-constant, site-specific unobserved sources of heterogeneity.

There are several reasons for choosing model (1). First, including AGE^2 allows for

[17] Similar approaches have been used by Hausman (2014) and Bizet et al. (2022).

nonlinear patterns.[18] A flexible polynomial is often assumed because of its ability to arbitrary well approximate continuous functions (the so-called Weierstrass Theorem) and because of its efficiency compared to fully nonparametric approaches. Second, interacting the age and the site fixed effects allows for heterogeneous "effects" of age. In particular, technical differences in the design of reactors, as well different effectiveness of maintenance and replacement measures might lead to different effects of aging on the safety of nuclear reactors. In addition, as discussed above, climatic and seismic conditions, as well as management organization differ between reactors, and these differences could also interact with the effect of age. As a robustness check, we estimate a homogeneous model of the conditional expectation, and the results remain qualitatively the same, see Section 4.2.[19] Third, the nonlinear expectation accounts for the counting nature of the data and is superior to a linear model. In Section C in the online appendix, we provide an extensive model diagnostics.

The estimation results are displayed in table 3. Columns 3 and 4 contain the results obtained on the subsamples of the two types of PDD events (ASD and SFG , respectively), while column 5 presents estimates on the full sample. These three specifications are estimated with a Negative Binomial estimator using unconditional fixed-effects (see appendix C and Allison and Waterman (2002) for a comprehensive discussion). For a reference, we also present results obtained under a linear fixed effects specification on the subsample of ASD events.

Our estimates reveal several regularities. First, the majority of the coefficients of an interacted AGE variable are negative and significant in all four specification. In addition, the coefficients of an interacted AGE^2 variable are positive and significant, and are of smaller magnitude than the corresponding coefficients of the interacted AGE variables. Since the estimates in all three columns are of comparable sign and significance, we focus in

[18]In an alternative specification, we also included higher polynomial terms, which were insignificant. Results are available upon request.

[19]Including AGE as a separate additive variable would lead to an equivalent specification. The coefficient of AGE in this case would simply measure the effect of the non-included dummy variable (the reference category).

Table 3: Main regression results: non-linear, heterogeneous treatment effects

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ALL
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

ASN data. This table presents the result of four estimations of the effect of age on safety, and using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In all specifications, the age and squared-age variables have been interacted with site dummies to allow for non-linear and heterogeneous treatment effects. All specifications include site fixed effects and year dummies. Column (OLS) presents the result of an OLS estimation using automatic shut-downs (ASDs) as a dependent variables. Columns (1) to (3) present the results of a NB regression with quadratic over-dispersion. Column (1) uses counts of ASDs as the dependent variable. Column (2) uses counts of safeguard events as the dependent variable, to test whether the trends observed on ASDs can be generalized to other PDD events. Column (3) uses the counts of significant safety events as the dependent variable, to test whether the trends observed on PDD events can be generalized to non-PDD events. Standard-errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

our discussion on the OLS results, whose interpretation is easiest. Consider the Chooz site as a representative example. The coefficient of AGE is -0.29, which implies a reduction of -0.29 in the total number of events for each additional year. This effect is offset for high values of AGE by the effect of AGE^2, whose estimated effect is 0.014. Figure 4 shows the estimated pattern of the causal effect for that reactor. On the x-axis, the AGE is measured with values between 0 and 20. This span approximately equals the time span of observation of the reactors of this station in our dataset. On the y axis, we measure safety as captured by Y. The plotted graph is generated by the function $\hat{\alpha}_1 AGE + \hat{\alpha}_2 AGE^2$ with $(\hat{\alpha}_1, \hat{\alpha}_2)$ here equal to $(-0.29, 0.014)$.[20] The number of accidents goes at first down with increasing time, driven by the coefficient of AGE. Roughly at the age of 7, the relationship becomes reversed and safety worsens with passing time. The initial level of safety is achieved at approximately age of 15 years. This pattern is compatible with the so-called "bathtub effect". The engineering reliability literature describes the bathtub effect as a three-state process, see e.g. (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 relatively constant. In the final state, reliability decreases as the system wears out.[21] As our results show, the reactors with significant AGE variable all exhibit this non-linear pattern.

