A Large-Scale Empirical Study of Code Smells In JavaScript Projects

David Johannes, Amir Saboury, Pooya Musavi, Foutse Khomh and Giulio Antoniol
Polytechnique Montreal, Quebec, Canada
{david.johannes, amir.saboury, pooya.musavi, foutse.khomh, giuliano.antoniol}@polymtl.ca

Abstract*—JavaScript is a powerful scripting programming language that has gained a lot of attention this past decade. Initially used exclusively for client-side web development, it has evolved to become one of the most popular programming languages, with developers now using it for both client-side and server-side application development. Similar to applications written in other programming languages, JavaScript applications contain *code smells*, which are *poor* design choices that can negatively impact the quality of an application. In this paper, we extend the work of Amir Saboury et al. [1] by investigating code smells in more JavaScript server-side applications with the aim to understand how they impact the fault-proneness of applications, and how they survive all along the projects. We detect 12 types of code smells in 1807 releases of fifteen popular JavaScript applications (*i.e.,* express, grunt, bower, less.js, request, jquery, vue, ramda, leaflet, hexo, chart, webpack, webtorrent, moment, and riot) and perform survival analysis, comparing the time until a fault occurrence, in files containing code smells and files without code smells. In a different way than our predecessors, we do the survival analysis with a line grain approach (wich means considering the lines where the code smells and the potential bugs appear), and with a line grain approach including dependencies (which means considering the lines where functions, objects, variables are called). Finally, we perform a survival analysis on code smells to know how long they survive. Results show that (1) on average, files without code smells have hazard rates 20% lower than files with code smells in our line grain analysis, and 38% lower in our line grain analysis considering dependencies. (2) Among the studied smells, "Variable Re-assign", "Assignment In Conditional statements", and "Complex Code" smells have the highest fault hazard rates. (3) Code smells, and particularly "Variable Re-assign", tend to be created at the file creation, are not enough removed from the system, and have a high chance of surviving a very long time after their introduction; "Variable Re-assign" is also the most proliferated code smells. Overall, code smells affect negatively the quality of JavaScript applications and developers should consider tracking and removing them early on before the release of applications to the public.

I. Introduction

"Any application that can be written in JavaScript, will eventually be written in JavaScript."
— Jeff Atwood —

JavaScript is a highly dynamic scripting programming language that is becoming one of the most important programming languages in the world. Recent surveys by Stack Overflow [2] show JavaScript topping the rankings of popular programming languages for four years in a row. Many developers and companies are adopting JavaScript related technologies in production and it is the language with the largest number of active repositories and pushes on Github [3]. JavaScript is dynamic, weakly-typed, and has first-class functions. It is a class-free, object-oriented programming language that uses prototypal inheritance instead of classical inheritance. Objects in JavaScript inherits properties from other objects directly and all these inherited properties can be changed at run-time [4]. This trait can make JavaScript programs hard to maintain. Moreover, JavaScript being an interpreted language, developers are not equipped with a compiler that can help them spot erroneous and unoptimized code. As a consequence of all these characteristics, JavaScript applications often contain code smells [5], *i.e.,* poor solutions to recurring design or implementation problems. However, despite the popularity of JavaScript, very few studies have investigated code smells in JavaScript applications, and to the best of our knowledge, there is no work that examines the impact of code smells on the fault-proneness of JavaScript applications. This paper aims to fill this gap in the literature. Specifically, we detect 12 types of code smells in in 1807 releases of fifteen popular JavaScript applications (*i.e.,* express, grunt, bower, less.js, request, jquery, vue, ramda, leaflet, hexo, chart, webpack, webtorrent, moment, and riot) and perform survival analysis, comparing the time until a fault occurrence, in files containing code smells and files without code smells. We do the survival analysis following the work of Amir Saboury et al. [1], but with a line grain approach (wich means considering the lines where the code smells and the potential bugs appear), and with a line grain approach including dependencies (which means considering the lines where functions, objects, variables are called). Finally, we perform a survival analysis on code smells to know how long they survive. We address the following three research questions:

(RQ1) Is the risk of fault higher in files with code smells in comparison with those without code smell? Previous works [6], [7] have found that code smells increase the risk of faults in Java classes. In this research question, we compare the time until a fault occurrence in JavaScript files that contain code smells and files without code smells, computing their respective hazard rates. Results show that on average, across our fifteen studied applications, JavaScript files without code smells have hazard rates 20% lower than files with code smells in our line grain analysis, and 38% lower in our line grain analysis considering dependencies. These results hold those found by our predecessors [1].

(RQ2) Are JavaScript files with code smells equally fault-prone? A major concern of developers interested in improving the design of their application is the prioritization of code and design issues that should be fixed, giving their limited resources. This research question examines faults in files affected by different types of code smells, with the aim to identify code smells that developers should refactor in priority. We do this research through our line grain, and line grain including dependencies analysis. Our findings show that "Variable Re-assign", "Assignment in Conditional Statements", and "Complex Code" smells are consistently associated with high hazard rates across the fifteen studied systems. Developers should consider removing these code smells, in priority since they make the code more prone to faults.

(RQ3) How do the smells survive over time? It is interesting to know how long the smells of a project survive, when they are introduced (at the creation of a file or during a revision), and what type smell are likely to survive the most. Indeed, having a specific knowledge on the smells of a project could help us to determine what smell types are the most dangerous. Results show that smells are created at the file birthdate, and are persistant because a considerable proportion still survive today in the studied systems, and because they have a high chance to survive even a very long time after their introduction into the codebase. Especially, "Variable Re-assign" is the most sizable code smells with one of the highest probability of surviving over time, and thereby we strongly recommend to developers to remove this code smells, at least to reduce their number.

The remainder of this paper is organized as follows. Section II describes the type of code smells we used in our study. Section III describes the design of our case study. Section IV presents and discusses the results of our case study. Section V discusses the limitation of our study. Section VI discusses related works on code smells and JavaScript systems, while Section VII concludes the paper.

II. Background

To study the impact of code smells on the fault-proneness of server-side JavaScript applications, and to study the smells's survival, we first need to identify a list of JavaScript bad practices as our set of code smells. Hence, we select 12 popular code smells from different JavaScript Style Guides [4], [8]–[12], which are the same than those chosen in [1]. Thereby, we will focus on the following code smells:

- **Lengthy Lines** appear when there are too many characters in a single line of code.
- **Chained Methods** appear when there is a "chain" of method calls (repeted chaining method). Chaining method is a common practice in object-oriented programming languages, that consists in using an object returned from one method invocation to make another method invocation.
- **Long Parameter List** happens when a function has too many parameters.
- **Nested Callbacks** are introduced in the code when multiple asynchronous tasks are invoked in sequence (*i.e.,* the result of a previous one is needed to execute the next one) [13], [14].
- **Variable Re-assign** corresponds to the reuse of variables in the same scope for different purposes.
- **Assignment in Conditional Statements** occurs when the = operator is used in conditions. For example: $if(a = b)$
- **Complex Code** smell appears when a Javascript File is characterized by high cyclomatic complexity values (*i.e.,* high numbers of linearly independent paths through the code [15]).
- **Extra Bind** occurs most of time when we let "`.bind(ctx)`" on a function after removing a `this` variable from the body of the inner function, which is an unnecessary overhead.
- **This Assign** happens specifically when a `this` variable is stored in another variable to access to the parent scope's context.
- **Long Methods** is a well-known code smell [4], [16], [17] which consists in writing a method with too many statements.
- **Complex Switch Case** happens when there are too many switch statements.
- **Depth** smell comes when the number of nested blocks of code (or the level of indentation) is too high.

For more clarifications about the twelve studied code smells, please look at the work of Amir Saboury et al. [1], section II.

III. Study Design

The *goal* of our study is to investigate the relation between the localisation of code smells in JavaScript files and lines or line blocks fault-proneness of those JavaScript files, as well as the smells survival all along the projects. The *quality focus* is the source code fault-proneness, which, if high, can have a concrete effect on the cost of maintenance and evolution of the system. The *perspective* is that of researchers, interested in the relation between code smells and the quality of JavaScript systems. The results of this study are also of interest for developers performing maintenance and evolution activities on JavaScript systems since they need to take into account and forecast their effort, and to testers, who need to know which files should be tested in priority. Finally, the results of this study can be of interest to managers and quality assurance teams, who could use code smell detection techniques to assess the fault-proneness of in-house or to-be-acquired systems, to better quantify the cost-of-ownership of these systems. The *context* of this study consists of 12 types of code smells identified in fifteen JavaScript systems. In the following, we introduce our research questions, describe the studied systems, and present our data extraction approach. Furthermore, we describe our model construction and model analysis approaches.

(RQ1) Is the risk of fault higher in files with code smells in comparison with those without code smell? Prior works show that code smells increase the fault-proneness of Java classes [6], [7]. Since JavaScript code smells are different

Table I: Descriptive statistics of the studied systems.

Module	Domain	# Commits	# Contributors	# Github stars	# Releases	# Closed issues	# Forks	Project start date
Express	Web framework	5300+	209	32500+	268	2400+	5900+	Jun 26, 2009
Request	HTTP client utility	2100+	272	16000+	130	1200+	1900+	Jan 23, 2011
Less.js	CSS pre-processor	2600+	209	14500+	49	2100+	3300+	Feb 20, 2010
Bower.io	Package manager	2600+	211	15000+	101	1600+	1900+	Sep 7, 2012
Grunt	Task Runner	1400+	66	11000+	11	1000+	1500+	Sep 21, 2011
Jquery	JavaScript library	6200+	265	45500+	146	1300+	13000+	Apr 3, 2009
Vue.js	JavaScript framework	2100+	122	60500+	207	4800+	8500+	Jul 29, 2013
Ramda	JavaScript library	2400+	160	8500+	45	800+	500+	Jun 21, 2013
Leaflet	JavaScript library	6300+	503	18500+	35	3100+	3200+	Sep 22, 2010
Hexo.io	Blog framework	2300+	100	17000+	119	2100+	2500+	Sep 23, 2012
Chart.js	JavaScript charting	2300+	277	31000+	37	3000+	7900+	Mar 17, 2013
Webpack	JavaScript bundler	4300+	327	30000+	244	3300+	3700+	Mar 10, 2012
Webtorrent.io	Streaming torrent client	2000+	89	13500+	257	700+	1200+	Oct 15, 2013
Moment	JavaScript date manager	3400+	413	32000+	62	2400+	4700+	Mar 1, 2011
Riot	Component-based UI library	3000+	159	12000+	96	1600+	900+	Sep 27, 2013

from the code smells investigated in these previous studies on Java systems, we are interested in examining the impact that JavaScript code smells can have on the fault-proneness of JavaScript applications. The work of our predecessors [1] showed that JavaScript files with code smells are more likely to be fault-proneness than those without code smells. In this research question, we will refute or confirm this conclusion analysing approaches with finer grain, which means a line grain approach and a line grain including dependencies approach.

