
Article:
Invernizzi, DC orcid.org/0000-0001-8178-9557, Locatelli, G orcid.org/0000-0001-9986-2249 and Brookes, NJ (2019) An exploration of the relationship between nuclear decommissioning projects characteristics and cost performance. Progress in Nuclear Energy, 110. pp. 129-141. ISSN 0149-1970

https://doi.org/10.1016/j.pnucene.2018.09.011

Exploration of the Relationship between Nuclear Decommissioning Projects Characteristics and their Cost Performance

Diletta Colette Invernizzi

School of Civil Engineering, University of Leeds,

Woodhouse Lane, Leeds LS2 9JT (UK)

e-mail: cndci@leeds.ac.uk

CORRESPONDING AUTHOR

Dr Giorgio Locatelli

School of Civil Engineering, University of Leeds,

Woodhouse Lane, Leeds LS2 9JT (UK)

e-mail: g.locatelli@leeds.ac.uk

Prof Naomi J Brookes

School of Civil Engineering, University of Leeds,

Woodhouse Lane, Leeds LS2 9JT (UK).

e-mail: N.J.Brookes@leeds.ac.uk

Please cite this work as "*Diletta Colette Invernizzi, Giorgio Locatelli, Naomi J. Brookes, 2019, An exploration of the relationship between nuclear decommissioning projects characteristics and cost performance, Progress in Nuclear Energy, Volume 110, Pages 129-141,"*

Abstract

Nuclear Decommissioning Projects and Programmes (NDPs) are characterized by high complexity and variety, and a time schedule that can take decades. Moreover, NDPs estimates at completion can reach billions of Euro and (for many of these projects) keep increasing, while there is a limited understanding of why this happens. To address this knowledge gap, this paper describes how to statistically test the association between the NDP characteristics and the NDP cost performance. The implementation of statistics on a pool of European NDPs highlights the significance of several country-specific and site-specific characteristics (e.g. respectively, the governance system and the availability of facilities to deal with radioactive material on site). Hence, the original contribution of this paper consists in (i) the selection of statistical tests suitable for analysing small sample sizes (i.e. NDPs) and (ii) the presentation of the results from the implementation of these tests on a pool of 24 European NDPs with an illustrative purpose.

Keywords

Nuclear Decommissioning, Benchmarking, Statistical Analysis, Fisher's Exact Test, Barnard's Test

Highlights

- Nuclear decommissioning projects and programmes often suffer from cost overruns
- There is a limited understanding of what causes these cost overruns
- The association between project characteristics and performance needs investigation
- This paper presents a quantitative approach to statistically test this association
- Some country-specific and site-specific characteristics emerge as relevant results

1 Introduction

Until now, the nuclear sector and its stakeholders (industry, academia, policy maker etc.) have mostly focused on the design and construction of new nuclear infrastructure while, in comparison, the body of knowledge on decommissioning is more limited. Indeed, more than 500 Nuclear Power Plants (NPPs) have been built in the world, but only 16 NPPs have been fully decommissioned (OECD/NEA 2016). However, due to safety, security, economic, environmental, social and ethical reasons, in the near future, more and more nuclear facilities will need to be decommissioned, and a number of new challenges will arise.

Decommissioning encompass all the "*administrative and technical actions taken to allow the removal of some or all the regulatory controls from a facility, except a repository which is closed and not decommissioned*" (IAEA 2017). Nuclear Decommissioning Projects and Programmes (NDPs) are therefore here intended as site-level projects and programmes undertaken to restore the site to new use.

NDPs are characterized by extremely diverse inventories of radiological material, whose handling increases the project complexity and uncertainties. NDPs range from smaller projects like the decommissioning of Vandellós-1 NDP (in Spain), whose final costs of the work to reach dormancy in 2003 was €94.6million (IAEA 2011, p.55), to major national multibillion projects, like Sellafield NDP (in the UK). Indeed, Sellafield alone reaches almost £120bn (€136bn), i.e. more than 70% of the decommissioning cost estimates of the whole UK nuclear legacy, which is estimated at £163bn (€185bn) (NDA 2017b). Additionally, year after year, the estimates at completion for some of these NDPs keep increasing (see Table 2 in section 2.2), and there is only a limited understanding of why this happens. Consequently, there is a need to systematically investigate which are the NDP characteristics that mostly impact on the NDP cost performance.

NDP characteristics encompass country-specific characteristics (such as the governance, the funding and the regulatory environment, etc.), site-specific characteristics (such as the age and the operational history of the nuclear facility, etc.) and management-related characteristics (such as scope definition and planning of the NDPs). For illustrative reasons, the NDP performance are assessed in this paper in terms of the NDP cost performance, however this approach can be applied to other project performance (such as time, safety, etc.). The NDP characteristics and the NDP performance are described in more detail in section 2.

Until now, only limited research has investigated NDPs from the project management perspective, and the literature still lacks a systematic analysis to assess the association between NDP characteristics and

NDP performance[1]. Therefore, this paper describes a methodology based on benchmarking to analyse NDPs, focusing on the selection and application of suitable statistical tests to address this knowledge gap. Indeed, benchmarking is ideal to compare actual or planned practices in order to identify best practices and generate ideas for improvement (PMBOK 2013), as it is a flexible approach that can address the alleged uniqueness of NDPs. Indeed, every project can be argued to be "unique" (PMBOK 2013). NDPs can be seen as "more unique" than other projects due to their complexity and variety of their design, the legal requirements to decommissioning them, the stakeholders involved, etc. However, lessons from benchmarking NDPs can still be learned, but benchmarking needs to firstly be adapted to the context of NDPs (Invernizzi et al. 2018a).

Indeed, in parallel with the growth of the decommissioning industry, the information available on decommissioning will also increase in the next decades. This information will be both qualitative and quantitative in nature, so there is a need to develop a robust methodology to guarantee a systematic analysis, in which both qualitative and quantitative data are used, and that lessons can be learned and re-applied to seemingly unique projects.

This aim of this paper is to present a systematic approach to test the association between the NDP characteristics and the NDP performance through statistics. Therefore, two statistical tests that are suitable for investigating NDPs (which consists of a small sample size) are selected and applied on 24 European NDPs with an illustrative purpose.

The remaining part of the paper proceeds as follows: section 2 reports the methodology based on benchmarking developed to investigate NDPs, detailing the process of selection of the statistical tests suitable for small sample sizes. Then, these statistical tests are applied on European NDPs; results are presented in section 3 and discussed in section 4; section 5 highlights the limitations of this analysis, while section 6 concludes the paper, paving the way for future research.

[1] If statistical analysis is applied to test the "relationship" between categorical variables, the term "association" is used.

2 Adapting Benchmarking to Nuclear Decommissioning

Invernizzi et al. (2018a; 2017a) presented a selection of benchmarking studies both in the nuclear and non-nuclear sector, highlighting that the meaning of the term "benchmarking" has been widely discussed in the last decades, and that a number of different benchmarking processes are presented in the literature (e.g. see (Anand & Kodali 2008)). Invernizzi et al. (2018a; 2017a) also proposed a methodology based on benchmarking and tailored for NDPs, based on 5 steps:

1. Research initiation, which refers to the gathering of information to understand the context in which the NDP progress;
2. Data collection, which is a systematic recording of information on the NDPs;
3. Operationalisation of the NDP characteristics and the NDP performance (i.e. respectively the independent and dependent variables of this analysis). This consists of creation of a systematic list of the NDP characteristics that impact on the NDP performance, and their codification into non-arbitrary constructs;
4. Implementation, which refers to the actual "problem solving", and it is split into two stages:
   4.1. Cross-comparison of NDPs
   4.2. Statistical analysis implemented on NDPs
5. Validation and dissemination, which provides confirmation of the findings and enables sharing both the methodological and practical learnings, which will be further developed in future work.

Step 4.2, i.e. the statistical analysis, is a fundamental part of this research, as it highlights potential association between the NDP characteristics and the NDP performance. This paper focuses on step 4.2. The choice and implementation of the statistical analysis is grounded on previous research (Locatelli, Mikic, et al. 2017; Locatelli, Invernizzi, et al. 2017; Brookes & Locatelli 2015), which this paper develops both in terms of the selection of the statistical tests and their application on NDPs.

The five steps of the methodology based on benchmarking and described above, the selection of the Barnard's test alongside the Fisher's exact test, and their implementation on NDPs are described in detail in the next sections.

2.1 Research Initiation

The research initiation is the first step to benchmark NDPs, and includes a scrutiny of the information available on NDPs, early scoping interviews with experts and site visits (section 2.1.1), as well as the selection of suitable statistical tests to be implemented (section 2.1.2). This lays the foundation for a sound understanding of the context in which NDPs progress, sets the boundaries of the research and enables a systematic collection of information.