Second, for the SFG (ASD) subsample, there are 3 (6) sites whose AGE effect estimates are positive, two of them being significant. Analogous observation with a reversed sign can be made for the AGE^2 variable.[22] These estimates imply that if homogeneity of the age effect is assumed (i.e. the age variable is not interacted with site effects), there would be a loss of information which might bias the results. We study this possibility below.

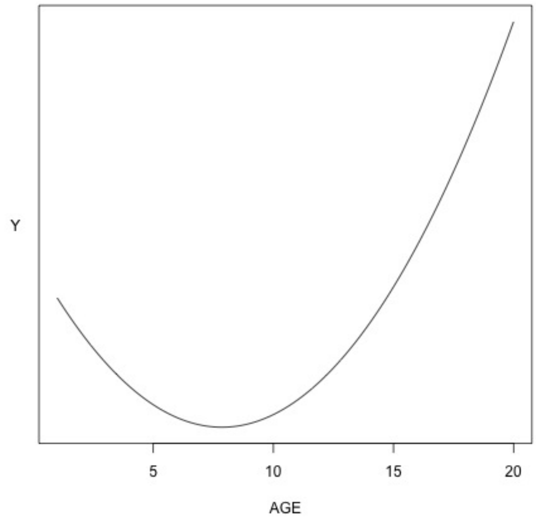
Third, the estimates of $FOAS$ and $FOAK$ are not significant. This is somewhat surprising for the following reason. The literature on the costs of construction of the

[20]The intercept and hence the values of the dependent variable Y are suppressed.

[21]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).

[22]In the Penly power station, occurrences of safeguard events exhibit a *reverse bathtub* trend. Safeguard events 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.

Figure 4: Age and safety: a bathtub effect



French nuclear reactors provides evidence that FOAS/FOAK reactors in general cost more. This suggests the presence of learning-by-doing effects in terms of lead-time and construction costs within sites and technological designs (see e.g. Berthélemy and Rangel (2015); Rangel and Lévêque (2015)). This is the main motivation for including these two variables into our regressions: if these learning-by-doing effects have an effect on the safety of the reactors, then omitting the FOAS/FOAK variables would violate our site cohorts assumption. One possible explanation for the insignificant estimates is that 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 have leveled out safety across reactors within sites. To verify this conjecture, we run a regression equivalent to the main specification but without site FE. The results are displayed in table 7 in appendix E. The estimate of the coefficient of $FOAS$ in this regression becomes negative and statistically significant, thus supporting our conjecture. Moreover, this finding provides an additional evidence that all unobserved reactor-specific endogeneity is at the site level and therefore captured by site FE.

Fourth, the estimates on the full sample and on the PDD samples are qualitatively similar. As an example, these two regressions yield the same sign in 13 out of 19 estimates for the AGE variable. However, the estimates on the full sample are almost always smaller in magnitude and in roughly half of the cases insignificant. There are two possible (non-competing) explanations for this finding. First, it is possible that the aging has different effects for the PDD and all events. The second one is related to transparency problems described in the previous section. In particular, it is possible that for the full sample, actual changes in safety over time are less frequently detected or reported. Without further data, it is impossible to distinguish between these two explanations and the true explanation might be a combination of them.

4.2 Challenging the assumptions

Challenging the fixed-effects assumption. We now challenge the assumption that it is sufficient to define a cohort (fixed effects) at the site level. Our robustness checks address two related problems.

First, as discussed above, including reactor-specific fixed effects is not possible due to perfect multicollinearity. We therefore challenge our assumption in an alternative way: instead of using more coarse definition of cohorts (as in the main results), we aggregate on the time scale. In particular, instead of adding time dummies for each year to the regression, we add 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. This strategy allows us to include reactor-specific FE. The results are shown in table 8 in appendix E. They are in line with our main results.

Second, even if the unobserved heterogeneity is at the site level, it is possible that it varies over time. In this situation, site fixed effects and year dummies would not be sufficient to capture this unobserved heterogeneity, and the estimated coefficient of the age variable could be biased. To challenge the time-constant nature of unobserved heterogeneity, we compare the results of the linear within estimation presented in table 3 with the results of a first difference (FD) estimation, formally described in appendix D.1. As the within and FD estimators are asymptotically equivalent under the fixed-effects assumption, one should not expect major differences in their results. On the contrary, if the fixed-effects assumption is violated, omitting a time-varying first differenced variable in the FD approach could provide a different estimate than when omitting an averaged variable as in the within approach.[23] Results of the FD estimation are presented in table 9 in appendix E, and are also in line with our main results.