(RQ2) Are JavaScript files with code smells equally fault-prone? During maintenance and quality assurance activities, developers are interested in identifying parts of the code that should be tested and–or refactored in priority. Hence, we are interested in identifying code smells that have the most negative impact on JavaScript systems, *i.e.,* making JavaScript applications more prone to faults.

(RQ3) How do the smells survive over time? We are interested here in knowing the genealogy of the smells of project, in order to have a better idea of how long those smells survive, if they are persistent, when they are created during the process life of files, and which are the most dangerous.

A. Studied Systems

In order to address our research questions, we perform a case study with the following fifteen open source JavaScript projects. Table I summarizes the characteristics of our subject systems.

Express[1] is a minimalist web framework for Nodejs. It is one of the most popular libraries in NPM [18] and it is used in production by IBM, Uber and many other companies[2]. Its Github repository has over 5,300 commits and more than 200 contributors. It has been forked 5,900 times and starred more than 32,500 times. Express is also one of the most dependent upon libraries on NPM with over 8,800 dependents. There are more than 2,400 closed Github issues on their repository.

Bower.io[3] is a package manager for client-side libraries. It is a command line tool which was originally released as part of Twitter's open source effort[4] in 2012 [19]. Its Github repository has more than 2,600 commits from more than 210 contributors. Bower has been starred over 15,000 times on Github and has over 1,600 closed issues.

LessJs[5] is a CSS[6] pre-processor. It extends CSS and adds dynamic functionalities to it. There are more than 2,600 commits by over 200 contributors on its Github repository. LessJs's repository has more than 2,100 closed issues and it is starred more than 14,500 times and forked over 3,300 times.

Request[7] is a fully-featured library to make HTTP calls. More than 8,300 other libraries are direct dependents of Request. Over 2,100 commits by more than 270 contributors have been made into its Github repository and 16,000+ users starred it. There are more than 1,200 closed issues on its Github repository.

Grunt[8] is one of the most popular JavaScript task runners. More than 1,600 other libraries on NPM are direct dependents of Grunt. Grunt is being used by many companies such as Adobe, Mozilla, Walmart and Microsoft [20]. The Github repository of Grunt is starred by more than 11,000 users. More than 60 contributors made over 1,400 commits into this project. They also managed to have more than 1,000 closed issues on their github repository. We selected these projects because they are among the most popular NPM libraries, in terms of the number of installs. They have a large size and possess a Github repository with issue tracker and wiki. They are also widely used in production.

Jquery[9] is a famous JavaScript library, created to make easier the writing of client-side scripts in the HTML of web pages. It makes also easier the way to write Ajax (asynchronous JavaScript and XML) code. More than 6,200 commits have been made into its Github repository by over 260 contributors, and 45,500+ users starred it. Plus, it is forked more than 13,000 times, and there are more than 1,300 closed issues. Jquery is likely one of the most popular and biggest project of JavaScript ones.

VueJs[10] is a performant and progressive JavaScript framework for building user interfaces. It has the big advantage (in com-

[1] https://github.com/expressjs/express
[2] https://expressjs.com/en/resources/companies-using-express.html
[3] https://github.com/bower/bower
[4] https://engineering.twitter.com/opensource
[5] https://github.com/less/less.js
[6] Cascading Style Sheet
[7] https://github.com/request/request
[8] https://github.com/gruntjs/grunt
[9] https://github.com/jquery/jquery
[10] https://github.com/vuejs/vue

parison with other JavaScript frameworks) to be incrementally adoptable. Over 120 contributors made over 2,100 commits into its Github repository, and they closed more than 4,800 issues. It is forked more than 8,500 times and starred more than 60,500 times, which makes it so popular.

Ramda[11] is a functional library, which makes easier the creation of functional pipelines and functions (as sequences for example), and doesn't mutate user data. It is starred more than 8,500 times, and 160 contributors made over 2,400 commits into its Github repository.

Leaflet[12] is used for mobile-friendly interactive maps, and is designed in order to be simple, efficient, easily extended (with plugins), easy to use, and usable across desktop and mobile platforms. Its Github repository is starred by more than 18,500 users and forked by over 3,200 users. More than 500 people contribute to over 6,300 commits, and managed to have more than 3,100 closed issues on their github repository.

Hexo.io[13] is a very fast, powerful, and simple framework designed for blog's creation. It has 100 contributors, who made more than 2,300 commits, and closed over 2,100 issues. Its Github repository is forked over 2,500 times and starred over 17,000 times.

ChartJs[14] is a flexible and very simple HTML5 charting that offers to designers and developers the chance to see their data in 8 different ways, possibly scalable, customisable and animated. Its Github repository joins over 270 contributors, who closed more than 3,000 issues in over 2,300 commits. Plus, more thant 7,900 users fork it and over 31,000 users star it.

Webpack[15] is a module blunder designed for modern applications. It allows the browser to load only a few number of bundles as small as possible. Those bundles correspond to the packaged modules that the application needs. Webpack is easy to configure and to take in hand. Its Github repository has over 4,300 commits and more than 320 contributors, who closed more than 3,300 issues. It has been forked 3,700 times and starred more than 30,000 times.

Webtorrent.io[16] is a streaming torrent client especially designed for the desktop and the web browser. Almost 90 contributors made over 2,000 commits and help to solve and close more than 700 issues on its Github repository. It is starred over 13,500 times.

Moment[17] allows users to do whatever they want with dates and times in JavaScript (which means manipulate, parse, validate, display, etc.) in a very easy way. Its Github repository has more than 3,400 commits, over 400 contributors, and more than 2,400 closed issues. It is forked over 4,700 times and starred more than 32,000 times.

[11]https://github.com/ramda/ramda
[12]https://github.com/Leaflet/Leaflet
[13]https://github.com/hexojs/hexo
[14]https://github.com/chartjs/Chart.js
[15]https://github.com/webpack/webpack
[16]https://github.com/webtorrent/webtorrent
[17]https://github.com/moment/moment

Riot[18] is a simple, minimalistic, and elegant component-based UI library that offers to users the necessary building blocks for modern client-side applications, some custom tags, and an elegant syntax and API. Almost 160 people contribute to its Github repository, made more than 3,000 commits, and close over 1,600 issues. It is starred more than 12,000 times.

B. Data Extraction

To answer our research questions, we need to mine the repositories of our fifteen selected systems to extract information about the *smelliness* of each file at commit level, identifying whether the file contains a code smell or not. In addition, we need to know for each commit, if the commit introduces a bug, fixes a bug, or introduces a vulnerability, or just modifies the file in a way that a code smell is removed or added. Figure 1 provides an overview of our approach to answer RQ1 and RQ2, and Figure 2 to answer RQ3. We describe each step in our data extraction approach below. We have implemented all the steps of our approach into a framework available on Github[19].

Snapshot Generation: Since all the fifteen studied systems are hosted on Github, at the first step, the framework performs a `git clone` to get a copy of a system's repository locally. It then generates the list of all the commits and uses it to create snapshots of the system that would be used to perform analysis at commits level.

Identification of Fault-Inducing Changes: Our studied systems use Github as their issue tracker and we use Github APIs to get the list of all the resolved issues on the systems. We leverage the SZZ algorithm [21] to detect changes that introduced faults. We first identify fault-fixing commits using the heuristic proposed by Fischer et al. [22], which consists in using regular expressions to detect bug IDs from the studied commit messages. Next, we extract the modified files of each fault-fixing commit through the following Git command:

```
git log [commit-id] -n 1 --name-status
```

We only take modified JavaScript files into account. Given each file F in a commit C, we extract C's parent commit C'. Then, we use Git's `diff` command to extract F's deleted lines. We apply Git's `blame` command to identify commits that introduced these deleted lines, noted as the "candidate faulty changes". We eliminate the commits that only changed blank and comment lines. Then, we filter the commits that were submitted after their corresponding bugs' creation date. Considering the file F in a fault-fixing commit and its commit that introduced faults, we use again Git's `diff` command to extract F's changes between both commits, in order to retrieve the "candidate fault lines" (useful for our line grain analysis). For the next step, we use UglifyJS[20] to get an F's Abstract Syntax Tree (AST) that gives the dependencies of all F's variables, objects and functions (which means their declaration

[18]https://github.com/riot/riot
[19]https://github.com/DavidJohannesWall/smells_project
[20]https://github.com/mishoo/UglifyJS

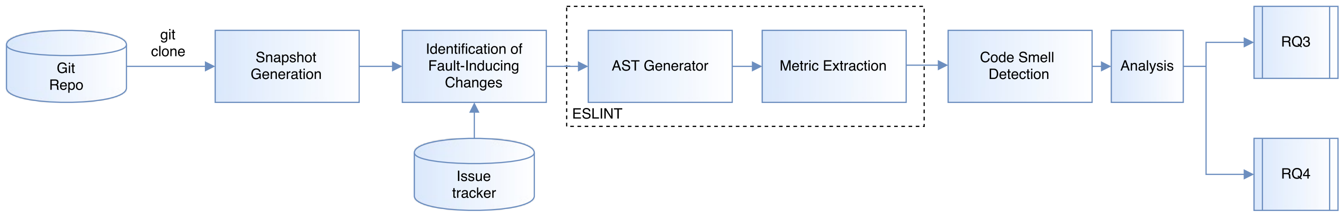


Figure 1: Overview of our approach to answer RQ1 and RQ2.