2.1.1 Exploration of the Literature and Collection of Primary Data

The exploration of the literature showed the limited attention posed by academics on the infrastructure end-of-life and management of NDPs. Conversely, publications by international organizations, such as the International Atomic Energy Agency (IAEA/OCED-NEA 2017; IAEA 2011), the OECD/Nuclear Energy Agency (OECD/NEA 2016; OECD/NEA 2015; OECD/NEA 2012) and the European Commission (EU 2015) on this topic have recently flourished. These publications are some of the most relevant sources of information used to understand the NDPs context and collect the NDP characteristics that are recognized to have an impact on the NDP performance. Relevant publications reviewed for this research also include:

- The European Court of Auditors reports (2016; 2011), which discuss the progress of the decommissioning in Lithuania, Bulgaria and Slovakia, stressing (among others NDP characteristics) the consequences of not having a storage facility available;
- the Öko-Institut report (2013), which compares French NDPs by EDF, the Sellafield/NDA case and Greisfwald NDP in Germany;
- The reports by the UK National Audit Office (NAO 2018; NAO 2015; NAO 2012), which describe major projects in Sellafield and the technical and organizational issues that they are facing, as well as contractual challenges concerning the governance of the Magnox NDPs (NAO 2017).
- Laraia's book (2012), which describes several aspects of nuclear decommissioning, ranging from technical to managerial ones, even providing a list of empirical cases;
- The paper by Torp & Klakegg (2016), that explains the challenges in cost estimation under uncertainty in the context of nuclear decommissioning;
- The paper by Invernizzi et al. (2017), where a cross-comparison between two NDPs, i.e. Rocky Flats (US) and Sellafield (UK) was performed;

These publications allowed to build a preliminary list of NDP characteristics that impact on the NDP performance. Nevertheless, none of these publications statistically tests the association between the NDP characteristics and the NDP performance.

To complement the information gathered from the literature, primary data were also collected, and a questionnaire based on the publications listed above and preliminary scoping interviews was prepared. The questionnaire contained one open question (i.e.: "*in your opinion, which NDP characteristics mostly impact on the NDP performance, in terms of cost and time?")* and 29 closed questions. The complete list of NDP characteristics collected both through secondary and primary data are presented in the appendix in Table 4, Table 5 and Table 6, while Table 3 in section 3 summarizes the results.

Table 4, Table 5 and Table 6 respectively list the country-specific, site-specific and management-related NDP characteristics that have been highlighted by the respondent either in the first (and only) open question of the questionnaire (data collection – A), and that have been discussed during the interviews, as included in the closed questions of the questionnaire (data collection – B).

Interviewees were chosen primarily according to their experience of at least one of the NDPs of Table 2, and at least one person with experience of one of the NDPs was interviewed. In total, 35 semi-structured interviews with NDP experts were performed. The interviewees covered the following countries: UK, France, Italy, Spain, Germany, Lithuania, Bulgaria, Slovakia, Sweden, Finland, Switzerland and the Netherlands. More than 80% of the interviewees had more than 10 years' experience in the nuclear decommissioning industry. The collection of primary data was fundamental to make explicit the recent, "tacit knowledge" gained on-field by practitioners.

The list of NDP characteristics was used to describe NDPs systematically. To do this, the NDP characteristics were operationalized into binary, categorical variables. So, for example, for the NDP characteristic "There is an ILW storage available on site", the binary answer Yes/No was used to differentiate NDPs that have a ILW storage available on site, from the ones that did not. Similarly, the NDP performance were operationalized into binary, categorical variables, as explained in section 2.2. First of all, however, the statistical tests suitable to investigate small sample sizes and categorical variables need to be selected. This is described in section 2.1.2.

2.1.2 Selection of Statistical Tests Suitable for Small Sample Size

The selection of statistical tests that are suitable to be implemented on small sample sizes, which is the case of NDPs, is fundamental. The Fisher's exact test is appropriate to test the association between variables in the context of nuclear decommissioning (Invernizzi, et al. 2017a).

The Fisher's exact test uses binary categorical data in the form of contingency tables as input, i.e. tables showing the distribution of one variable in the rows and the other in the columns, as illustrated by a generic contingency table in Table 1. The table reports the number of cases belonging to each of the four cells. The Fisher's exact test is then able to identify whether a single NDP characteristic (i.e. an "independent variable") presents an association (or not) with the NDP performance (i.e. the "dependent variable"), which in this paper consists of (the loosely termed) "NDP cost overruns".

Contingency Table		NDP Performance (i.e. is the project within 10% cost overruns?	
		Yes	No
NDP Characteristic (e.g. is the NDP in the UK?)	Yes	a	c
	No	b	d

Table 1. Example of a 2×2 contingency table

The Fisher's exact test is suitable to be applied to the context of NDPs since (Leach 1979; Freeman & Campbell 2007):

- It investigates the association between variables in the presence of a small data sets (<30 cases), which is the case of NDPs;
- It uses categorical binary data in the form of a contingency table, which is a way to be more objective in the operationalization of independent variables;
- It is a non-parametrical statistical significance test, i.e. it does not require assumptions about distributions (in particular, no normality is assumed);
- It is an exact test, i.e. the probability of an association existing between the variables can be calculated exactly.

Moreover, Kroonenberg & Verbeek (2017) recently quoted the specific recommendation for 2×2 tables by (Cochran 1952, p.334) and (Cochran 1954, p.420), explaining: "*Use Fisher's exact test (i) if the total N of the table < 20, (ii) if 20 < N < 40 and the smallest expectation is less than 5. [..]. If N > 40 use* $\chi 2$*, corrected for continuity"*. Additionally, McDonald (2014, p.77) stated that the "*Fisher's exact test is more accurate than the chi-square test or G–test of independence when the expected numbers are small. I recommend you use Fisher's exact test when the total sample size is less than 1000*". In 1995, Martin (1995) already pointed out how the Fisher's exact test is simple to compute, available in almost all statistical packages, and it is valid from the unconditional point of view (Martin 1995, p.590). These are some of the main reasons why the Fisher's exact test has been traditionally used to test the relationship between two variables when dealing with small sample sizes.

Nonetheless, the Fisher's exact test has also been often criticized for being too conservative (Routledge 1992; Hasselblad & Lokhnygina 2007; Lydersen et al. 2009), and other tests have been suggested by the literature to overcome this drawback. Hasselblad & Lokhnygina (2007) compare five tests for 2×2 tables in clinical trials, among which are the Fisher's exact test and the Barnard's Test. These two tests are both suitable for small sample sizes, and their difference lays on the fact that the Fisher's exact test is a conditional test, while Barnard's test is an unconditional test.

Conditional tests assume that the marginal of the rows and columns (i.e. the row and columns totals) are fixed (or "conditioned"), while in unconditional experiments, none of the row or column totals are pre-specified by the experimenter (Ruxton & Neuhäuser 2010, p.1508). For example, if the researcher decides to explore the potential association between sex of some birds and their willingness to try a novel food type, and he/she selects ten female birds and ten male birds, and introduces them into an experimental arena in which there is a novel food type, and the experiment is stopped after ten birds have consumed the food, both the total numbers of male and females and the total numbers of feeders and non-feeders have been fixed beforehand. So this is a (doubly) conditioned experiment (Ruxton & Neuhäuser 2010, p.1508). Interestingly, a major part of the discussion about 2×2 tables is concerned with which approach, i.e. the conditional or the unconditional one, is the most suitable one (Andres 2006) and there is still great controversy as to whether Fisher's exact test is effective when applied to non-conditional situations (Ruxton & Neuhäuser 2010).

Martín Andrés et al. (2015) explain that conditional exact tests are well known to be more conservative and less powerful than unconditional ones "*because the loss of information as a result of conditioning may be as high as 26% (Zhu & Reid 1994)"* (Martín Andrés et al. 2015, p.1). Andres advocates the Barnard's test as the principal alternative for the Fisher's exact test (Andres 2006, p.4) as well. However, this author also underlines that *"the differences between the two methodologies are greatly diminished for sizes above 50, and even more so in contingency tables of order higher than 2 x 2"* (Andres 2006, p.1). Mehta & Senchaudhuri (2003) compare Fisher and Barnard, explaining more in detail the difference between the two tests.

Traditionally, the Fisher's exact test has been used more often than the Barnard's Test, the latter one being only recently employed in the area of medical research: for example, Shan et al. (2013) and Behrends et al. (2012) presented the results from the implementation of the Barnard's test in the medical field, while Proschan et al. (2016) applied both statistical tests on a research on the Ebola virus.

The recent interest in the Barnard's test is probably due to the fact that, in its earlier development, the Barnard's test was computationally too heavy. Indeed, in 2009, Lydersen et al (2009, p.1159) explained

how "*unconditional tests preserve the significance level and generally are more powerful than Fisher's exact test for moderate to small samples, but previously were disadvantaged by being computationally demanding. This disadvantage is now moot, as software to facilitate unconditional tests has been available for years* […]". These authors also stated that, at the time of writing (i.e. 2008-2009), they were not aware of the Barnard's test being included in any available software (Lydersen et al. 2009, p.1166). Conversely, nowadays Barnard's test can be performed both with R (Tal 2010) and Matlab (Cardillo 2010). For these reasons, the Barnard's test is also deemed to be suitable to be applied in the context of NDPs.