To round up our discussion on the age-cohorts problem, it would be useful to see in which direction the APC bias would be in our case. This would help learn more about the mechanics of the separate components. To do so, we run a regression with capacity FE instead of site FE. This exercise is closely related to our study of heterogeneity above. The results are shown in table 7 in appendix E. In column (2), age is shown to have a significant linear effect on occurrences of safeguard events in all capacity cohorts, contrarily to the results of the estimation presented in column (1). These results suggest that the frequency of safeguard events increases when these reactors get older. In addition, nearly all year fixed effects become statistically significant, which suggests an increase in safety over time, whereas year fixed effects were not significant when site fixed effects were included. Similar variations are observed when comparing the estimations presented in columns (3) and (4), when allowing for non-linear treatment effects. This is an illustration of the APC problem: 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.

Imposing effect homogeneity. We first study whether allowing heterogeneity of the age effect is necessary in the first place. To do so, we modify model (1) in which the AGE, AGE^2 are not interacted with site fixed effects. The results can be found in table 6 in appendix D.2. The estimates are qualitatively largely the same with negative signs for the AGE variable and positive for the AGE^2.

[23] A detailed description of this novel approach is provided in appendix D.1.

5 Concluding remarks

In this paper, we estimate the relationship between age and safety of a nuclear reactor. We use a novel dataset that contains more than 13000 safety incidents from the French nuclear power fleet. We find that the relationship follows a so-called bathtub effect: safety initially increases, perhaps due to learning effects, and eventually decreases, possibly due to amortization.

Since age is not a true cause of safety, but merely a compound approximation of all usage-related factors of safety that are correlated with age, these results must be interpret with caution. A promising path forward for future research is therefore opening the black box and trying to measure directly these usage-related factors.

A Sample selection

Not all reported significant safety events were used for our analysis. First, generic events characterizing the whole fleet or particular cohorts of reactors are excluded. These generic events represent the detection 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. 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, although our dataset initially included events reported between 1973 and 1996, we excluded this period of observation 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 (and numerical storage) process. Conversely, when aggregating safety events into counts of reports per reactor and per year, the events that affected some systems common to multiple reactors within a given power station were counted in each affected reactor.

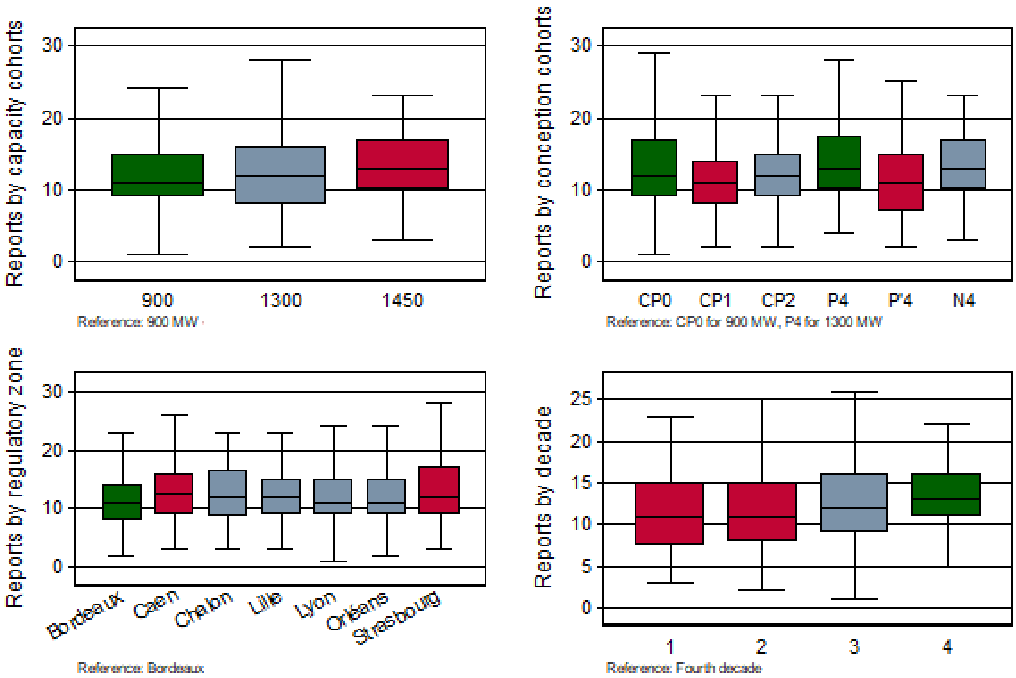
B Additional descriptive evidence

In this appendix, we compare the frequencies of reports of safety significant events (without any restriction to the PDD subsample, and using all available data regarding events reported between 1973 and 1997) across groups of reactors. [24] Figure 5 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, 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

[24] 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.

pairwise comparisons[25].