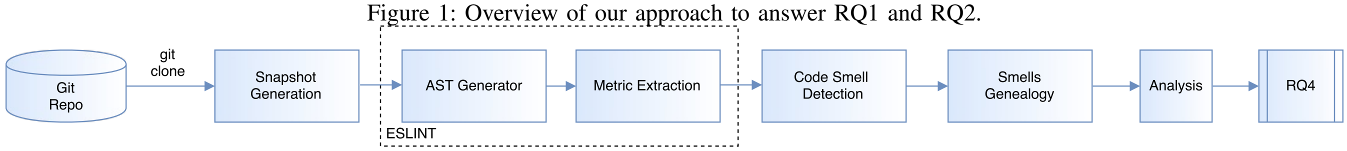


Figure 2: Overview of our approach to answer RQ3.

and use lines). We then match F's dependencies with the "candidate fault lines" to extend them: given an F's element (variable, object, or function), if one of its declaration or use lines is found into the "candidate fault lines", then we add these declaration and use lines to the "candidate fault lines". We finally obtain the "extended candidate fault lines" (usefull for our line grain analysis including dependencies).

AST Generation and Metric Extraction: To automatically detect code smells in the source code, we first extract the Abstract Syntax Tree from the code. AST are being used to parse a source code and generate a tree structure that can be traversed and analyzed programmatically. ASTs are widely used by researchers to analyze the structure of the source code [23]–[25]. We used ESLint[21] which is a popular and open source lint utility for JavaScript as the core of our framework. Linting tools are widely used in programming to flag the potential non-portable parts of the code by statically analyzing them. ESLint is being used in production in many companies like Facebook, Paypal, Airbnb, etc. ESLint uses espree[22] internally to parse JavaScript source codes and extracts Abstract Source Trees based on the specs[23]. ESLint itself provides an extensible environment for developers to develop their own plugins to extract custom information from the source code. We developed our own plugins and modified ESLint built-in plugins to traverse the source tree generated by ESLint to extract and store the information related to our set of code smells described in section II. Table II summarizes all the metrics our framework reports for each type of code smell.

Smells Genealogy: Thanks to our previous extraction methods, we easily get, for each Javascript file of a project, the history of the commits that modified those files. Given the history H of a JavaScript file F, we identifiy and track F's smells through each commit of H. Given two consecutive commits $C1$ and $C2$ of H: if one smell appears in $C2$ (and not in $C1$), we consider it as a new smell and keep its date of creation ($C2$'s date); if one smell disappears in $C2$ (and was present in $C1$), we consider it was killed, and keep its date of destruction ($C2$'s date). If a smell if never killed (present in the last commit of H), we consider its presence until the last project's commit. To measure the similarity degree between two smells, they first need to be from the same smell type, and then we use SequenceMatcher[24] from difflib (a Python library) that gives us a number between 0 and 1 as a similarity degree (1: both smells are the same; 0: they are totally different). We consider two smells as the same if they are from the same smell type (among the 12 studied), and if their similarity degree is greater than 0.7. If one smell of $C1$ gets a similarity degree greater than 0.7 with two smells of $C2$, we keep the maximum in account. We tried our survival analysis of smells with different thresholds of similarity degree (0.8 and 0.9), but we observe no significant difference with the use of 0.7 threshold.

Code Smell Detection: Among of 12 metric values reported by our framework, 4 are boolean. The boolean metrics concern *This Assign*, *Extra Bind*, *Assignment in Conditional Statements*, and *Variable Re-assign* smells. The 8 remaining metrics are integers. To identify code smells using the metric values provided by our framework, we follow the same approach as previous works [1], [26], [27], defining threshold values above which files should be considered as having the code smell. We define the thresholds relative to the systems using Box-plot analysis. We chose to define threshold values relative to the projects because design rules and programming styles can vary from one project to another, and hence it is important to compare the characteristics of files in the context of the project. For each system, we obtain the threshold values as follows. We examined the distribution of the metrics and observed a big gap around the first 70% of the data and the top 10%. Hence, we decided to consider files with metric values in the top 10% as containing the code smell. For files that contain multiple functions, we aggregated the metric values reported for each functions using the maximum to obtain a single value characterizing the file.

C. Analysis

To assess the impact of code smells on the fault-proneness of JavaScript files, or to assess the smells survival over project lifetime, we perform survival analysis, comparing the time until a fault occurrence, in files containing code smells and files without code smells, or comparing the time until a type

[21]http://eslint.org/
[22]https://github.com/eslint/espree
[23]https://github.com/estree/estree
[24]https://docs.python.org/2/library/difflib.html

Table II: Metrics computed for each type of code smell.

Smell Type	Type	Metric
Lengthy Lines	Number	The number of characters per line considering the exceptions described in section II.
Chained Methods	Number	The number chained methods in each chaining pattern.
Long Parameter List	Number	The number of parameters of each function in source code.
Nested Callbacks	Number	The number of nested functions present in the implementation of each function.
Variable Re-assign	Boolean	The uniqueness of variables in same scope.
Assignment in Conditional Statements	Boolean	The presence of assignment operator in conditional statements.
Complex code	Number	The cylcomatic complexity value of each function defined in the source code.
Extra Bind	Boolean	Whether a function is explicitly bound to a context while not using the context.
This Assign	Boolean	Whether `this` is assigned to another variable in a function.
Long Methods	Number	The number of statements in each function.
Complex Switch Case	Number	The number of case statements in each switch-case block in the source code.
Depth	Number	The maximum number of nested blocks in each function.

smell occurrence in files containing code smells, for each of the 12 studied type smell.

Survival analysis is used to model the time until the occurrence of a well-defined event [28]. One of the most popular models for survival analysis is the Cox Proportional Hazards (Cox) model. A Cox hazard model is able to model the instantaneous hazard of the occurrence of an event as a function of a number of independent variables [29] [30]. Particularly, Cox models aim to model how long subjects under observation can *survive* before the occurrence of an event of interest (a fault occurrence in our case) [30] [31].

Survival models were first introduced in demography and actuarial sciences [32]. Recently, researchers have started applying them to problems in the domain of Software Engineering. For example, Selim et al. [31] used the Cox model to investigate characteristics of cloned code that are related to the occurrence of faults. Koru et al. [33] also used Cox models to analyze faults in software systems. In Cox models, the hazard of a fault occurrence at a time t is modeled by the following function:

$$\lambda_i(t) = \lambda_0(t) * e^{\beta * F_i(t)} \quad (1)$$

If we take log from both sides, we obtain:

$$log(\lambda_i(t)) = log(\lambda_0(t)) + \beta_1 * f_{i1}(t) + ... + \beta_n * f_{in}(t) \quad (2)$$

Where:

- $F_i(t)$ is the time-dependent covariates of observation i at the time t.
- β is the coefficient of covariates in the function $F_i(t)$.
- λ_0 is the baseline hazard.
- n is the number of covariates.

When all the covariates have no effect on the hazard, the baseline hazard can be considered as the hazard of occurrence of the event (*i.e.,* a fault). The baseline hazard would be omitted when formulating the relative hazard between two files (in our case) at a specific time, as shown in the following Equation 3.

$$\lambda_i(t)/\lambda_j(t) = e^{\beta * (f_i(t) - f_j(t))} \quad (3)$$

The proportional hazard model assumes that changing each covariate has the effect of multiplying the hazard rate by a constant.

Link function. As Equation 2 shows, the log of the hazard is a linear function of the log of the baseline hazard and all the other covariates. In order to build a Cox proportional model, a linear relationship should be available between the log hazard and the covariates [34]. Link functions are used to transform the covariates to a new scale if such relationship does not exist. Determining an appropriate link function for covariates is necessary because it allows changes in the original value of a covariate to influence the log hazard equally. This allows the proportionality assumption to be valid and applicable [34].

Stratification. In addition to applying a link function, a stratification is sometimes necessary to preserve the proportionality in Cox hazard models [29]. For example, if there is a covariate that needs to be controlled because it is of no interest or secondary, stratification can be used to split the data set so that the influence of more important covariates can be monitored better [29].

Model validation. Since Cox proportional hazard models assume that all covariates are consistent over time and the effect of a covariate does not fluctuate with time, hence, to validate our model, we apply a non-proportionality test to ensure that the assumption is satisfied [34] [31].

In this paper, we perform our analysis at commit level. For each file, we use Cox proportional hazard models to calculate the risk of a fault occurrence over time, considering a number of independent covariates. We chose Cox proportional hazard model, as well as our predecessors [1], for the following reasons:

(1) In general, not all files in a commit experience a fault. Cox hazard models allow files to remain in the model for the entire observation period, even if they don't experience the event (*i.e.,* fault occurrence). (2) In Cox hazard models, subjects can be grouped according to a covariate (*e.g.,* smelly or non-smelly). (3) The characteristics of the subjects might change during the observation period (*e.g.,* size of code), and (4) Cox hazard models are adapted for events that are recurrent [34], which is important because software modules evolve over time and a file can have multiple faults during its life cycle.

IV. Case Study Results

In this section, we report and discuss the results for each research question. For each research question, we collect information about smell, and fault hazard codes of JavaScript files of the studied systems, and more specifically those which end with the *.js* extension. Also, we don't take in account the JavaScript files with *.min.js* extension, because they are a minified version of *.js* files that we already keep in our study. In this way, we avoid redundancy in our analyzes.

(RQ1) Is the risk of fault higher in files with code smells in comparison with those without code smell?

Approach. We use our framework described in Section III-B (Figure 1) to collect information about the occurrence of the 12 studied code smells in our fifteen subject systems. For each file and for each revision r (*i.e.,* corresponding to a commit), we also compute the following metrics:

- **Time:** the number of hours between the previous revision of the file and the revision r. We set the time of the first revision to zero.
- **Smelly:** this is our covariate of interest. It takes the value 1 if the revision r of the file contains a code smell and 0 if it doesn't contain any of the 12 studied code smells.
- **Event:** For the line grain and line grain including dependencies approaches, this metric takes the value 1 if the revision r is a fault-fixing change and if there is at least one match between the fault lines and the smell lines, and 0 otherwise. Indeed, if there is no matching, we consider that the fault-fixing change doesn't fix any code smells. We use the SZZ algorithm to insure that the file contained a code smell when the fault was introduced.