In summary, drawing from previous research and the considerations mentioned above, as well as aiming to provide the reader with the most complete and transparent results, while still being aware of the limitations of the sample size of 24 NDPs and the quality of the input data, results of both the implementation of the Fisher's exact test and the Barnard's test are presented. This choice also derives from:

- the understanding and agreement with McDonald's concerns (2014), who writes "*If your data weren't significant with Fisher's but were significant with your fancy alternative test, they would suspect that you fished around until you found a test that gave you the result you wanted, which would be highly evil. Even though you may have really decided on the obscure test ahead of time, you don't want cynical people to think you're evil, so stick with Fisher's exact test*" (McDonald 2014, p.80);
- having read the work by Camilli (1990), who compares different tests for 2×2 contingency tables, showing that the Barnard's test is theoretically superior, but the Fisher's exact test is still advocated as "*the most rational choice"* and *"the most defensible statistical test available*" (Camilli 1990, p.135); but also
- acknowledging the work of Martin et al. (2004), which concludes that the Fisher's exact test *"can be used as an approximation to Barnard's exact test for a table with 0 or 1 fixed marginals, when the sample size is >= 100 or when the smaller sample size is >= 80, respectively […]*"(Martin et al. 2004, p.745).

Therefore, the output of both the Fisher's exact test and the Barnard's test is a p-value, which represents how likely it is that the results detected by the implementation of this statistical analysis could have resulted from chance rather than due to a actual association between the variables in question.

In this respect, the smaller the p-value, the more significant are the results. Since this paper investigates the NDP cost performance, drawing from (Ruxton & Neuhäuser 2010), the one-sided p-value is reported. Consistently with the literature, the results from the implementation of the Barnard's test show lower p-values than the ones from the implementation of the Fisher's exact test (see section 3). This is thought-

provoking, and can trigger the discussion concerning what is the "right" threshold to define the significance of the result (a plea that has been often made, but only rarely heard). The significance threshold for p-values can vary. In this paper, consistently with (Brookes & Locatelli 2015; locatelli et al. 2017; Locatelli, Mikic, et al. 2017), and following the considerations of Camilli (1990), who envisages to report significance levels only, rather that the dichotomous decision of either "significant" or "non-significant, the authors present the results from the implementation of the Fisher's exact test and the Barnard's test with a p-value lower than 10%. This means that the results regarding the association of variables must be dealt in a circumspect fashion.

2.2 Data Collection based on the NDP Cost Performance

As mentioned in section 1, the units of analysis are European NDPs, intended as site-level projects, i.e. "one nuclear site undergoing decommissioning" is referred to as "one NDP". In the effort of collecting information on the maximum number of European NDPs undergoing decommissioning, publications in English, French, German and Italian were reviewed. The NDPs initially selected after this review are reported in Table 2, which collects and lists the publicly available information on the development of the estimates at completion of NDP. All the cost data refer to estimates at completion (i.e. "*the expected total cost of completing all work expressed as the sum of the actual cost to date and the estimate to complete*", as defined by the Project Management Body of Knowledge (PMBOK 2013, p. 539)), let alone the one referring to Vandellós-1 NDP (in Spain), whose "final cost of work" is reported by the (IAEA 2011, p.55).
During the collection process, it emerged that:

- All the NDPs in Table 2 were nuclear facilities that produced electricity, let alone Harwell NDP, that was nevertheless included in the pool of selected NDPs because (i) of the availability of data and (ii) it is coupled with Winfrith NDPs (which also included one reactor producing electricity), as both Harwell and Winfrith were managed together by Research Site restoration Ltd, and earlier estimates are provided in conjunction;
- In the UK, the Nuclear Decommissioning Authority yearly publishes the cost estimates of the UK nuclear legacy. As mentioned in section 1, Sellafield NDP stands out as a complete outlier, being Sellafield's decommissioning estimates more than 70% than the UK overall. Therefore, Sellafield is not considered in the current analysis (see the last column on the right of Table 2);
- Greifswald is the only German NDPs selected and listed in Table 2, as in Germany the utilities are the reactor owners and there is only scattered information publically available about the development of

decommissioning cost estimates in time. Some updates from the German approach to decommissioning are available on the World Nuclear Association Website (WNA 2018), but cost information are very limited.

- There is no recent public information on the estimate at completion of the Italian NDPs. The latest information regarding the estimate at completion of the four Italian cases come from local news, in the years 2012- 2013. Therefore, because of the unreliability and date of reference of these data, it was deemed inappropriate to include the italian NDPs in the statistical analysis.

The last column on the right of Table 2 highlights which are the NDPs that have been ultimately selected for the current analysis.

In summary, in the effort to generate evidence as well as to guarantee the maximum possible reliability of the results with the limited data and limited quality of the information available, 24 European NDPs have been selected through purposive sampling (Palinkas et al. 2015) for the implementation of the statistical analysis.

United Kingdom	UK NDPs	2006 discounted nuclear provisions [£million] (NDA 2006)	2016 discounted nuclear provisions [£million] (NDA 2016)	2016/17 decommissioning & clean-up costs from the discounted lifetime plan (NDA 2017a)	Is this NDP included in the statistical analysis?
	Berkeley	360	589	1,658	Yes
	Bradwell	506	210	1,736	Yes
	Chapelcross	527	664	2,852	Yes
	Dounreay	2,091	2,394	2,697	Yes
	Dungeness A	504	525	2,035	Yes
	Harwell and Winfrith	1,103	1,174	855	Yes
	Hinkley Point A	543	651	2,102	Yes
	Hunterston A	482	600	2,044	Yes
	Oldbury	444	873	2,072	Yes
	Sellafield	17,831	53,200	119,930	No
	Sizewell A	354	709	1,982	Yes
	Trawsfynydd	413	288	1,859	Yes
	Wylfa	442	728	2,550	Yes
Spain	Vandellós-1	Decommissioning Projects (to reach C&M) completed with 4% cost overruns (IAEA 2011)			Yes
	Jose Cabrera	Progress of the Decommissioning: on time and within the budget (IAEA 2016), while using (ENRESA 2016) cost overruns result >10% but < than 25%.			Yes
France	French NDPs	2001 estimate [€million 2001] (CdC 2012)	2008 estimate [€million 2008] (CdC 2012)	2012 estimate [€million 2013] (CdC 2014)	-
	Chinon A	694,7	810,0	930,3	Yes
	St. Laurent	822,1	803,0	997,6	Yes
	Bugey - 1	348,4	412,0	585,9	Yes
	Brennilis	254,0	373,0	458,6	Yes
	Chooz A	245,1	220,0	344,4	Yes
	Super-Phoenix (Creys-Malville)	941,6	943,0	1311,5	Yes
Germany	German NDP	Greifswald Decommissioning [€million] (EWN 2011) (Wuppertal Institute 2007)		Rückbau-Monitoring 2015 [€million] (Wealer et al. 2015)	-
	Greifswald (Germany)	3,200		4,000	Yes
Bulgaria, Lithuania, Slovakia	Bulgarian, Lithuanian, Slovakian NDPs	Estimate 2010 [€million] (European Court of Auditors 2011; 2016)	Estimate 2011 [€million] (European Court of Auditors 2011; 2016)	Estimate 2015 [€million] (European Court of Auditors 2016)	-
	Kozloduy 1-4 (Bulgaria)	1,118	1,243	1,107	Yes
	Ignalina (Lithuania)	2,019	2,930	3,376	Yes
	Bohunice 1-2 (Slovakia)	950	1,146	1,239	Yes
Italy	Italian NDPs	Estimates "till deactivation" [€million]		Estimate at completion [€million]	-
	Enrico Fermi - Trino	291 (SOGIN 2003)		234 (iBasilicata 2012)	No
	Caorso	568 (SOGIN 2003)		240 (ANSA 2013)	No
	Latina	615 (SOGIN 2003)		704 (LatinaNotizie 2012)	No
	Garigliano	311 (SOGIN 2003)		360 (LatinaNotizie 2012)	No

Table 2. Input costs data in UK, Spain, France, Germany, Bulgaria, Lithuania, Slovakia, and Italy

2.3 Operationalisation of the NDP Characteristics and the NDP Performance

The NDP characteristics collected through the literature review and the semi-structured interviews, as well as the NDP performance in terms of "cost overruns", are categorical variables that need to be operationalized into binary constructs to allow the implementation of both the Fisher's exact test and the Barnard's test. The operationalization of these variables, that consists in coding real data (quantitative, qualitative, complex and uncertain) into "formalised constructs" (as defined by Lee and Lings (2008)) to describe NDPs through a list of binary categorical variable (i.e. boht the NDP caracteristics and the NDP performance), is challenging. In fact, characteristcs such as the location and physical characteristics of the NDP can be operationalized into constructs in a "non-arbitrary way" through concrete objects and attributes (as explained by Rossiter (2002)), while other characteristics, such as "the stakeholders' engagement", consists of a mix of qualitative and quantitative information.

For example, in the attempt to operationalize the stakeholders' engagement, it will result extremely hard to answer the question: "*was the local community surrounding the NDP engaged early and timely*?" with either a "Yes" or a "No". In fact: what does "local community" exactly mean? What is the meaning of "engagement"? What does "early and timely" refer to? To what extent the response of the local community to this "engagement" was actually considered during the development of the NDPs? These are only some of the questions that arose when trying to investigate NDPs social aspects. This is to exemplify the reasons why some NDP characteristics cannot be operationalized. Social aspects, however, have been discussed at a "macro-level" in (Invernizzi et al. 2017b) and at a "micro-level" in (Invernizzi et al. 2018d).

Hence, not all the NDP characteristics that impact on the cost overruns either according to the literature or the interviewees (or both) have been operationalized in a binary way, and are therefore not tested through statistical analysis. The complete list of NDP characteristics that emerged from the literature and/or in the semi-structured interviews is discussed in (Invernizzi et al. 2018b) and presented in Table 4, Table 5 and Table 6, together with comments on their operationalization.