Figure 5: Annual declarations by reactor for different groups of reactors



For the upper two sub-figures in figure 5, 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.[26] 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 5 is dedicated to France's regional regulatory subdi-

[25] The significance of the difference in means among the pairs identified in figure 5 is robust to several methods of adjustment for multiple comparisons (i.e. Tukey, Sidak and Bonferonni's methods).

[26] 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.

visions.[27] 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.[28] 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.

C Statistical model specification and model diagnostics

Our model selection is based on standard arguments, well established in the empirical and theoretical literature. First, 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.

Second, concerning the estimation procedure, we estimate this model using a Negative

[27]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.

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

Binomial estimator with quadratic over-dispersion and site-clustered standard errors. We perform model selection using the Akaike and Bayesian information criterion, whose values are reported in table 4. Calculations of the Pearson statistics for the Poisson regressions (on the ASD and SFG subsamples) 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.

Table 4: Test-statistics for regression model selection

Regression model	Over-dispersion	S-E clusters	Ln L	Pearson	AIC	BIC
Poisson	-	Reactor	-1473	1198	2956	2981
Poisson	-	Site	-1473	1198	2982	3072
Neg. Bin.	NB1	Reactor	-1471	-	2952.2	2977.2
Neg. Bin.	NB1	Site	-1471	-	2978.1	3068.1
Neg. Bin.	NB2	Reactor	-1471	-	2952.1	2977.1
Neg. Bin.	NB2	Site	-1471	-	2978.2	3068.3

ASN data. This table presents test statistics associated with different count estimators using the model specification presented in equation (5) and automatic shut-downs as a dependent variable (i.e. regression (1) in table 4). Two regression models, e.g. the Poisson and Negative Binomial model are compared using the Pearson over-dispersion test statistics. For the Negative Binomial model, we also compare linear and quadratic over-dispersion specifications. For all models, we compare standard-error clustering at the reactor or site level. The log-likelihood, Akaike and Bayesian Information Criteria are computed for each regression models.

As a result, the Negative Binomial estimator with quadratic over-dispersion and site-clustered standard errors dominates the Poisson-QMLE estimator and the Negative-Binomial estimator with linear over-dispersion. One advantage of the Negative Binomial estimator is that it better fits over-dispersed data, consistently with the descriptive evidence provided in table 2 on page 14. 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.[29]

The interpretation of the coefficients obtained through a Negative-Binomial regression

[29] 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. Results are in line with the main results and available upon request.

model can be done 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. 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. As an example, using the notations defined in equation (5), $e^{\beta_{Age}}$ represents the (multiplicative) average effect of a unit increase in the age of a reactor on the expected number of occurrences of events.

D Robustness checks

D.1 Is unobserved heterogeneity site-specific and time-constant? Some robustness checks

Assumption A1 consists of two components: (a) that the unobserved heterogeneity is at the site level, and (b) that this unobserved heterogeneity is time constant. To fix ideas, suppose that the unobserved factors r_{it} can be decomposed as $r_{it} = q_{it} + \varepsilon_{it}$, where q_{it} is potentially correlated with age and ε_{it} is the independent component of the error term. Assumption A1 postulates that $q_{it} = q_j$, where j indexes the different sites. Part (a) of the assumption requires that for any pair of reactors (i, i') in site j, it holds $q_{it} = q_{i't} = q_{jt}$. Part (b) of the assumption amounts to omitting the dependence of q_{it} on t.