Using the smelly metric, we divide our dataset in two groups: one group containing files with code smells (*i.e.,* smelly = 1) and another group containing files without any of the 12 studied code smells (*i.e.,* smelly = 0). For each group we create an individual Cox hazard model. In each group, the covariate of interest (*i.e.,* smelly) is a constant function (with value either 1 or 0), hence, there is no need for a link function to establish a linear relationship between this covariate and our event of interest, *i.e.,* the occurrence of a fault. We use the *survfit* and *coxph* functions from R [35] to analyze our Cox hazard models.

In addition to building Cox hazard models, and as well as our predecessors [1], we test the following null hypothesis: H_0^1: *There is no difference between the probability of a fault occurrence in a file containing code smells and a file without code smells*. We use the *log-rank* test (which compares the survival distributions of two samples), to accept or refute this null hypothesis.

Findings. Table III (line grain results) and Figure 3 (line grain including dependencies results) show that files containing code smells experience faults faster than files without code smells, as observed by our predecessors (for their file grain approach). The Y-axis in Figure 3 represents the probability of a file *surviving* a fault occurrence. Hence a low value on the Y-axis means a low *survival* rate (*i.e.,* a high hazard or high risk of fault occurrence). For all fifteen projects, and for each approach (line grain, and line grain including dependencies), we calculated relative hazard rates (using Equation 3 from Section III-C) between files containing code smells and files without code smells. Results show that, on average, files without code smells have hazard rates 20% lower than files with code smells in our line grain analysis, and 38% lower in our line grain analysis including dependencies. It is normal to see this pourcentage decreasing in line grain approach (in comparison with the results of our predecessors [1]), because we add an additional matching condition to set the *event* to 1. Between the line grain and the line grain including dependencies analyzes, this pourcentage increases due to the increasing of the fault lines number considering during the compute of *event*. We performed a *log-rank* test comparing the survival distributions of files containing code smells and files without any of the studied code smells and obtained p-values lower than 0.05 for most of the fifteen studied systems. Hence, we reject H_0^1. Since our detection of code smells depends on our selected threshold value (*i.e.,* the top 10% value chosen in Section III-B), we conducted a sensitivity analysis to assess the potential impact of this threshold selection on our result. More specifically, we rerun all our analysis with threshold values at top 20% and top 30%. We observed no significant differences in the results. Hence, we conclude that:

> *JavaScript files without code smells have hazard rates 20% lower than JavaScript files with code smells in the line grain approach, and this difference is statistically significant. Plus, this difference still remains significant and increases in the line grain including dependencies approach, because hazard rates reach 38%.*

(RQ2) Are JavaScript files with code smells equally fault-prone?

Approach. Similar to **RQ1**, we use our framework from Section III-B ((Figure 1)) to collect information about the occurrence of the 12 studied code smells in our fifteen subject systems. For each file and for each revision r (*i.e.,* corresponding to a commit), we also compute the **Time** and **Event** metrics defined in **RQ1**. For each type of code smell i we define the metric **Smelly$_i$:** which takes the value 1 if the revision r of the file contains the code smell i and 0 if it doesn't contain any of the 12 studied code smells. Also, in respect with our line grain and line grain including dependencies approaches, we define the metric **Event$_i$:** which takes the value 1 if the revision r is a fault-fixing change and if the code smell i is in the intersection between the fault lines and the smell lines, and 0 otherwise. When computing the **Event** and **Event$_i$** metrics, we used the SZZ algorithm to ensure that the file contained the code smell i when the fault was introduced. Because size, code churn, and the number of past occurrence of faults are known to be related to fault-proneness, we add the following metrics to our models, to control for the effect of these covariates : (i) LOC: the number of lines of code in the file at revision r; (ii) Code Churn: the sum of added, removed and modified lines in the file prior to revision r; (iii) No. of Previous-Bugs: the number of fault-fixing changes experienced by the file prior to revision r. We perform a stratification considering the covariates mentioned above, in order to monitor their effect on our event of interest, *i.e.,* a fault occurrence. Next, we create a Cox hazard model for each of our fifteen studied

Table III: Fault hazard ratios for each project with the line grain approach. $exp(coef)$ values means higher hazard rates.

module	$exp(coef)$	p-value (Cox hazard model)	p-value (Proportional hazards assumption)
express	1.341	0.002	0.192
request	2.538	0.028e-2	0.602
less	1.791	0.003	0.419
bower	1.321	0.027	0.982
grunt	0.594	0.005	0.019
jquery	3.436	0	1.197e-8
vue	0.062	0.011e-9	0.843
ramda	0.460	0.01e-8	3.691e-5
leaflet	0.725	0.088e-4	0.739
hexo	1.199	0.077	0.945
chart	0.711	0.136	1.46e-9
webpack	0.603	0	0
webtorrent	1.222	0.047e-9	0.045
moment	0.941	0.401	5.046e-5
riot	1.047	0.586	0.797

systems. In order to build an appropriate link function for the new covariates considered in this research question (*i.e.,* LOC, Code churn, and No. of Previous-Bugs), we follow the same methodology as [29] [31] and plot the log relative risk vs. each type of code smell, the No. of Previous-Bugs, LOC and Code Churn in each of our fifteen datasets (corresponding to the fifteen subject systems). We generated summaries of all our Cox models and removed insignificant covariates, *i.e.,* those with p-values greater than 0.05. Finally, for each system, we performed a non-proportional test to verify if the proportional hazards assumption holds.

Findings. Tables IV, and V summarize the fault hazard ratios for the 12 studied code smells for respectively the line grain, and line grain including dependencies approach. The value in the column exp(coef) shows the amount of increase in hazard rate that one should expect for each unit increase in the value of the corresponding covariate. The last column of Tables IV, and V show that the p-values obtained for the non-proportionality tests are above 0.05 for all the fifteen systems; meaning that the proportional hazards assumption is satisfied for all the fifteen studied systems. Actually, we removed from the tables the insignificant covariates, which means those with non-proportionality test p-values less than 0.05, and we will only consider the covariates with *exp(coef)* greater than 1 (for those the corresponding files are more fault-proneness when they are smelly).

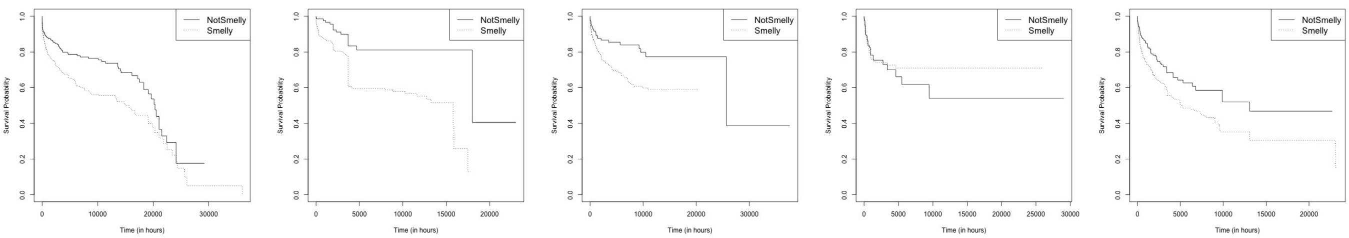
Overall, the hazard ratios of the studied code smells vary across the systems and accross the approaches (line grain and line grain including dependencies). With our line grain approach, *Variable Re-assign* has one of the highest hazard ratio in most systems, that is to say in four out of fifteen systems (27%); *Assignment in Conditional Statements* has one of the highest hazard rate in two out of fifteen systems (13%); *Lenghty Lines*, and *Complex Code* are the most hazard code smell in only one out of fifteen systems (7%); the other smells don't appear in any of the studied systems as having a high hazard ratio. With the last approach (line grain including dependencies), the results are a little different. *Variable Re-assign* still is one of the most hazard code smell in most systems, in seven out of fifteen systems (47%); *Complex Code* has one of the highest hazard rate in three out of fifteen systems (20%); *Lenghty Lines*, *Assignment in Conditional Statements*, and *Long Methods* are the most hazard code smells in only one out of fifteen systems (7%); the other smells don't have an enough high hazard rate in any of the studied systems. Furthermore, the most hazard types of code smell seem not to vary accross the approaches, and this observation particularly affects *Variable Re-assign*, *Assignment in Conditional Statements*, and *Complex Code* smells, which is consistent with the work of Amir Saboury and al. [1].

As we expected, in our both approaches, the covariates No.Previous-Bugs is significantly related to fault occurrence, because it appears in at least eight out of fifteen systems with an *exp(coef)* greater than 1 and good p-values. However, its hazard rate is lower than those of many of the studied code smells. LOC is significantly related to fault occurrence in seven systems in our both approaches (less than half of the studied systems) with a very low hazard rates, meaning that JavaScript developers cannot simply control for size and monitor files with previous fault occurrences, if they want to track fault-prone files effectively. Since *Variable Re-assign*, *Assignment in Conditional Statements*, and *Complex Code* are related to high hazard ratios in respectively 37%, 10% and 13% of the cases (which means fifteen studied systems and two approches, that is to say 30 cases), we strongly recommend that developers prioritize files containing these three types of code smells during testing and maintenance activities.

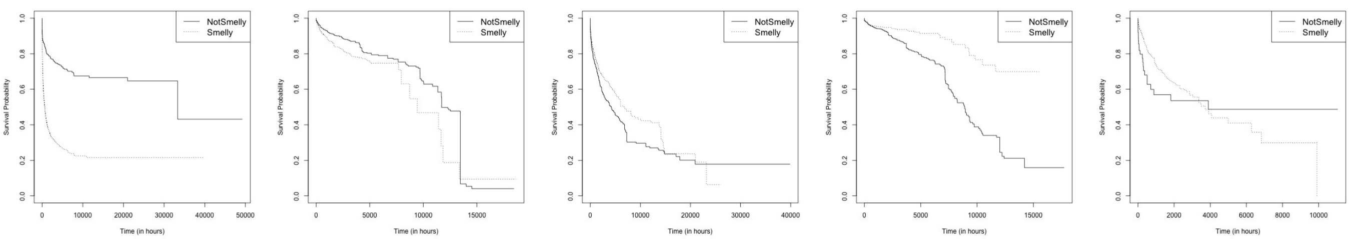
JavaScript files containing different types of code smells are not equally fault-prone. Developers should consider refactoring files containing either Variable Re-assign code smell, or Assignment in Conditional Statements code smell, or Complex Code smell in priority since they seem to increase the risk of faults in the system.