The NDP performance is assessed in terms of cost overruns, which should ideally be rigorously calculated as discussed and described by Invernizzi et al. (2017c). To calculate the cost overruns, drawing from Thompson (2009), the earlier ("initial") estimates are adjusted for the yearly inflation measured by the consumer price index (that can be found in the OECD official website (OECD 2017)). The yearly inflation of Bulgaria comes from the World Bank official website (The World Bank 2018) as it is not available in the OECD official website.

Costs are firstly expressed in costs in 2015 currency, i.e. they are actualized using Eq 1, where C_t is the time when the estimation are defined (see Table 2) and C_{t+1} is the costs actualized using the annual inflation $\boldsymbol{i_{t+1}}$, iterated until all costs refer to 2015. Cost overruns are then calculated as in Eq 2, where $\mathrm{C_{end}}$ refers to the latest estimates and $\mathrm{C_{initial}}$ refers to earlier estimates. Regarding the UK, data from 2006 and 2016 are the ones taken into account as respectively the "initial" NDP estimates at completion ($\mathrm{C_{initial}}$) and the "latest" NDP estimate at completion ($\mathrm{C_{end}}$). UK data from 2016/17 have not been used since the denomination of the costs presented in these reports changed from "nuclear provision" to "decommissioning and clean-up costs", which implies a possible change of scope in the decommissioning activities that would make the data not comparable.

$$\text{Eq 1.}\quad \boldsymbol{C_{t+1} = C_t * (1 + i_{t+1})}$$

$$\text{Eq 2. Cost Overruns [\%]} = \frac{\boldsymbol{C_{end}[currency] - C_{initial}[currency]}}{\boldsymbol{C_{initial}[currency]}}$$

Figure 1 plots the NDP cost overruns against their latest estimates at completion, showing that there is no evident correlation between their estimates and their cost overruns. Cost overruns range from -67% to +60% and estimates range from the €94.6 million of Vandellós-1 (Spain) to the €4billion of Greifswald (Germany). As there is limited agreement on what is the threshold after which a project should be considered affected by cost overruns (e.g. does cost overruns occur when final costs are 2% higher than the initial estimates? Or 10% higher? Or 50% higher?), NDPs are grouped according to their cost overruns, following the literature. More specifically, NDPs are grouped using the following arbitrary threshold, i.e. if there is no cost overrruns, if their cost overrun is within 10%, as in (Brookes & Locatelli 2015), within 25%, as in (Merrow 2011), and within 50%, because Figure 1 shows a considerable gap between Bugey NDP, compared to Brennilis NDP and Iganlina NDP.

After both the NDP characteristics and the NDP performance are coded into binary variables, each NDP characteristic is tested against the NDP performance. For each NDP characteristic, a contingency table (like Table 1 in section 2.1) is built, and both the Fisher's exact test and the Barnard's test are applied. The results from this implementation are in section 3.

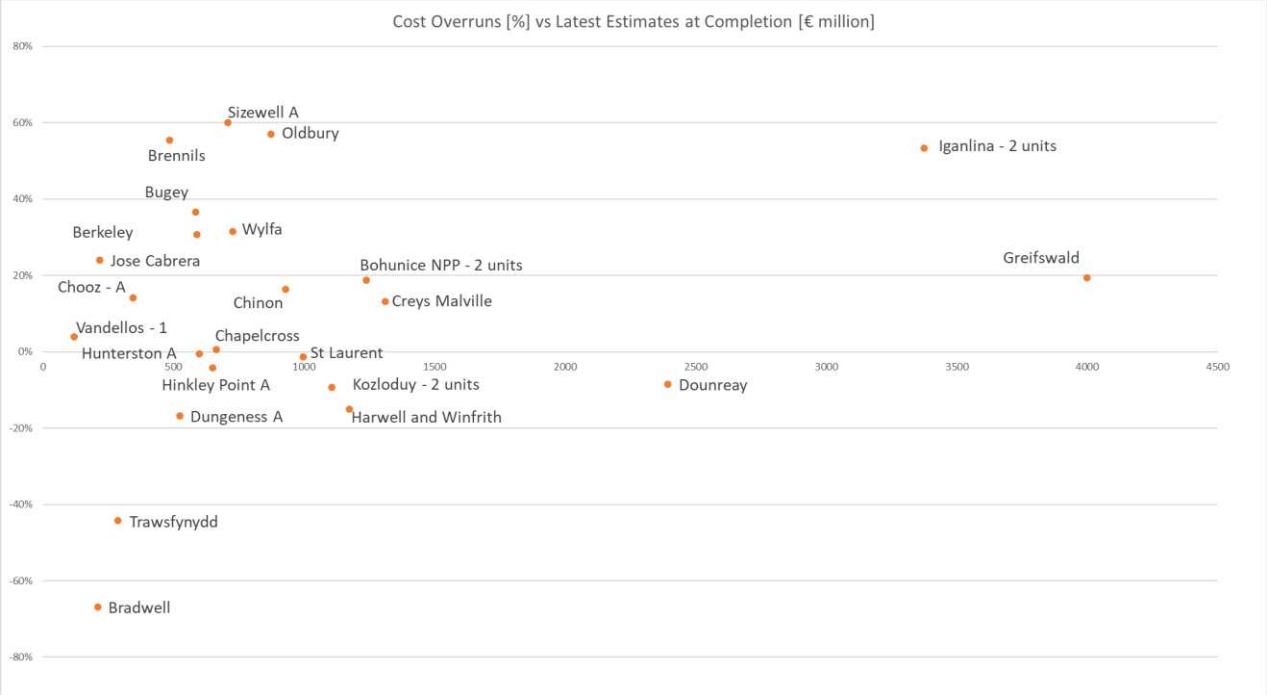

Figure 1. NDPs Cost Overruns [%] vs NDPs Latest Estimates at Completion [€ million]

3 Results of the Statistical Analysis

Table 3 lists the p-values that result from the implementation of the Fisher's exact test and the Barnard's test, and that are lower than 10%. Several considerations can be drawn from these results.

The first consideration is that of the ~80 independent variables (i.e. the NDP characteristics) that have been collected, operationalized, clustered, summarised and tested against the dependent variable "NDP cost overruns" using four different thresholds, only 17 NDP characteristics show a p-value lower than 10% according either to the Fisher's exact test, the Barnard's test or both. This means that for each of these 17 NDP characteristics, it is not possible to hypothesise that there is no association between each single NDP characteristic and the NDP performance assessed in terms of cost overruns. In other terms, the implementation of these statistical tests provides means to highlight the NDP characteristics that present a possible association with the NDP performance.

The second consideration consists in the fact that, as expected and explained in section 2.1, the p-values from the Fisher's exact test are usually higher than the ones calculated using the Barnard's test. This is caused by the fact that Fisher's exact tests is generally more conservative. Therefore, it is possible to underline (once again) that the choice of the statistical tests to implement is fundamental, and has to be clearly and transparently presented. Indeed, as in the example of this paper, some NDP characteristics have a p-value lower than 10% only according to the Barnard's test and would not have emerged if only the Fisher's exact test was implemented. This is for example the case of the variable "stable funding is guaranteed until the end of the NDP" and "there is a 'buffer zone' available on site", which emerge from the Barnard's test. In this situation, probably even more than in others, it is necessary that the knowledge of the researchers comes into play to discuss more in-depth the meaning of the lower p-values, as well as the actual relevance of the operationalized NDP characteristics. Similarly, it is important to underline that the absence of an association does not mean that the corresponding NDP characteristic is not relevant, but simply that this association does not emerge from the implementation of the statistical tests on the specific sample of European NDPs that have been selected, and/or that the information available for the operationalization were not sufficient to highlight an association. In other terms, it is not possible to reach any conclusion on the NDP characteristics that do not have a p-value higher than 10%.

The third consideration derives from the need to stress the importance of the researcher's role in the techno-socio-economic explanation of the actual relevance of the NDP characteristics that emerge from the statistical analysis.

Some NDP characteristics that emerge from the application of the statistical tests, are particularly interesting food for thoughts. For example, the NDP characteristic "there is a complex and multi-layered

governance" could trigger the thought that a complex governance system could support the management of complex projects, e.g. by supporting the NDP progress at different levels (e.g. long-term strategic level vs day-to-day operations) and from different perspectives (e.g. from the financial perspective, to the project-control perspective, etc.). Conversely, it could be conjectured that a complex governance is actually causing cost overruns due to the additional indirect costs. This dilemma can be solved only by going back each single NDPs, and deepening the investigation about governance in each one of them.

Other examples of NDP characteristics that emerged from this analysis and are worth further investigation are surely two site-specific NDP characteristics, i.e. the fact that "the NDP consists of a group of facilities" and that "there is a ILW storage available in the country and on site". Indeed, the first one can trigger the thought that if there is more than one reactor to decommission on site, lessons learned from the first one can be transferred to the second one, and the possibility to re-employ the same team can be a considerable advantage; the second one can suggest to check the interdependencies between waste management operations and decommissioning before undertaking (or progressing with) the NDP.