In the absence of a quasi-experimental setting, we first provide indirect evidence for part (a). In principle, allowing for reactor-specific fixed effects would be desirable since there could exist some unobserved heterogeneity at the reactor level. However, allowing for reactor-specific FE is not possible due to perfect multicollinearity of age, time dummies, and reactor FE. We tackle this problem in an alternative way: instead of using more coarse definition of cohorts (as in the main results), we now aggregate on the time scale. In particular, instead of adding time dummies for each year to the regression, we add 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. This strategy allows us to include reactor-specific FE. By doing so, and if there exist unobserved heterogeneity correlated with age at the reactor level, then the results of our main results should be biased and thus differ from those obtained in this aggregated time-span specification. Nevertheless, the results are shown in table 7 in appendix F, and are in line with our main results.

We now target part (b). While this assumption is not testable, indirect evidence can be provided. To do so, we use a novel approach that compares the estimates produced with a standard linear within and first difference (FD) approaches. Under assumption A1, the within and FD estimators are asymptotically equivalent. Therefore, although in finite samples the estimates produced with these two approaches under A1 are likely to differ[30], one should not expect substantial qualitative differences in the implications. On the other hand, if A1 is violated, then the two estimators might produce very different results. In particular, omitting a time varying first differenced variable q_{it} could lead to a different bias then when the omitted variable is an averaged q_{it} as in the within approach. A very large difference in the conclusions produced with these two approaches would therefore point at violation of A1.

We therefore reestimate the main linear specification (whose results are presented in table 4 on page 17) with an FD estimator. The estimates are shown in table 8 on page 33. The results in 14 out of 19 nuclear plants have the same sign as in the main regression, and in 12 out of those 14 they have also they same statistical significance indication. The bathtub is found consistently with the main results. In addition, the magnitudes of the coefficients remain very close to the initial ones. Thus, both qualitative and quantitative implications derived on basis of the two methods are comparable. As a representative example, we again look at the Chooz site. The estimated age effect for this reactor exhibits a bathtub pattern as in the initial result, with a maximum safety achieved after approximately 12 years (compared to the 7 years with the within estimation).

[30] This is because the FD uses less observations: the first period of observation cannot be lagged and is therefore dropped; in addition, there is a pure mechanical difference in how the values are calculated, which leads to deviations in small samples and vanishes asymptotically).

Remark. Note that this exercise cannot be taken as a formal test for time varying unobserved heterogeneity. In particular, there could be cases of time patterns of q_{it} for which within and FD estimators remain very similar. Put differently, it is not clear against which alternatives there will be statistical power. Nevertheless, the results presented above add additional evidence for the plausibility of assumption A1.

D.2 Imposing homogeneity

To learn about the sensitivity of the heterogeneity assumption, we now impose the following exponential specification of the conditional mean:

$$\mathbb{E}(Y_{it}|W_{it}) = \exp\left(\alpha_1 AGE_{it} + \alpha_2 AGE_{it}^2 + X_{it}\beta + \sum_{Year} \beta_{Year} \cdot \mathbb{1}_{Year} + \sum_{Site} \beta_{Site} \cdot \mathbb{1}_{Site}\right). \tag{2}$$

Note that the main independent variables, AGE and AGE^2, appear in equation (2) as an additive argument of the exponential function. As a result, their (multiplicative) effects on safety are not allowed to depend on the specific values of the remaining variables (homogeneous treatment effect).

The results are displayed in table 5 in section E below. Column 2 contains results obtained under a linear FE specification on the subsample of ASD events (i.e. Y is defined as the count of ASD events), while columns 3 and 4 contain the results obtained using the exponential specification presented in equation (2) on the subsamples of ASD and SFG events, respectively. The results in columns 3 and 4 are produced with a Negative Binomial estimator using unconditional fixed-effects (see appendix D and Allison and Waterman (2002) for a comprehensive discussion). In all three specifications, the estimates of the effect of AGE are negative and the estimates of the effect of AGE^2 are positive. This pattern is consistent with a bathtub effect, which we discuss in detail below. These estimates, however, are of small magnitude and insignificant in all three specifications. In addition, the order of commissioning of a reactor (FOAK and FOAS) is also shown to have no significant effect on safety.