Similar to **RQ1**, we conducted a sensitivity analysis to assess the potential impact of our threshold selection (performed during the detection of code smells) on the results; rerunning the analysis using threshold values at top 20% and top 30%. We did not observed any significant change in the results.

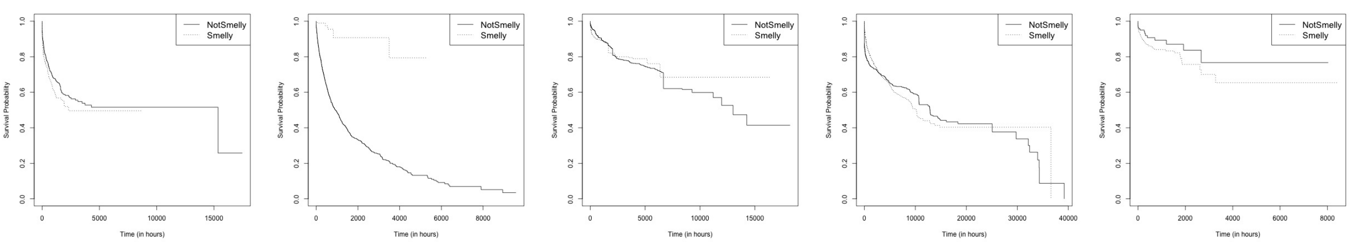
(RQ3) How do the smells survive over time?



(a) express.js (b) request.js (c) less.js (d) grunt.js (e) bower.js

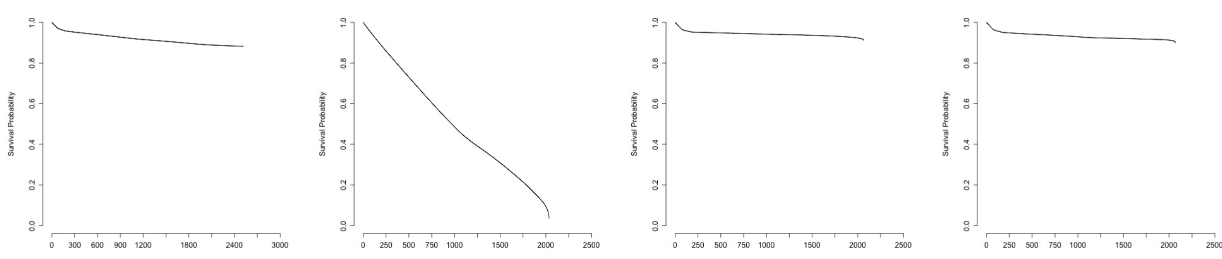


(f) jquery.js (g) hexo.js (h) leaflet.js (i) ramda.js (j) chart.js



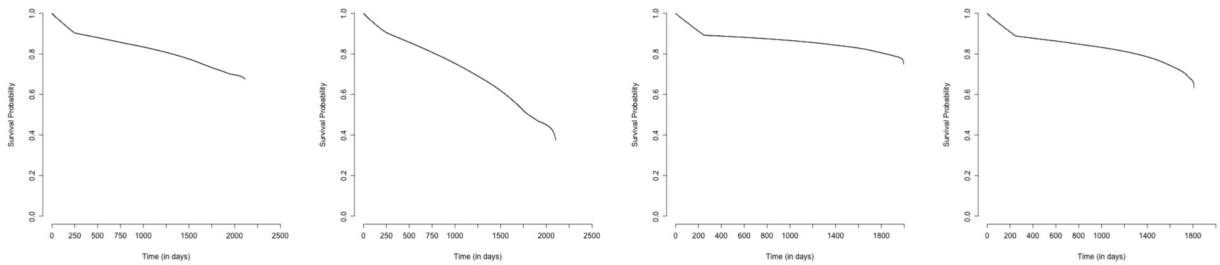
(k) riot.js (l) vue.js (m) moment.js (n) webpack.js (o) webtorrent.js

Figure 3: Survival probability trends of smelly codes vs. non-smelly codes in our fifteen JavaScript projects with the line grain including dependencies approach.



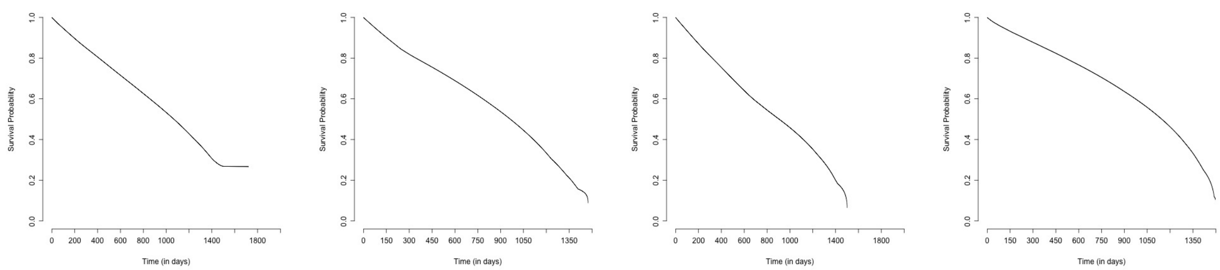
(a) Variable Re-Assign (b) Nested Callbacks (c) Complex Code (d) Long Methods

Figure 4: Survival analyzes of the largest smells of express.js.



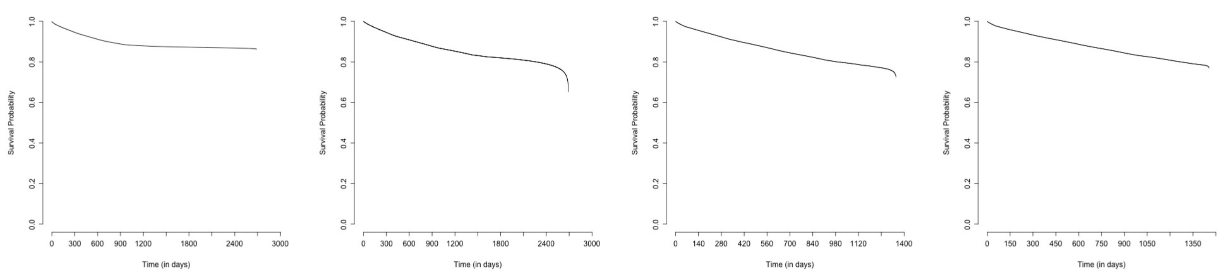
(a) Variable Re-Assign (b) Lengthy Lines (c) Complex Code (d) Long Methods

Figure 5: Survival analyzes of the largest smells of grunt.js.



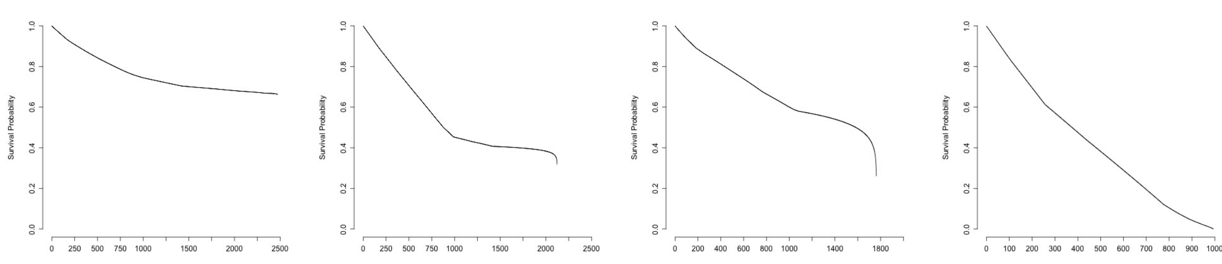
(a) Variable Re-Assign (b) Chained Methods (c) Lengthy Lines (d) Nested Callbacks

Figure 6: Survival analyzes of the largest smells of bower.js.



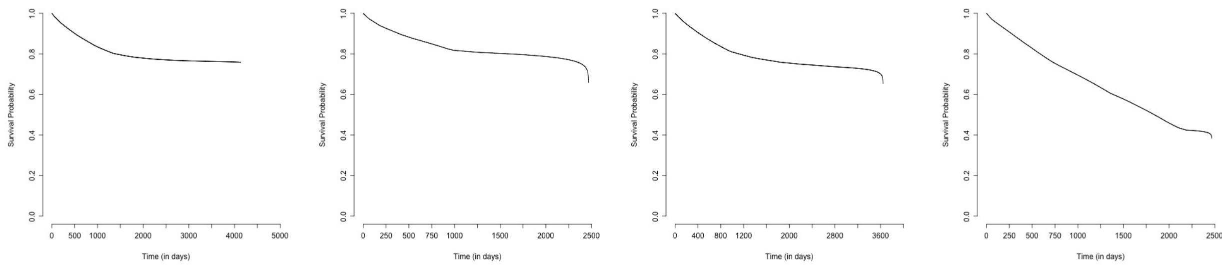
(a) Variable Re-Assign (b) Nested Callbacks (c) Complex Code (d) Long Methods

Figure 7: Survival analyzes of the largest smells of less.js.



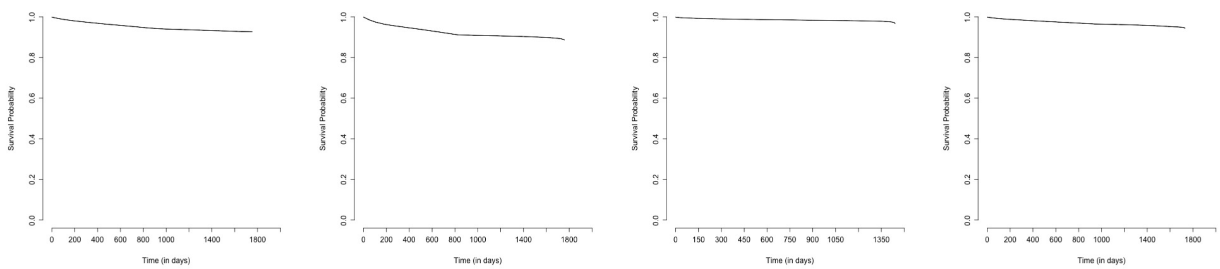
(a) Variable Re-Assign (b) This Assign (c) Chained Methods (d) Nested Callbacks

Figure 8: Survival analyzes of the largest smells of request.js.