Similarly, the need to have sufficient space on site (a "buffer zone") for the decommissioning and waste management activities is envisaged to be further analysed, both in light of previous research (Invernizzi et al. 2017), and because it was particularly stressed during the interviews. One interviewee, for instance, explained: "*In Fountain-aux-Rose [NDP, in France], the site is so small! Inside the city! It is a huge struggle for them because they don't have enough room to make a new building, to make an interim storage for waste… so they have to create special access to remove directly waste, as soon as the waste is packed, they send it!*".

Another interviewee also provided an empirical example: "*In Jose Cabrera [NDP, in Spain], cooling towers went down… one each day! They had been built in the 90ies, they were not concrete-based and they have been dismantled one by one, one day after the other. The objective of their demolition was to create more space on site! Only a little amount of the material from the demolition has been re-used. And the scrap metal was sent to a melting facility."*

This suggests the need to plan for decommissioning, even before the start of the construction of a nuclear facility.

NDPs characteristics that show an association with the NDP performance		p-values lower than 10% with Fisher's Exact Test				p-values lower than 10% with Barnard's Test			
		No Cost Overruns	Cost Overruns Within 10%	Cost Overruns Within 25%	Cost Overruns Within 50%	No Cost Overruns	Cost Overruns Within 10%	Cost Overruns Within 25%	Cost Overruns Within 50%
Country-Specific NDP characteristics	The NDP is in the UK	4.47%	4.89%			2.68%	2.47%		
	The NDP is in France						6.87%		
	The country scores a corruption perception index > 80	8.36%				4.95%	5.50%		
	The legal timeframe for review of decommissioning plans is less 2 years	5.23%	8.20%			2.71%	4.14%		
	There is a complex and multi-layered governance	4.81%	8.36%			2.68%	4.28%		
	There is a separate external funding		5.95%			8.02%	3.56%		
	There is a regulated and separate internal fund of the NPP operator, with some protection against insolvency of the operator						6.87%		
	The government funds the whole NDP	8.36%				4.95%	5.50%		
Site-Specific NDP characteristics	The facility started construction before 1960	1.27%	0.61%			0.54%	0.21%		
	Stable funding is guaranteed until the end of the NDP					8.9%			
	The NDP consists of a group of facilities (i.e. more than one reactor on site has to be decommissioned)	3.72%				2.67%	6.87%		
	The facility has a net capacity higher than 200MW but lower than 600 MW		4.96%				2.46%		
	The facility has a net capacity higher than 1000 MW		6.73%			8.0%	2.86%		
	There is a 'buffer zone' available on site								7.56%
	There is a ILW repository available in the country	4.81%	8.36%			2.68%	3.56%		
	There is a ILW storage available in the country and on site	7.13%			8.27%	3.88%	9.79%		8.44%
	There is a HLW storage available on country and on site							6.33%	

Table 3. P-values of the Fisher's exact test and the Barnard's test lower than 10%

4 Discussion

This paper investigates the association between project characteristics and project performance in the nuclear decommissioning industry through statistical analysis. The originality of this research lays on both the methodological approach developed to investigate NDPs (and described in this paper) and on its application on NDPs (which has an illustrative purpose).

In terms of methodological development, compared to previous research, this paper implements the Barnard's test alongside the Fisher's exact test to investigate the NDP characteristics that mostly impact on the NDP performance: as the Barnard's test is less conservative than the Fisher's exact test, the p-values derived with Barnard are lower than the ones derived from the implementation of Fisher. This was anticipated by McDonald's comment, who suggests to use the Fisher's exact test to avoid critics regarding the deliberate choice to use less conservative tests (McDonald 2014). The Barnard's test is an unconditional test that is suitable to test the association of categorical variables as it is more powerful than other tests (Ruxton & Neuhäuser 2010). Therefore, in order to provide the reader with the most complete and transparent set of results, p-values lower than 10% resulting from the implementation of both tests are presented.

Moreover, the NDP performance in terms of cost overruns has been operationalized following the literature using three different thresholds, i.e. 0%, 10%, 25% and 50% of cost overruns, because there is no "universal" agreement in the literature regarding which percentage of over budget can actually be considered a "cost overrun" (e.g. 10% according to (Brookes & Locatelli 2015), 25% according to (Merrow 2011), etc.). Also, different threshold-percentages were considered to suggest that larger thresholds should be used to capture the presence of greater uncertainties regarding the initial and final estimates at completion. Indeed, "*the determination of an accurate cost overrun can only be made by excluding cost increases during project elaboration. These are costs that occur between the initial budget established at definition phase and the final approved budget before work commences. Such costs should be regarded as part of the project initiation process prior to establishing final budget*" (Olaniran et al. 2016, p. 128).

Additionally, the NDP characteristics have been operationalized not only using secondary data, but also using primary data from semi-structured interviews, which can be used to update the information previously collected and to "make explicit" the tacit knowledge of experienced practitioners. Indeed, the information collected through the semi-structured interviews has been firstly analysed through qualitative content analysis (Hsieh & Shannon 2005), used to complement the information from the literature, and then operationalized (when possible) into binary, categorical variables, so that the selected statistical tests can be applied.

Concerning the operationalization of the NDP characteristics, however, it is important to highlight that, even complementing the literature with information from semi-structured interviews, not all the NDP characteristics listed in Table 3 could be operationalized for all the 24 NDPs. More specifically:

- "The legal timeframe for review of decommissioning plans is less than 2 years" has been operationalized for 22 NDPs;
- "There is a ILW storage available in the country and on site" has been operationalized for 21 NDPs;
- "There is a HLW storage available on country and on site" has been operationalized for 17 NDPs.

This is due to either a lack of information in the references used to operationalized the NDP characteristics and/or is caused by the situation where, in the absence of recent and reliable, publically available documents, even the answer of the interviewee(s) was still too vague to guarantee a transparent operationalization of the NDP characteristic for the specific NDP under scrutiny. For example, the NDP characteristics "there is a ILW storage available in the country and on site" received firstly a positive answer that was disproved soon after, when the interviewee specified that the ILW storage was currently under construction, while in this investigation, the focus was put only on already operational facilities (as specified during the interviews).

The possibility to operationalize each NDP characteristics only for some of the 24 NDP obviously affected some NDP characteristics more than others, and further research could tackle this limitation by systematically collecting new, updated information that will be available in the future, as NDPs progress. However, even without having operationalized 24 NDPs, p-values lower than 10% for the abovementioned NDP characteristics were reached.

In terms of practical implication, this paper suggests a way to investigate the project characteristics that impact on the project performance in a systematic way. Therefore, the statistical tests presented in this paper could be re-applied to other contexts (e.g. Oil and Gas decommissioning) and provide new, fascinating insights.

Lastly, as often happens, collecting data through interviews provides a large amount of unexpected and relevant information, which were not possible to operationalize into binary variables. For instance, among others, two of the site-specific NDP characteristics raised the interested of the researchers as particularly stressed by some interviewees. These were: (i) the removal of a layer of contaminated concrete to dispose of the remaining concrete as conventional waste, and (ii) back-filling the voids created on site with non-contaminated material from the demolition were discussed. The activity of removing contaminated concrete to dispose of the remaining concrete as conventional waste received mixed answers, while back-

filling was judged positively by most of the interviewees. Indeed, although the removal of a layer of concrete is envisaged as a way to reduce the waste volume and therefore ultimately reduce the costs of waste storage and disposal, it was judged by more than one interviewee not to be as efficient as expected. Back filling, on the other hand, can both reduce the amount of material to be transported off-site and the amount of material that is needed on-site to fill the voids left after the removal of underground structures, and was therefore seen positively by the interviewees.

Removal of contaminated concrete and back-filling, however, are just two of the characteristics that were particularly stressed during the interviews. Therefore, a follow-up work consists in the systematic analysis of the qualitative information collected during the interviews.

5 Limitations and Scope for Future Developments

Despite marking a major step towards the rigorous investigation of decommissioning projects, this paper has four main limitations. The first one that affects this study is the quality of the cost data. Indeed, cost data have been collected from publicly available sources and often only a limited explanation regarding the assumptions underlying the calculation of these cost data is available: for example, the NDP estimates at completion refer to different stages of the NDP development (an information which is rarely specified in publically-available documents), and there is very limited knowledge about how cost escalation is taken into account and how contingencies have been calculated. Moreover, in this paper, for the calculation of the cost overruns, the consumer price index has been used to account for inflation, even if this index is a "particularly poor choice" (Hollman 2016, p.68), e.g. compared to other indices (such as the chemical engineering plant cost index (Hollman 2016, p.68-69)). Indeed, many indices exist (e.g. commodity indices, labour price and economic indices (Hollmann & Dysert 2007, p.3)). However, due to the unavailability of decommissioning-related indices, the authors preferred to avoid to add further assumptions and complexity on poor-quality input data, and strongly suggest future research on this topic.

Indeed, the aim of this paper is to present an approach to test the association between the NDP characteristics and the NDP performance, and presents its implementation with the available data with a purely illustrative purpose.

Greater understanding of the specific assumptions underlying cost data could be a development of this research, including the specifications of the items that are included in the estimates, boundary conditions and limitations, decommissioning strategy description, end point state, changes in the regulations and technologies (Varley & Rush 2011), through the descritpion of how uncertainties in the cost estimating process are addressed (Torp & Klakegg 2016), how currency (Love et al. 2005) and escalation are taken into account (ackowledging that it is not driven by practices used companies or project managers (Hollmann & Dysert 2007, p.2), and how uncertainties are tackled (IAEA/OCED-NEA 2017).