Table 5: Preliminary regression results: homogeneous non-linear treatment effects

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ALL
AGE	-0.0045	-0.0066	-0.037	0.013
	(0.081)	(0.073)	(0.067)	(0.019)
AGE^2	0.00091	0.0012	0.00019	0.000077
	(0.0012)	(0.00081)	(0.0017)	(0.00038)
First of a Site	-0.059	-0.063	-0.0052	0.033
	(0.12)	(0.11)	(0.13)	(0.040)
First of a Kind	0.086	0.068	-0.068	-0.080**
	(0.14)	(0.12)	(0.27)	(0.034)
Year FE	Y	Y	Y	Y
Site FE	Y	Y	Y	Y

ASN data. This table presents the result of four estimations of the effect of age on safety, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In all specifications, the age and squared-age variables are included, allowing for non-linear but homogeneous treatment effects. All specifications include site fixed effects and year dummies. Column (OLS) presents the result of an OLS estimation using automatic shut-downs (ASDs) as a dependent variables. Columns (1) to (3) present the results of a NB regression with quadratic over-dispersion. Column (1) uses counts of ASDs as the dependent variable. Column (2) uses counts of safeguard events as the dependent variable, to test whether the trends observed on ASDs can be generalized to other PDD events. Column (3) uses the counts of significant safety events as the dependent variable, to test whether the trends observed on PDD events can be generalized to non-PDD events. We provide site-clustered standard errors in parentheses. Intercepts are omitted. Significance: ***1%; **5%; *10%.

These results are largely consistent in terms of sign with our main results.

E Additional tables

Table 6: Additional regresssion results: analysis of the ACP bias

VARIABLES	(1) SFG	(2) SFG	(3) SFG	(4) SFG
900×AGE	-0.041	0.12***	-0.17	-0.066
1300×AGE	-0.022	0.12***	-0.068	0.049
1450×AGE	-0.045	0.11*	-0.024	0.11
900×AGE^2			0.0029	0.0040
1300×AGE^2			0.0017	0.0023
1450×AGE^2			-0.00036	0.00043
FOAS	-0.0026	-0.23**	-0.0023	-0.22**
FOAK	-0.074	0.18	-0.089	0.13
Capacity = 1300MW		1.25**		-0.22
Capacity = 1450MW		2.51***		0.46
Blayais	-0.40		0.68	
Bugey	0.54		1.59	
Cattenom	-0.69***		-0.69***	
Chinon B	-0.48		0.55	
Chooz B	-0.043		-0.44	
Civaux	-0.60		-0.99	
Cruas	-1.00		0.059	
Dampierre	-0.33		0.74	
Fessenheim	0.86		1.86	
Flamanville	0.18		0.15	
Golfech	-0.98***		-0.98***	
Gravelines	-1.16		-0.098	
Nogent	-0.045		-0.044	
Paluel	0.16		0.12	
Penly	-0.39***		-0.37***	
Saint-Alban	-0.15		-0.18	
Saint-Laurent	-0.025		1.05	
Tricastin	-0.42		0.65	
1998	-0.25	-0.42	-0.23	-0.39
1999	-0.20	-0.51	-0.17	-0.44
2000	-0.21	-0.68**	-0.16	-0.59**
2001	-0.11	-0.72**	-0.056	-0.62*
2002	0.077	-0.68**	0.13	-0.56
2003	-0.31	-1.23***	-0.26	-1.11**
2004	-0.62	-1.69***	-0.57	-1.57***
2005	0.014	-1.22***	0.044	-1.11**
2006	-1.19**	-2.57***	-1.17**	-2.47***
2007	-0.71	-2.24***	-0.72	-2.15***
2008	-1.53**	-3.21***	-1.56***	-3.16***
2009	-0.20	-2.03***	-0.26	-2.01***
2010	-0.69	-2.67***	-0.78	-2.69***
2011	-0.40	-2.53***	-0.52	-2.59***
2012	-0.0071	-2.30***	-0.17	-2.40***
2013	-0.60	-3.04***	-0.81	-3.20***
2014	0.0050	-2.59***	-0.25	-2.81***
2015	-1.03	-3.76***	-1.34	-4.06***

ASN data. This table presents four estimations, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In specifications (1) and (2), AGE is interacted with capacity group dummies. Specification (1) includes site fixed effects and year dummies. Specification (2) includes capacity group and time dummies, to measure the biases associated with the omission of cohort controls. Specifications (3) and (4) are respectively identical to (1) and (2), but allow for non-linear treatment effects. Standard errors are clustered at the site level. Intercepts