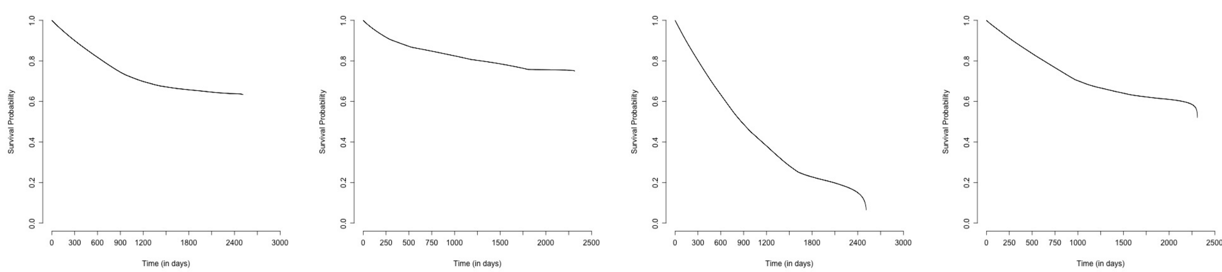
(a) Variable Re-Assign (b) Complex Switch Case (c) This Assign (d) Chained Methods

Figure 9: Survival analyzes of the largest smells of jquery.js.



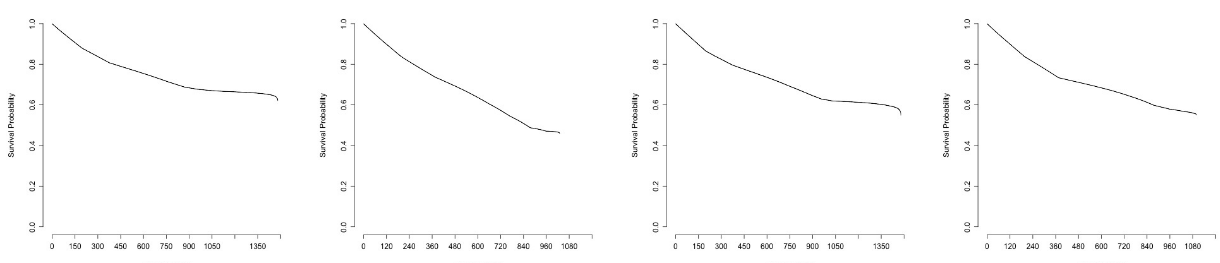
(a) Variable Re-Assign (b) Lengthy Lines (c) Long Parameter List (d) Complex Code

Figure 10: Survival analyzes of the largest smells of hexo.js.



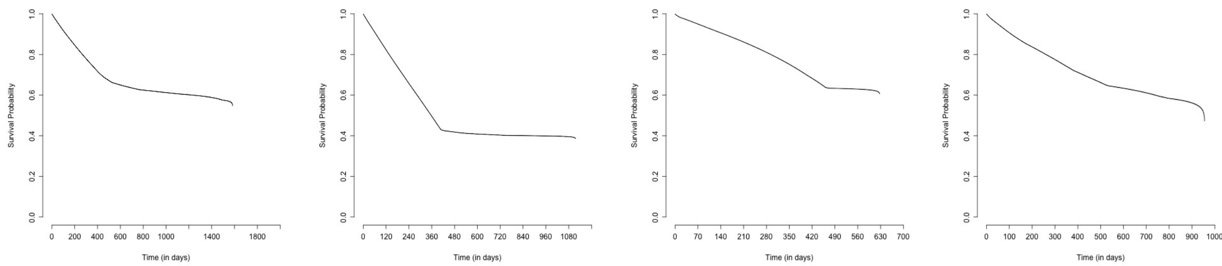
(a) Variable Re-Assign (b) Lengthy Lines (c) Nested Callbacks (d) Long Methods

Figure 11: Survival analyzes of the largest smells of leaflet.js.



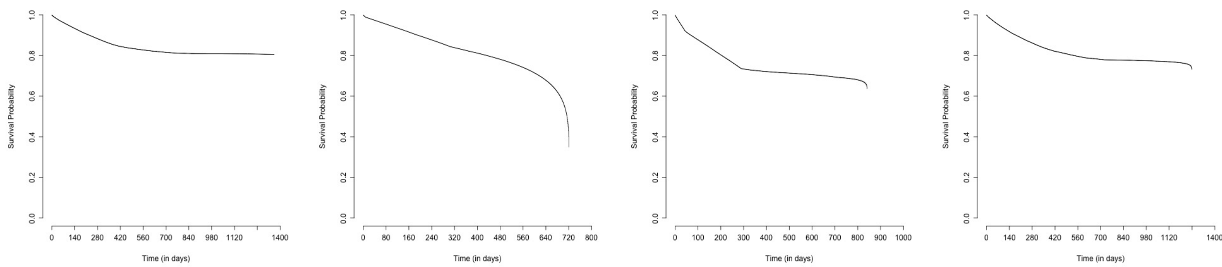
(a) Variable Re-Assign (b) Chained Methods (c) Long Parameter List (d) Complex Code

Figure 12: Survival analyzes of the largest smells of ramda.js.



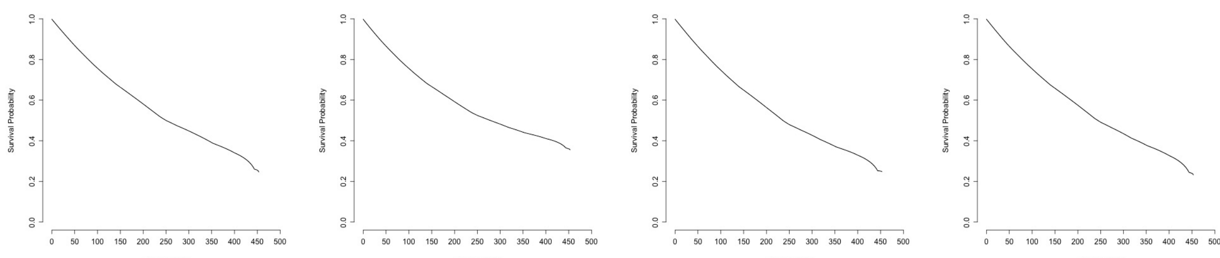
(a) Variable Re-Assign (b) This Assign (c) Lengthy Lines (d) Complex Switch Case

Figure 13: Survival analyzes of the largest smells of chart.js.



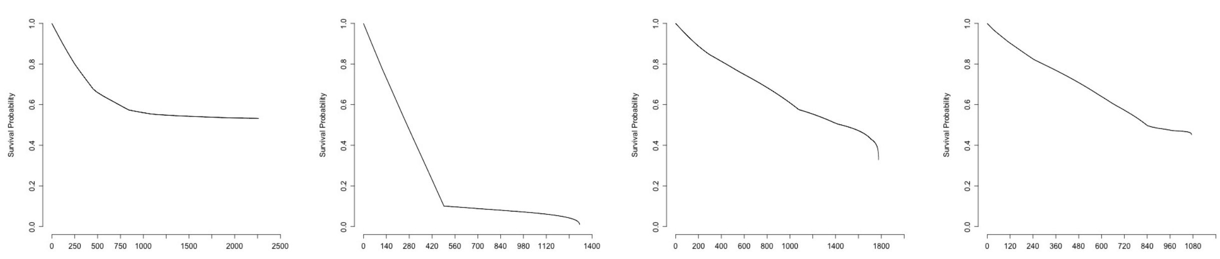
(a) Variable Re-Assign (b) Lengthy Lines (c) This Assign (d) Assignment in Conditional Statement

Figure 14: Survival analyzes of the largest smells of riot.js.



(a) Variable Re-Assign (b) Lengthy Lines (c) Complex Code (d) Long Methods

Figure 15: Survival analyzes of the largest smells of vue.js.



(a) Variable Re-Assign (b) Nested Callbacks (c) Complex Switch Case (d) Chained Methods

Figure 16: Survival analyzes of the largest smells of moment.js.

Table IV: Hazard ratios for each type of code smells with line grain approach. Higher $exp(coef)$ values means higher hazard rates.

module	covariate	$exp(coef)$	p-value (Cox hazard model)	p-value (Proportional hazards assumption)
express	No.Previous-Bugs	1.034	0	0.126
	Variable Re-assign	1.22	0.024	0.814
grunt	No.Previous-Bugs	1.064	0.009	0.693
	Assign. in Cond. State.	0.315	0.047	0.628
	Chained Methods	0.206	0.002	0.94
	Lengthy Lines	0.154	0.036e-3	0.053
bower	No.Previous-Bugs	1.051	0	0.679
	LOC	1.001	0.064e-11	0.480
	Variable Re-assign	1.335	0.02	0.832
	This Assign	0.394	0.012e-5	0.998
less	No.Previous-Bugs	1.024	0	0.566
	Variable Re-assign	1.446	0.033	0.652
	Assign. in Cond. State.	1.342	0.036	0.052
request	No.Previous-Bugs	1.061	0.013e-13	0.185
	LOC	1.001	0	0.517
	Variable Re-assign	1.913	0.003	0.455
jquery	LOC	1.0001	0	0.164
	Lengthy Lines	2.002	0	0.385
	Assign. in Cond. State.	1.881	0.014e-6	0.123
	Complex Code	1.684	0	0.164
hexo	No.Previous-Bugs	1.25	0	0.296
	LOC	1.001	0.015e-8	0.251
leaflet	No.Previous-Bugs	1.016	0	0.08
	LOC	1.0001	0.01e-2	0.192
	Variable Re-assign	0.712	0.025e-4	0.578
	Complex Code	0.485	0.003	0.179
	Chained Methods	0.405	0.034e-4	0.201
ramda	LOC	1.0001	0.037	0.225
	Nested Callbacks	0.399	0.014e-2	0.271
	Complex Code	0.242	0.046	0.322
	Chained Methods	0.231	0.038	0.996
chart	No.Previous-Bugs	1.151	0	0.113
	Nested Callbacks	0.321	0.034e-7	0.747
	This Assign	0.265	0	0.198
	Long Parameter List	0.191	0.036e-4	0.306
riot	This Assign	0.125	0.047e-6	0.8
	Long Parameter List	0.105	0.069e-4	0.277
vue	Variable Re-assign	0.069	0.065e-9	0.678
	This Assign	0.017	0.049e-3	0.156
moment	No.Previous-Bugs	1.015	0	0.496
	Long Methods	0.261	0.003	0.425
webpack	Extra Bind	0.295	0.035	0.969
webtorrent	No.Previous-Bugs	1.054	0.075e-5	0.058
	LOC	1.001	0.018e-2	0.095
	Nested Callbacks	0.17	0.013	0.405