The second limitation is that this analysis is bounded to European NDPs. This limitation stems from the choice of the authors to limit this research geographically, as both country-specific and site-specific physical characteristics are considerably different in non-European contexts. For example, the regulatory environment and the number of NPPs that have been built in the US are considerably different than the ones in Europe; moreover, the size of the licensed site and the free space available to progress with the decommissioning activities (generally bigger in the US than in Europe) are also dissimilar. Additionally, the US Nuclear Regulatory Commission (NRC) has specific requirements regarding the funding adequacy, i.e.

every US NRC licensed NPP has to estimate decommissioning costs every other year and to submit the estimates to the US NRC to assure that adequate funding provisions are being made into approved decommissioning trust funds. Unregulated NPPs, however, are not required to submit cost estimates publicly (LaGuardia 2016). These differences were also stressed during the interviews. For example, one interviewee from the UK compared the US regulations to the UK ones and emphasized: *"What was found to be hugely different, and the Americans were a bit shocked, was [that] our waste arrangement are entirely different here! We work at Best Practicable Means principle […]. An example of that would be: one of their waste strategies for major reprocessing facilities was that there was lots of concrete, [so] they would fill the hot cells up with concrete, slice them up in one thousand tons or two thousand tons pieces and just place them in the Low Level Waste repositories. And because they added so much concrete to it, everything was Low Level Waste. Now, that's unacceptable in this country! Unless it can be shown that it's Low Level Waste before you add concrete to it, you can't dispose of it that way!*". Future analysis could therefore also consider non-European NDPs and highlight new similarities, differences and potentially new lessons to be learned.

A third limitation of this research is that the results of the statistical tests only provide an "indication" of which NDP characteristics to scrutinize first. This means that no additional conclusion can be derived from the p-values, but conversely the low p-values only play the role of a "sieve" that provides an indication of the first NDP characteristics to look at (to begin with). Indeed, it is important to avoid to be affect by what has been called the "illusion of causality" by (Ahiaga-dagbui et al. 2015, p.866), as finding associations or correlations between factors does not necessarily mean that there is a relationship of causality between them.

The fourth limitation is the fact the statistical tests selected in this paper test the association of each single NDP characteristics against their performance in terms of cost overruns, without considering what a combination of two or more NDP characteristics could show. This latter point could be tackled through Qualitative Comparative Analysis (Schneider & Wagemann 2012) in future research.

Lastly, corrections and/or controlling procedures such as false discovery rate or the familywise error rate could also be considered in further development of the statistics used in this paper. Meanwhile, each, single NDP characteristics that have been collected and listed in this paper can be scrutinized in-depth through single-case study or cross-case study. For example, in light of the results presented in Table 3, a cross-comparison of the different governance systems of the Bulgarian, Lithuanian and Slovakian NDPs (which are managed through the European Bank for Reconstruction and Development, but through different implementing bodies) is envisaged.

6 Conclusions

NDPs are a novel class of projects that has emerged in recent years, issuing new challenges to a number of different stakeholders, including policy makers, project managers, employees on site, and the local community surrounding the NDP. Moreover, the NDP estimates at completion for many of these projects keep increasing, and there is a limited research embracing this area. Benchmarking is a way to tackle these challenges and understand which NDP characteristics mostly impact on the NDP performance, but it needs to be tailored to the case of NDPs. For this reason, a methodology based on benchmarking which includes cross-comparison and statistics has been developed specifically for NDP. This paper focuses on statistics and presents an approach to investigate the association between the NDP characteristics and the NDP performance, through the selection and application of two statistical tests.

The NDP characteristics that have been tested in the paper have been collected from the literature, complemented with empirical information from semi-structured interviews, analysed systematically and operationalized into binary variables (when possible). Then, the Fisher's exact test and the Barnard's test have been applied to test the association between NDP characteristics and NDP performance.

Results highlight the significance of several country-specific and site-specific characteristics (e.g. respectively, the governance system and the availability of facilities to deal with radioactive material on site). However, low p-values from statistical tests can only provide a first indication regarding which NDP characteristics to look at (to begin with), and it is the researcher that plays a pivotal role in discussing and further investigating the NDP characteristics that emerged from the application of statistics.

It is also necessary to iterate that the aim of this paper is neither to discuss the process of estimating costs, nor to propose a new model for costs estimation. This paper examines the relationship between the NDP characteristics and the NDP cost performance, applying two statistical tests suitable for small sample sizes on 24 NDPs with an illustrative purpose. However, the decommissioning industry is a growing industry, and more and more data and information on NDPs will be generated and collected in the very near future and could be fed into this approach, whose results could also ultimately inform project planners and cost estimators. Research on how to improve the NDPs performance has only recently started, and it will be a long journey, which needs to start with a first step. This research represents the first step towards a better selection, planning and delivery of NDPs.

Acknowledgements

This research has been supported by the grant NNL/UA/002. The authors are extremely grateful to all the NDA and NNL experts for all the support received. The opinions in this paper represent only the point of view of the authors, and only the authors are responsible for any omission or mistake. This paper should not be taken to represent in any way the point of view of NDA, NNL or any other organization involved in the decommissioning process of nuclear facilities either in the UK or abroad. The authors are also grateful to Mr John K Hollmann, Dr Tristano Sainati, Mr Dario Domingo, Mr Antonino Arini and Dr Sarah Fores for their useful inputs during the review of the paper.

References

Ahiaga-dagbui, D.D., Smith, S.D. & Love, P.E.D., 2015. Spotlight on construction cost overrun research: superficial, replicative and stagnated. In 31st Annual ARCOM Conference, Raiden A and Aboagye-Nimo E (Eds), Association of Researchers in Construction Management 7-9th Sept, 2015 Lincoln, UK. pp. 863–872.

Anand, G. & Kodali, R., 2008. Benchmarking the benchmarking models. Benchmarking: An International Journal, 15(3), pp.257–291.

Andres, M., 2006. Fisher's exact and Barnard's test - Encyclopedia of Statistical Sciences.

ANSA, 2013. Sogin, al via smantellamento centrale nucleare Caorso. ANSA Official Website, pp.5–7. Available at: http://www.ansa.it/web/notizie/canali/energiaeambiente/nucleare/2013/02/05/Sogin-via-smantellamento-centrale-nucleare-caorso_8194672.html [Accessed July 12, 2017].

Behrends, M., Niemann, C.U. & Larson, M.D., 2012. Infrared pupillometry to detect the light reflex during cardiopulmonary resuscitation : A case series. Resuscitation, 83(10), pp.1223–1228.

Brookes, N.J. & Locatelli, G., 2015. Power plants as megaprojects: Using empirics to shape policy, planning, and construction management. Utilities Policy, (36), pp.57–66.

Camilli, G., 1990. The test of homogeneity for 2 × 2 contingency tables: A review of and some personal opinions on the controversy. Psychological Bulletin, 108(July), pp.135–145.

Cardillo, G., 2010. MyBarnard - File Exchange MathWorks. Mathlab Offcial Website. Available at: https://uk.mathworks.com/matlabcentral/fileexchange/25760-mybarnard [Accessed January 30, 2017].

CdC, 2014. *Le coût de production de l'électricité nucléaire* - actualisation 2014, Paris, 13 rue Cambon 75100. Available at: http://data.over-blog-kiwi.com/0/87/23/97/20150120/ob_55ea8d_cour-des-comptes-cout-production-elect.pdf.

CdC, 2012. The costs of the nuclear power sector - Courtes des Comptes, Paris, 13 rue Cambon 75100. Available at: https://www.ccomptes.fr/sites/default/files/EzPublish/thematic_public_report_costs_nuclear_power_sector_012012.pdf.

Cochran, W.G., 1954. Some Methods for Strengthening the Common χ2 Tests. Biometrics, 10, pp.417–451.

Cochran, W.G., 1952. The χ2 Test of Goodness of Fit. The Annals of Mathematical Statistics, 23, pp.315–345.

ENRESA, 2016. Comparison of estimated and actual decommissioning cost of José Cabrera NPP - International Conference on the Financing of Decommissioning Stockholm, 20-21 September 2016. , p.15.

EU, 2015. Report from the Commission to the European Parliament and the Council - European Commission, Brussels. Available at: http://ec.europa.eu/transparency/regdoc/rep/1/2015/EN/1-2015-78-EN-F1-1.PDF.

European Court of Auditors, 2011. EU financial assistance for the decommissioning of nuclear plants in Bulgaria, Lithuania and Slovakia: achievements and future challenges, Available at: https://www.eca.europa.eu/Lists/ECADocuments/SR11_16/SR11_16_EN.PDF.

European Court of Auditors, 2016. EU nuclear decommissioning assistance programmes in Lithuania, Bulgaria and Slovakia: some progress made since 2011, but critical challenges ahead, Available at: https://www.eca.europa.eu/Lists/ECADocuments/SR16_22/SR_NUCLEAR_DECOMMISSIONING_EN.pdf.

EWN, 2011. The Greifswald Decommissioning Project - Energiewerke Nord GmbH - ppt presentation.

Freeman, J. V. & Campbell, M.J., 2007. The Analysis of Categorical Data: Fisher's Exact Test - tutorial. , pp.11–12. Available at: http://www.sheffield.ac.uk/polopoly_fs/1.43998!/file/tutorial-9-fishers.pdf.