Table 7: Robustness checks: aggregated time periods and reactor fixed-effects

VARIABLES	(1) ASD	(2) SFG
Belleville×AGE	-0.38***	0.0017
Blayais×AGE	0.18**	-0.52***
Bugey×AGE	0.024	-0.095
Cattenom×AGE	-0.00027	0.23***
Chinon×AGE	-0.18***	0.069
Chooz×AGE	-0.25***	-0.21***
Civaux×AGE	-0.14***	0.26***
Cruas×AGE	-0.065	-0.19**
Dampierre×AGE	-0.44***	0.25**
Fessenheim×AGE	-0.63***	0.28*
Flamanville×AGE	0.095	-0.16*
Golfech×AGE	-0.14***	-0.54***
Gravelines×AGE	-0.18***	0.50***
Nogent×AGE	-0.14**	0.045
Paluel×AGE	0.30***	-0.45***
Penly×AGE	0.042	0.62***
St-Alban×AGE	0.13*	-0.55***
St-Laurent×AGE	0.050	-1.01***
Tricastin×AGE	-0.18**	-0.20*
Belleville×AGE^2	0.0081***	0.00040
Blayais×AGE^2	-0.0053***	0.011***
Bugey×AGE^2	-0.0018	-0.00067
Cattenom×AGE^2	-0.00054	-0.0087***
Chinon×AGE^2	0.0033***	-0.0020
Chooz×AGE^2	0.0088***	0.0068***
Civaux×AGE^2	0.00068	-0.014***
Cruas×AGE^2	0.0014	0.0026
Dampierre×AGE^2	0.0077***	-0.0068***
Fessenheim×AGE^2	0.011***	-0.0031
Flamanville×AGE^2	-0.0044***	0.0033
Golfech×AGE^2	0.0043***	0.017***
Gravelines×AGE^2	0.0031***	-0.014***
Nogent×AGE^2	0.0015	-0.000097
Paluel×AGE^2	-0.0091***	0.0095***
Penly×AGE^2	-0.0013	-0.025***
St-Alban×AGE^2	-0.0040**	0.013***
St-Laurent×AGE^2	-0.0018	0.018***
Tricastin×AGE^2	0.0030*	0.0039*
$\mathbb{1}_{1997-2003}$	-0.19	1.00
$\mathbb{1}_{2004-2009}$	-0.019	0.69
$\mathbb{1}_{2010-2015}$	-0.021	0.94*
Reactor FE	Y	Y

ASN data. This table presents the result of two estimations of the effect of age on safety, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In both specifications, non-linear and heterogeneous treatment effects are allowed for. The first specification uses automatic shut-downs (ASDs) as the dependent variable. The second specification uses safeguard events as the dependent variable to test whether results obtained on ASDs can be generalized to other PDD events. Both specifications include aggregated time dummies and reactor fixed effect, to test the possibility of reactor-level unobserved heterogeneity. Standard errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Table 8: Robustness checks: first-difference estimation

VARIABLES	(1) ASD	(2) SFG	(3) ALL
Belleville*AGE	-0.53***	-0.24***	-4.16***
Blayais*AGE	0.79***	-0.080	1.19*
Bugey*AGE	-0.24	-0.65***	-3.43***
Cattenom*AGE	-0.16*	0.084*	-0.10
Chinon B*AGE	-0.16	-0.074	-0.41
Chooz B*AGE	-0.88***	-0.20***	-1.27***
Civaux*AGE	-0.17***	0.065*	1.34***
Cruas*AGE	-0.71***	-0.16**	-5.07***
Dampierre*AGE	-0.67***	0.062	-2.15***
Fessenheim*AGE	-0.66***	-0.23**	3.78***
Flamanville*AGE	0.32**	-0.23***	0.60
Golfech*AGE	-0.26***	-0.46***	-1.55***
Gravelines*AGE	-0.53***	-0.061	-3.24***
Nogent*AGE	-0.68***	-0.20***	-2.59***
Paluel*AGE	0.31**	-0.31***	-1.07*
Penly*AGE	-0.053	0.0077	-1.63***
Saint-Alban*AGE	0.10	-0.82***	-2.19***
Saint-Laurent B*AGE	-0.62***	-1.53***	-4.13***
Tricastin*AGE	-0.46***	-0.065	-2.50***
Belleville*AGE^2	0.012***	0.0057***	0.13***
Blayais*AGE^2	-0.017***	0.0014	-0.020
Bugey*AGE^2	0.0028	0.010***	0.068***
Cattenom*AGE^2	0.0042	-0.0024*	0.010
Chinon B*AGE^2	0.0033	0.0014	0.015
Chooz B*AGE^2	0.031***	0.0062***	0.038**
Civaux*AGE^2	0.0058*	-0.0037**	-0.053***
Cruas*AGE^2	0.016***	0.0033**	0.12***
Dampierre*AGE^2	0.011***	-0.0012	0.043***
Fessenheim*AGE^2	0.013***	0.0055***	-0.059***
Flamanville*AGE^2	-0.0084**	0.0036**	-0.0072
Golfech*AGE^2	0.0069**	0.013***	0.053***
Gravelines*AGE^2	0.0096***	0.0012	0.064***
Nogent*AGE^2	0.016***	0.0062***	0.075***
Paluel*AGE^2	-0.0082**	0.0052***	0.026*
Penly*AGE^2	-0.0010	-0.0012	0.048***
Saint-Alban*AGE^2	0.00025	0.018***	0.063***
Saint-Laurent B*AGE^2	0.0100***	0.027***	0.078***
Tricastin*AGE^2	0.0081**	0.0015	0.055***
Year FE	Y	Y	Y