Table V: Hazard ratios for each type of code smells with line grain including dependencies approach. Higher $exp(coef)$ values means higher hazard rates.

module	covariate	$exp(coef)$	p-value (Cox hazard model)	p-value (Proportional hazards assumption)
express	No.Previous-Bugs	1.034	0	0.124
	Complex Code	1.955	0.03e-2	0.152
	Long Methods	1.674	0.024	0.227
	Variable Re-assign	1.354	0.042e-2	0.742
grunt	No.Previous-Bugs	1.072	0.035e-2	0.495
	Lengthy Lines	0.402	0.001	0.288
bower	No.Previous-Bugs	1.05	0	0.855
	LOC	1.001	0.07e-11	0.652
	Complex Code	2.314	0.006	0.734
	Variable Re-assign	1.579	0.019e-2	0.601
less	No.Previous-Bugs	1.023	0	0.522
	Variable Re-assign	1.616	0.005	0.463
request	No.Previous-Bugs	1.056	0.038e-13	0.188
	LOC	1.001	0.022e-14	0.664
	Variable Re-assign	2.316	0.089e-3	0.66
jquery	LOC	1.0001	0	0.161
	Lengthy Lines	2.002	0	0.385
	Assign. in Cond. State.	1.881	0.014e-6	0.123
	Complex Code	1.684	0	0.164
hexo	No.Previous-Bugs	1.254	0	0.325
	LOC	1.001	0.019e-11	0.269
	Variable Re-assign	1.321	0.004	0.896
leaflet	LOC	1.0001	0.081e-10	0.101
	Variable Re-assign	0.755	0.047e-3	0.435
	Complex Code	0.485	0.003	0.179
	Chained Methods	0.434	0.091e-4	0.222
ramda	LOC	1.0001	0.037	0.228
	Nested Callbacks	0.579	0.007	0.978
	Complex Code	0.242	0.046	0.322
chart	No.Previous-Bugs	1.15	0	0.122
	Nested Callbacks	0.321	0.034e-7	0.747
	This Assign	0.281	0	0.173
	Long Parameter List	0.191	0.036e-4	0.306
riot	Variable Re-assign	1.26	0.004	0.258
	Chained Methods	0.22	0.067e-3	0.602
	This Assign	0.125	0.046e-6	0.93
vue	Variable Re-assign	0.092	0.018e-9	0.149
	Depth	0.033	0.066e-2	0.187
moment	No.Previous-Bugs	1.015	0	0.493
	This Assign	0.384	0.003	0.565
webpack	Nested Callbacks	0.381	0.002	0.439
webtorrent	LOC	1.001	0.08e-6	0.054
	Variable Re-assign	1.654	0.043	0.647
	Nested Callbacks	0.255	0.019	0.491

Table VI: Descriptive statistics on survival over time of the largest smells of studied systems.

System	Smell	Not Survived	Survived	Number created at file birth	Median days or survival	Average days of survival
express	Variable Re-assign	6743	425	5783 (80.7%)	74	209
	Nested Callbacks	314	728	417 (40%)	1101	1152
	Complex Code	374	5	353 (93.1%)	74	122
	Long Methods	283	9	260 (89%)	74	143
	SUM	8430	1238	7348 (76%)		
grunt	Variable Re-assign	2210	317	1636 (64.7%)	248	411
	Lengthy Lines	243	108	172 (49%)	248	681
	Complex Code	91	5	83 (85.4%)	248	292
	Long Methods	55	5	41 (68.3%)	248	334
	SUM	2732	448	2017 (63.4%)		
bower	Variable Re-assign	1427	1801	1235 (38.3%)	797	777
	Chained Methods	82	96	35 (19.7%)	1231	819
	Lengthy Lines	32	50	28 (34.1%)	644	719
	Nested Callbacks	38	32	27 (38.6%)	163	656
	SUM	1647	2087	1368 (36.6%)		
less	Variable Re-assign	36979	3779	37303 (91.5%)	56	281
	Assign. in Cond. State.	1349	4	1261 (93.2%)	405	477
	Lengthy Lines	547	26	509 (88.8%)	36	139
	This Assign	392	33	387 (91.1%)	129	314
	SUM	40241	3927	40391 (91.4%)		
request	Variable Re-assign	1362	667	1140 (56.2%)	365	534
	This Assign	28	50	34 (43.6%)	874	743
	Chained Methods	32	22	38 (70.4%)	256	557
	Nested Callbacks	1	28	1 (3.4%)	774	592
	SUM	1455	777	1232 (55.2%)		
jquery	Variable Re-assign	13076	5156	12122 (66.5%)	528	694
	Complex Switch Case	146	15	130 (80.7%)	179	435
	This Assign	118	37	120 (77.4%)	539	716
	Chained Methods	59	58	35 (29.9%)	657	785
	SUM	13743	5356	12675 (66.4%)		
hexo	Variable Re-assign	19023	823	17626 (88.8%)	2	86
	Lengthy Lines	768	12	675 (86.5%)	10	138
	Long Parameter List	755	3	728 (96%)	2	20
	Complex Code	599	19	576 (93.2%)	2	51
	SUM	22522	980	20584 (87.6%)		
leaflet	Variable Re-assign	5856	498	2241 (35.3%)	789.5	734
	Lengthy Lines	733	270	358 (35.7%)	203	354
	Nested Callbacks	123	489	114 (18.6%)	986	1029
	Long Methods	77	5	32 (39%)	911	752
	SUM	6997	1278	2855 (34.5%)		
ramda	Variable Re-assign	4720	1078	4656 (80.3%)	375	391
	Chained Methods	365	101	344 (73.8%)	241	372
	Long Parameter List	176	90	208 (78.2%)	206	396
	Complex Code	194	28	200 (90.1%)	375	364
	SUM	5668	1340	5599 (79.9%)		
chart	Variable Re-assign	5297	5696	4538 (41.3%)	406	365
	This Assign	199	388	86 (14.7%)	406	339
	Lengthy Lines	119	14	34 (25.6%)	12	154
	Complex Switch Case	53	30	31 (37.3%)	169	258
	SUM	5740	6207	4746 (39.7%)		
riot	Variable Re-assign	8331	2625	7866 (71.9%)	52	188
	Lengthy Lines	193	33	206 (91.2%)	7	150
	This Assign	63	92	61 (39.4%)	43	110
	Assign. in Cond. State.	107	47	80 (51.9%)	100.5	196
	SUM	9119	2937	8637 (71.6%)		
vue	Variable Re-assign	4199	5833	6587 (65.7%)	139	175
	Lengthy Lines	2947	4208	3518 (49.2%)	125	143
	Complex Code	414	675	679 (62.4%)	139	171
	Long Methods	259	417	421 (62.3%)	139	171
	SUM	8612	11712	12129 (59.7%)		
moment	Variable Re-assign	5642	11063	6154 (36.8%)	450	486
	Nested Callbacks	19	335	12 (3.4%)	492	464
	Complex Switch Case	117	69	114 (61.3%)	299	634
	Chained Methods	113	42	74 (47.7%)	243	389
	SUM	6135	11590	6561 (37%)		
webpack	Variable Re-assign	4643	627	3192 (60.6%)	352	593
	Nested Callbacks	379	54	276 (63.7%)	104	359
	Chained Methods	371	16	174 (45%)	492	589
	Lengthy Lines	182	38	86 (39.1%)	236	518
	SUM	5970	813	3987 (58.8%)		
webtorrent	Variable Re-assign	709	424	471 (41.6%)	335	370
	This Assign	108	53	88 (54.7%)	335	427
	Nested Callbacks	28	55	23 (27.7%)	453	406
	Chained Methods	12	2	9 (64.3%)	19.5	122
	SUM	869	535	591 (42.1%)		

Approach. We use now the framework described in Section III-B, Figure 2, to collect information about the appearance of the 12 studied code smells in our fifteen subject systems, as well as their line localization, their content and their genealogy (which means their evolution over time from their creation to either their destruction, or the last revision of the studied system). For each studied system and for each smell type, we compute the following metrics:

- The number of created smells.
- The number of killed smells (over the system lifetime).
- The number of survived smells, which means the number of smell that presently appear in the system.
- The number of smells created at the file birthdate.
- The median days of survival of the smells.
- The average days of survival of the smells.

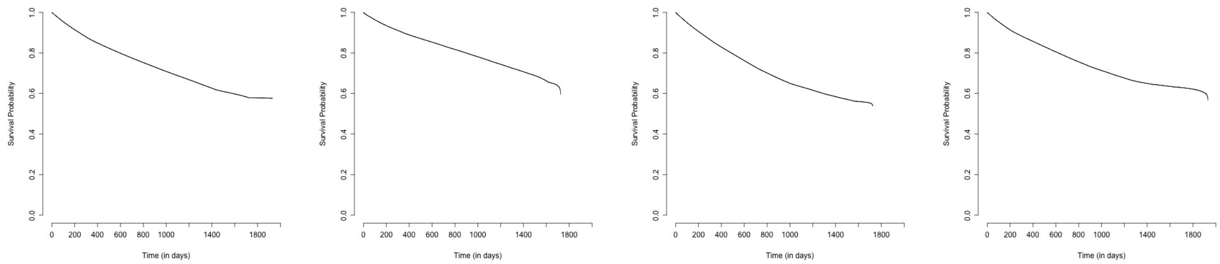
For each smell created (which means never encountered before), we also compute the **Time** and **Event** metrics thus defined:

- **Time:** the time in days since the smell creation.
- **Event:** this metric takes the value 1 if the studied smell is present at this time (which means not killed), and 0 otherwise.