Hasselblad, V. & Lokhnygina, Y., 2007. Tests for 2 x 2 Tables in Clinical Trials. Journal of Modern Applied Statistical Methods, 6(2), pp.456–468.

Hollmann, J.K., 2016. *Project Risk Quantification: Practitioner's Guide to Realistic Cost and Schedule Risk* Management, Probabilistic Publishing.

Hollmann, J.K. & Dysert, L.R., 2007. Escalation Estimation: Working With Economics Consultants. AACE International Transactions, pp.1–6.

Hsieh, H.F. & Shannon, S.E., 2005. Three Approaches to Qualitative Content Analysis. Qualitative Health Research, 15(9), pp.1277–1288.

IAEA/OCED-NEA, 2017. Addressing Uncertainties in Cost Estimates for Decommissioning Nuclear Facilities, Available at: https://www.oecd-nea.org/rwm/pubs/2017/7344-uncertainties-decom-cost.pdf.

IAEA, 2016. Decommissioning and Environmental remediation - IAEA Bulletin (papercopy), Austria.

IAEA, 2017. Glossary. IAEA official website. Available at: https://www.iaea.org/ns/tutorials/regcontrol/intro/glossaryd.htm#D [Accessed December 3, 2017].

IAEA, 2011. Selection and Use of Performance Indicators in Decommissioning, Vienna. Available at: http://www-pub.iaea.org/books/IAEABooks/8566/Selection-and-Use-of-Performance-Indicators-in-Decommissioning.

iBasilicata, 2012. SOGIN : AL VIA ATTIVITÀ SMANTELLAMENTO CENTRALE NUCLEARE - Regione Basilicata. Regione Basilicata Official Website. Available at: http://www.regione.basilicata.it/giunta/site/Giunta/detail.jsp?otype=1012&id=599065 [Accessed July 11, 2017].

Invernizzi, D.C. et al., 2017. Similar but different: A top-down benchmarking approach to investigate nuclear decommissioning projects. In International Conference on Nuclear Engineering, Proceedings, ICONE 25. Shanghai, China.

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2018a. A methodology based on benchmarking to learn across megaprojects: the case of nuclear decommissioning. International Journal of Managing Projects in Business, 11(1), pp.1–18.

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2018b. Characterising decommissioning projects: An exploration of the end-of-life of nuclear infrastructure - Submitted to "Energy Policy" in April 2018. Energy Policy.

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2018. Cost overruns - Helping to define what they really mean. Proceedings of the Institution of Civil Engineers: Civil Engineering, 171(2).

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2017a. How benchmarking can support the selection, planning and delivery of nuclear decommissioning projects. Progress in Nuclear Energy, 99, pp.155–164.

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2017b. Managing social challenges in the nuclear decommissioning industry: A responsible approach towards better performance. International Journal of Project Management, 35(7), pp.1350–1364.

Invernizzi, D.C., Locatelli, G. & Brookes, N.J., 2018c. The need to improve communication about scope changes: frustration as an indicator of operational inefficiencies. Production Planning and Control, (May), pp.1–14.

Irrek, W., 2016. Financing „ Stress Test " Methodology - Hochschule Ruhr West - presentation at the OECD-NEA/SSM International Conference on Financing Decommissioning. , (September), pp.1–25.

Kroonenberg, P.M. & Verbeek, A., 2017. The Tale of Cochran's Rule: My Contingency Table has so Many Expected Values Smaller than 5, What Am I to Do? The American Statistician. Available at: https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1286260.

LaGuardia, T., 2016. Decommissioning Cost Estimate Uncertainty: What is It, How Do You Deal with It? – 16527. In WM2016 Conference, March 6 – 10, 2016, Phoenix, Arizona, USA. pp. 1–9.

Laraia, M., 2012. Nuclear Decommissioning: Planning, Execution and International Experience M. Laraia, ed., Woodhead Publishing Series in Energy.

LatinaNotizie, 2012. Nucleare, la Sogin illustra il piano di decommissioning. www.latinanotizie.it, pp.2011–2013. Available at: http://www.latinanotizie.it/articolo.php?id=26034 [Accessed July 12, 2017].

Leach, C., 1979. Introduction to Statistics: a nonparametric approachfor the social science John Wiley., New York.

Lee, N. & Lings, I., 2008. Doing Business Research: A Guide to Theory and Practice, SAGE Publications.

Locatelli, G., Mariani, G., et al., 2017. Corruption in public projects and megaprojects: There is an elephant in the room! International Journal of Project Management, 35(3), pp.252–268.

Locatelli, G., Mikic, M., et al., 2017. The Successful Delivery of Megaprojects: A Novel Research Method. Project Management Journal, 48(5), pp.1–18.

Locatelli, G., Invernizzi, D.C. & Brookes, N.J., 2017. Project characteristics and performance in Europe: an empirical analysis for large transport infrastructure projects. Transportation Research Part A: Policy and Practice, 98, pp.108–122.

Love, P.E.D., Fong, P.S.W. & Irani, Z., 2005. Management of Knowledge in Project Environments,

Lydersen, S., Fagerland, M.W. & Laake, P., 2009. Recommended tests for association in 2 × 2 tables. Statistics in medicine, 28(January 2009), pp.1159–1175.

Martin, A.A. et al., 2004. Comparing the asymptotic power of exact tests in 2 × 2 tables. Computational Statistics and Data Analysis, 47, pp.745–756.

Martin, A.A., 1995. Is Fisher's exact test very conservative? Computational Statistics and Data Analysis, 19, pp.579–591.

Martín Andrés, A., Herranz Tejedor, I. & Álvarez Hernández, M., 2015. Conditional and Unconditional Tests (and Sample Size) Based on Multiple Comparisons for Stratified 2 × 2 Tables. Computational and mathematical methods in medicine, pp.1–8.

McDonald, J.H., 2014. Handbook of Biological Statistics, Baltimore, Maryland, U.S.A.: SPARKY HOUSE PUBLISHING. Available at: http://www.biostathandbook.com/.

Mehta, C.R. & Senchaudhuri, P., 2003. Conditional versus Unconditional Exact Tests for Comparing Two Binomials. , (September), pp.1–5.

Merrow, E.W., 2011. Industrial Megaprojects: Concepts, Strategies and Practices for Success 1st ed. John Wiley & sons, ed., Cambridge University Press.

NAO, 2012. Managing risk reduction at Sellafield, Available at: https://www.nao.org.uk/wp-content/uploads/2012/11/n1213630.pdf.

NAO, 2015. Progress on the Sellafield site: an update - UK Natinal Audit Office, Available at: https://www.nao.org.uk/wp-content/uploads/2015/03/Progress-on-the-Sellafield-Site-an-update.pdf.

NAO, 2018. The Nuclear Decommissioning Authority: progress with reducing risk at Sellafield Key facts - UK National Audit Office, Available at: https://www.nao.org.uk/report/the-nuclear-decommissioning-authority-progress-with-reducing-risk-at-sellafield/.

NAO, 2017. *The Nuclear Decommissioning Authority's Magnox Contract*, Available at: https://www.nao.org.uk/work-in-progress/the-nuclear-decommissioning-authority/.

NDA, 2006. Annual Report & Account 2005/6 - Nulcear Decommissioning Authority, Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/231625/1416.pdf.

NDA, 2017a. Annual Report & Account 2016/17 - Nuclear Decommissioning Authority, Available at: https://www.gov.uk/government/publications/nuclear-decommissioning-authority-annual-report-and-accounts-2016-to-2017/nda-annual-report-and-accounts-2016-to-2017.

NDA, 2016. Annual Report and Account 2015/2016 - Nuclear Decommissioning Authority, Available at: http://www.nda.gov.uk/documents/upload/Annual-Report-and-Accounts-2010-2011.pdf.

NDA, 2017b. Nuclear Provision: the cost of cleaning up Britain's historic nuclear sites. UK Government official website. Available at: https://www.gov.uk/government/publications/nuclear-provision-explaining-the-cost-of-cleaning-up-britains-nuclear-legacy/nuclear-provision-explaining-the-cost-of-cleaning-up-britains-nuclear-legacy [Accessed August 1, 2017].

OECD/NEA, 2010. Cost Estimation for Decommissioning, Available at: https://www.oecd-nea.org/rwm/reports/2010/nea6831-cost-estimation-decommissioning.pdf.

OECD/NEA, 2016. Costs of Decommissioning Nuclear Power Plants, Available at: http://www.oecd-nea.org/ndd/pubs/2016/7201-costs-decom-npp.pdf.

OECD/NEA, 2012. International Structure for Decommissioning Costing (ISDC) of Nuclear Installations, Available at: http://www.oecd-nea.org/rwm/reports/2012/ISDC-nuclear-installations.pdf.

OECD/NEA, 2015. The Practice of Cost Estimation for Decommissioning of Nuclear Facilities, Paris, France. Available at: https://www.oecd-nea.org/rwm/pubs/2015/7237-practice-cost-estimation.pdf.

OECD, 2017. Inflation measured by consumer price index (CPI). OECD official website. Available at: https://data.oecd.org/price/inflation-cpi.htm [Accessed November 3, 2017].

Öko-Institut, 2013. Nuclear Decommissioning: Management of Costs and Risks - Gerhard Schmidt, Veronika Ustohalova, Anne Minhans, Darmstadt. Available at: http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/490680/IPOL-JOIN_ET(2013)490680_EN.pdf.