ASN data. This table presents the result of three first-difference estimations of the effect of age on safety, using a panel of 1042 observed reactor.years (the FD operation requires to drop one observation per reactor) across 58 reactors and 19 years (1997-2015). In all specifications, non-linear and heterogeneous treatment effects are allowed for. The first specification uses automatic shut-downs (ASDs) as the dependent variable. The second specification uses safeguard events as the dependent variable to test whether results obtained on ASDs can be generalized to other PDD events. The third specification uses counts of safety significant events to check whether the results obtained for counts of PDD events can be generalized to non PDD events. All specifications include aggregated time dummies. Standard-errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Table 9: Bathtub effect from the OLS fixed-effect estimation on the ASD subsample

Site	Bathtub	Design	Conception	Size	Subdivision
Bugey	0	900	CP0	4	Lyon
Fessenheim	1	900	CP0	2	Strasbourg
Blayais	0	900	CP1	4	Bordeaux
Gravelines	1	900	CP1	6	Lille
Tricastin	1	900	CP1	4	Lyon
Dampierre	1	900	CP1	4	Orléans
Cruas	0	900	CP2	4	Lyon
Saint-Laurent B	0	900	CP2	2	Orléans
Chinon B	1	900	CP2	4	Orléans
Flamanville	0	1300	P4	2	Caen
Paluel	0	1300	P4	4	Caen
Saint-Alban	0	1300	P4	2	Lyon
Golfech	1	1300	P'4	2	Bordeaux
Penly	0	1300	P'4	2	Caen
Nogent	0	1300	P'4	2	Chalon
Belleville	1	1300	P'4	2	Orléans
Cattenom	0	1300	P'4	4	Strasbourg
Civaux	1	1450	N4	2	Bordeaux
Chooz B	1	1450	N4	2	Chalon

ASN Data. This table summarizes the estimation of the bathtub effect performed in the OLS fixed effect specification using the ASD subsample and presented in table 4. In the Bathtub column, a site is associated with the value 1 if the estimated coefficient of the $AGE \times Site$ variable is negative and significant and if the coefficient of the $AGE^2 \times Site$ variable is positive and significant. Otherwise, the site is associated with the value 0. Information regarding reactor design, conception, size and local regulatory subdivision are added to the table to search for potential patterns explaining the bathtub effect.

References

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

Allison, P. D. and Waterman, R. P. (2002). Fixed-effects negative binomial regression models. *Sociological methodology*, 32(1):247–265.

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.

Bizet, R., Bonev, P., and Lévêque, F. (2022). The effect of local monitoring on nuclear safety and compliance: Evidence from france. *Journal of Environmental Economics and Management*, 113:102651.

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 *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE)*, pages 110–114. IEEE.

Davis, L. W. and Wolfram, C. (2012). Deregulation, consolidation, and efficiency: Evidence from US nuclear power. *American Economic Journal: Applied Economics*, 4(4):194–225.

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.

Gray, W. B. and Shimshack, J. P. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 5(1):3–24.

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.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

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.

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.

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).

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Shimshack, J. P. (2014). The economics of environmental monitoring and enforcement. *Annu. Rev. Resour. Econ.*, 6(1):339–360.

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.

Tollefson, J. (2016). Nuclear power plants prepare for old age. *Nature*, 537(7618):16–17.

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.