Olaniran, O.J. et al., 2016. Cost Overruns in Hydrocarbon Megaprojects: a Critical Review and Implications for Research. Project Management Journal, 46(6), pp.126–138.

Palinkas, L.A. et al., 2015. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm. Policy Ment. Health, 42(5), pp.533–544.

PMBOK, 2013. A Guide to the Project Management Body of Knowledge - Fifth Edition, Project Management Institute.

Proschan, M.A., Lori, D.E. & Price, D., 2016. Statistical Considerations for a Trial of Ebola Virus Disease Therapeutics. Clinical Trials, 13 (1), pp.39–48.

Rossiter, J.R., 2002. The C-OAR-SE procedure for scale development in marketing. International Journal of Research in Marketing, 19(4), p.30.

Routledge, R.D., 1992. Resolving the conflict over Fisher' s exact test. The Canadian Journal of Statistics/La Revue Canadienne de Statistique, 20(2), pp.201–209.

Ruxton, G.D. & Neuhäuser, M., 2010. Good practice in testing for an association in contingency tables. Behavioral Ecology and Sociobiology, 64(9), pp.1505–1513.

Schneider, C.Q. & Wagemann, C., 2012. Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis, Cambridge University Press.

Shan, G. et al., 2013. Randomized Two-Stage Phase II Clinical Trial Designs Based on Barnard's Exact Test. Journal of Biopharmaceutical Statistics, 23:5, pp.1081–1090.

SOGIN, 2003. Bilancio SOGIN - Esercizio 2003. Available at: file:///C:/Users/cndci/Downloads/Bilancio-consuntivo-al-31.12.2003.pdf [Accessed December 3, 2017].

Tal, G., 2010. Barnard's exact test – a powerful alternative for Fisher's exact test (implemented in R). R-statistics blog.

The World Bank, 2018. Inflation, consumer prices (annual %) - Bulgaria. The World Bank official website.

Thompson, G., 2009. Statistical literacy guide: How to adjust for inflation - House of Commons Library, Available at: file:///C:/Users/cndci/Downloads/SN04962.pdf.

Torp, O. & Klakegg, O., 2016. Challenges in cost estimation under uncertainty—A case study of the decommissioning of Barsebäck Nuclear Power Plant. Administrative Sciences, 6(4), p.14.

Varley, G. & Rush, C., 2011. On Decommissioning Cost for Nuclear Power Plants,

Wealer, B. et al., 2015. Stand und Perspektiven des Rückbaus von Kernkraftwerken in Deutschland - Rückbau-

Monitoring 2015 - Deutsches Institut für Wirtschaftsforschung, Available at: https://www.diw.de/documents/publikationen/73/diw_01.c.519393.de/diw_datadoc_2015-081.pdf.

WNA, 2018. Nuclear Power in Germany. World Nuclear Association Official Website. Available at: http://www.world-nuclear.org/information-library/country-profiles/countries-g-n/germany.aspx [Accessed January 30, 2018].

Wuppertal Institute, 2007. Comparison among different decommissioning funds methodologies for nuclear installations - Country Report Germany, Available at: file:///C:/Users/cndci/Downloads/2604_EUDecommFunds_DE.pdf.

Zhu, Y. & Reid, N., 1994. Information, ancillarity, and sufficiency in the presence of nuisance parameters. The Canadian Journal of Statistics, 22(1), pp.111–123.

Appendix

Country - Specific NDP Characteristics		Data Collection		Data Analysis	Comments
		A: Answer of the first, open question	B: Explicit question in the questionnaire	Operationalised for the statistical analysis	
Location and National Strategy	The NDP is in the UK, France, Spain, Germany, Bulgaria, Lithuania, or Slovakia	No	No	Yes	There is enough public information available for the operationalization for the statistical analysis
	The country scores a corruption index lower than 60 and/or lower than 80, as scrutinized by (Locatelli, Mariani, et al. 2017)	No	No	Yes	
	The national strategy is "clearly defined" and/or did not change in the last 10 years	Yes	No	No	Even after the interviews, a univocal, unambiguous operationalization for the statistical analysis is extremely challenging
Regulatory Environment	New "regulations or "changes" in the regulations occurred and affected the NDP (e.g. regulations become "more strict")	Yes	Yes	No	
	The legal timeframe for review of decommissioning plans is less 4 years and/or also less than 2 years (as in (OECD/NEA 2010))	No	No	Yes	There is enough public information available for the operationalization for the statistical analysis
Ownership, Governance & Funding	The Government or the operator has the ownership and responsibility to decommission	Yes	No	Yes	
	There is a complex and multi-layered governance	Yes	No	Yes	
	Funding is allocated yearly or until the end of the NDP	Yes	Yes	Yes	
	There is a separate external funding; there is a regulated and separate internal fund of the NPP operator, with some protection against insolvency of the operator; there are internal restricted funds by the NPP operator governed by the state; there are internal restricted funds by NPP operators (no regulation by the state) – operationalized as in (Irrek 2016)	No	No	Yes	
Supply chain	There is an experienced and reliable supply chain	Yes	No	Yes	

Table 4. Country-Specific NDP characteristics

Site – Specific NDP Characteristics		Data Collection		Data Analysis	Comments
		A: Answer of the first, open question	B: Explicit question in the questionnaire	Operationalised for the statistical analysis	
Design & Construction	The design of the nuclear facility is a Pressurized Water Reactor (PWR), a Boiling Water Reactor (BWR), etc., with a capacity of less than 200 MW, less than 600MW, etc.	No	No	Yes	There is enough public information available for the operationalization for the statistical analysis
	The construction of the nuclear facility started in the 60ies/70ies/80ies/etc.	No	No	Yes	
Site Operations & Waste Management	There are other facilities still operating on site while the NDP takes place	No	Yes	Yes	The interviews provided the information for the operationalization for the statistical analysis
	The NDP collects waste from other sites and/or other countries	No	Yes	Yes	
	Incidents/Accidents occurred during operations or decommissioning in the International Nuclear and Radiological Event Scale (INES)	No	Yes	No	This characteristic emerged from the interviews. However a univocal, unambiguous operationalization was extremely challenging
	(Unexpected) chemical and physical risks are present, e.g. asbestos, sodium, etc.	Yes	No	No	
	There is a “buffer zone” on site, i.e. there is enough space available for the decommissioning activities	No	Yes	Yes	The interviews provided the information for the operationalization for the statistical analysis
	LLW, ILW, HLW storage facilities and/or repositories are available in the country and/or on site	Yes	Yes	Yes	
	Spent fuel is reprocessed in the country and/or on site	No	Yes	Yes	

Table 5. Site-Specific NDP characteristics

Management-Related NDP Characteristics:	Data Collection		Data Analysis	Comments
	A:Answer of the first, open question	B:Explicit question in the questionnaire	Operationalized for the statistical analysis	
The scope of the NDP is "clearly defined"	Yes	Yes	No	Even after the interviews, a univocal, unambiguous operationalization is extremely challenging.
The scope includes buildings remaining and/or includes the reuse of buildings for nuclear and non-nuclear purposes	Yes	Yes	Yes	The interviews provided the information for the operationalization for the statistical analysis
Planning started before the shutdown of the facilities	Yes	Yes	Yes	
Management tools like the Earned Value Management (EVM) are used to measure and report performance	No	Yes	No	Even after the interviews, a univocal, unambiguous operationalization is extremely challenging.
The NDP benefits from a knowledge management system to exchange information on site/in the country/globally and/or international organizations are supporting the NDP with publications and/or consultations and/or financially	Yes	Yes	No	
Incentives are allocated on key milestones, when the actual performance meet and/or exceed the expected performance and/or are allocated to single employees	Yes	Yes	No	Even after the interviews, an univocal, unambiguous operationalization is extremely challenging, and it was not possible to collect enough reliable information on SPE/SPVs
There is an Special Purpose Vehicle/Special Purpose Entity (SPV/SPE) involved in the contracting agreements	No	Yes	No	
Pilot projects and/or mock-ups are used on-site and/or off-site	Yes	Yes	No	Even after the interviews, a univocal, unambiguous operationalization is extremely challenging.
Technologies that are new on site/in the countries have been/are used	No	Yes	No	
Extensive characterization is planned and performed and/or resulted to be accurate	Yes	Yes	No	
Waste routes are "clearly defined" and the interface between the "decommissioning organization" and the "waste management organization" is "well-managed"	Yes	No	No	This characteristic emerged from the interviews. However, a univocal, unambiguous operationalization was extremely challenging.
Activities to reduce waste, such as stripping of concrete/back filling/segmentation in situ/etc. are planned/performed	No	Yes	No	Even after the interviews, a univocal, unambiguous operationalization is extremely challenging.
The NDP social culture needs to change during the transition from operation to decommissioning, as decommissioning is considered to be never ending; External project managers/consultants are employed to foster the "change of culture"; Employees are retrained for subsequent relocation/compensated, e.g. through a severance agreement; The local community is strongly dependent on the activities carried on; The local community has been/is engaged early and timely and no protest arose that caused delays; The authorities and the environmental agencies been engaged early and timely and no delays occurred; etc.	Yes	Yes	No	These characteristics have been discussed in (Invernizzi et al. 2017b)

Table 6. Management-Related NDP characteristics