In this way, if a particular smell s is killed x days after its introduction, we will have the corresponding event metric equal to 1 from 0 to $x-1$, and equal to 0 at the time x and after. When we report those information for a studied system, the maximum time that we take in account, for a particular smell type, corresponds to the maximum lifetime of the smells of this type. Thereby, for each smell type of the system and for each time, we will know the proportion of smells alive relatively to the number of smells created. This will particularly help us in the Cox survival model design.

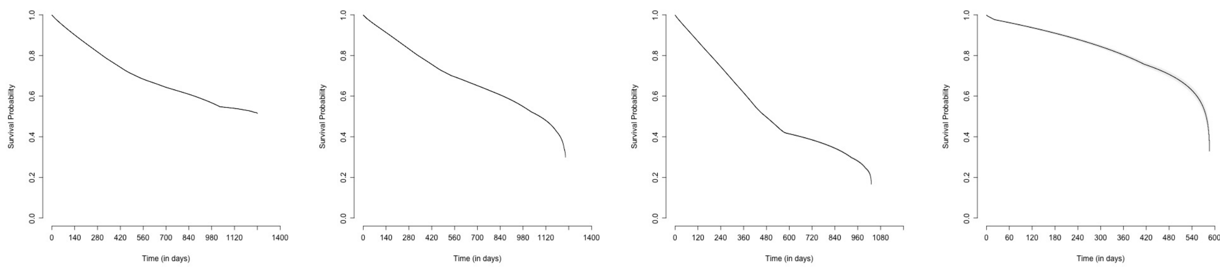
Then, for each of the twelve studied smells, and for each of the fifteen studied systems, we create an individual Cox survival model using the **Time** and **Event** metrics previously defined. We use the *survfit* and *coxph* functions from R [35] to analyze our Cox survival models.

Findings. Our results are presented in the Table VI for a density analysis, and in the Figures 4 to 18 for a survival analysis. For the Table VI, for each system and each smell, the third column corresponds to the number of killed smells, and the fourth to the number of survived smells. The sum of both columns gives us the number of created smells. The fifth column reports, in pourcentage, the proportion of smells created at the files birthdate, relatively to the number of created smells. In order to not overload the presentation of our results, we only report the descriptive statistics for the four most relevant smells, which means those for wich the number created is the most considerable. Finally, the Table reports, for each studied system, general statistics (*SUM* lines), computed by summing the statistics of the twelve studied smells. For the Figures 4 to 18, we plot the survival analysis for each studied systems, and for each smells reported in the Table VI, still in order to not overload the presentation of our results. The Y-axis corresponds to the chance of surviving of a given smell type, x days after its introduction into the codebase. The results presented in Table VI show that smells are not often introduced during files evolution and changes, but rather at the creation of files. Indeed, when we look at the *SUM* lines, from 34.5% (for *leaflet*) to 91.4% (for *less*) of the smells are introduced at the file birthdate, meaning that developers should be aware to their code when they create a JavaScript file, because it is precisely at this moment that most of the smells are introduced into the system. We also notice that, for the major part of the studied systems (eight out of fifteen), more than 20% of the smells created still survive presently; and for thirteen systems, over than 10% of the smells created are now present in those systems. It reveals that a significant part of the smells are never removed from the system once they are introduced in the code. Plus, after analyzing the commits of the studied systems, it is interesting to notice that most of time, the killed smells are removing at the same time than the file containing them (and not because of a file fix). The Table gives us also an overview of the smells lifetime, and we observe that for most of the systems (nine out of fifteen), the median and average days of survival of the most significant smell types are greater than 100 days; and for fourteen systems out of fifteen (except *hexo*), at least one of the most sizable smell types has an average and median lifetime greater than 100 days. This observation highlights that in general, smells tend to survive a very long time inside the system once they are introduced. Finally, Table VI presents an interesting result, which is that the smell *Variable Re-assign* is always the most considerable smell type (in our fifteen studied systems) in term of number of created smells, and its survival rate follows the trend of the sum of the smells (when we consider all the created smells of the system). For every studied system, over 1000 *Variable Re-assign* smells are created, and for eight systems out of fifteen, the number of created smells of this type exceeds 10000. Once again, *Variable Re-assign* is at the heart of our analysis, because as said previously, it is one of the most risky smell in terms of fault-proneness. According to our Figures 4 to 18, the four most significant smell types of each studied system have a considerable chance of surviving 500 days after their introduction. This is indeed the case for all the most significant smell types for nine systems out of fifteen (except *request*, *chart*, *moment*, *webtorrent*, and *vue*), with over 50% chance of surviving 500 days after the introduction of their largest smell types. Also, for fourteen studied systems out of fifteen (except *vue*), at least one of the most sizable smell types has more than 50% chance of surviving 500 days after its smells introduction. Plus, for twelve systems out of fifteen (except *bower*, *vue*, and *webtorrent*), the *Variable Re-assign* smell type has over than 50% chance of surviving 1500 days after its introduction. These observations show the trend of the smells of the studied systems to be persistent and survive a long time after their were introduced into the code, and also the significance of *Variable Re-assign* which is strongly linked to fault-proneness, and is the most proliferated smell type in the studied systems with a very high chance of surviving over time.



(a) Variable Re-Assign (b) Nested Callbacks (c) Chained Methods (d) Lengthy Lines

Figure 17: Survival analyzes of the largest smells of webpack.js.



(a) Variable Re-Assign (b) This Assign (c) Nested Callbacks (d) Chained Methods

Figure 18: Survival analyzes of the largest smells of webtorrent.js.

Most of the studied smells (from 34.5% to 91.4%) are introduced during the creation of JavaScript files. Once introduced, in most of half of the cases (eight systems), over than 20% of the studied smells are not removed and have a high chance of surviving a very long time. Plus, Variable Re-assign, which is the most subject to fault-proneness, is also the most sizable smell type with the highest chance of surviving over time.

V. Threats to Validity

In this section, we discuss the threats to validity of our study following common guidelines for empirical studies [36].

Construct validity threats concern the relation between theory and observation. In our study, threats to the construct validity are mainly due to measurement errors. The number of previous faults in each source code file was calculated by identifying the files that were committed in a fault fixing revision. This technique is not without flaws. We identified fault fixing commits by mining the logs searching for certain keywords (*i.e.,* "bug","fix"","defect" and "patch") as explained in Section III-B. Following this approach, we are not able to detect fault fixing revisions if the committer either misspelled the keywords or failed to include any commit message. Nevertheless, this heuristic was successfully used in multiple previous studies in software engineering [7], [37]. The SZZ heuristic used to identify fault-inducing commits is not 100% accurate. However, it has been successfully used in multiple previous studies from the literature, with satisfying results. In our implementation, we remove all fault-inducing commit candidates that only changed blank or comment lines. When analyzing the *smelliness* of files that experienced fault-inducing changes, we only tracked the presence of the smell in the file as a whole. Hence, the smell contained in the file may not have been involved in the changed lines that induced the fault.

Internal validity threats concern our selection of systems and tools. The metric extraction tool used in this paper is based on the AST provided by ESLint. The results of the study are therefore dependent on the accuracy of ESLint. However, we are rather assured that this tool functions properly as it is being used widely by big companies. *e.g.,* Facebook, Paypal, Airbnb. We chose a logarithmic link function for some of our covariates in the survival analysis. It is possible that a different link function would be a better choice for these covariates. However, the non-proportionality test implies that the models were a good fit for the data. Also, we do not claim causation in this work, we simply report observations and correlations and tries to explain these findings.

Threats to conclusion validity address the relationship between the treatment and the outcome. We are careful to acknowledge the assumptions of each statistical test.

Threats to external validity concern the possibility to generalize our results. In this paper, we have studied fifteen large JavaScript projects. We have also limited our study to open-source projects. Still, these projects represent different domains and various project sizes. Table I shows a summary of the studied systems, their domain and their size. Nevertheless, further validation on a larger set of JavaScript systems, considering more types of code smells is desirable.

Threats to reliability validity concern the possibly of replicating our study. In this paper, we provide all the details needed to replicate our study. All our fifteen subject systems are publicly available for study. The data and scripts used in this study is also publicly available on Github[25].

Threats to internal genealogy construction is about our way to get the smells genealogy of the studied smells, more specifically the recognition of the smells over time and commits. Indeed, we set a similarity threshold of 70%, meaning that if two smells of the same type have a similarity greater than 70%, there are likely the same. Obviously, this threshold is not perfect and can associate two different smells together, or dissociate two smells, which are in reality the same. However, we changed it in order to see if some significant differences would appear, but no relevant difference was revealed.

VI. Related Work

In this section, we discuss the related literature on code smell and JavaScript systems. Code Smells [5] are poor design and implementation choices that are reported to negatively impact the quality of software systems. They are opposite to design patterns [38] which are good solutions to recurrent design problems. The literature related to code smells generally falls into three categories: (1) the detection of code smells (e.g., [4], [39]); (2) the evolution of code smells in software systems (e.g., [40]–[43]) and their impact on software quality (e.g., [7], [43]–[46]); and (3) the relationship between code smells and software development activities (e.g., [46], [47]).

Our work in this paper, strongly related to the one of Amir Saboury et al. [1], falls into the second category. We aim to understand how code smells affect the fault-proneness of JavaScript systems. Li and Shatnawi [44] who investigated the relationships between code smells and the occurrence of errors in the code of three different versions of Eclipse reported that code smells are positively associated with higher error probability. In the same line of study, Khomh et al. [45] investigated the relationship between code smells and the change- and fault-proneness of 54 releases of four popular Java open source systems (ArgoUML, Eclipse, Mylyn and Rhino). They observed that classes with code smells tend to be more change- and fault-prone than other classes. Tufano et al. [43] investigated the evolution of code smells in 200 open source Java systems from Android, Apache, and Eclipse ecosystems and found that code smells are often introduced in the code at the beginning of the projects, by both newcomers and experienced developers. Sjoberg et al. [47], who investigated the relationship between code smells and maintenance effort reported that code smells have a limited impact on maintenance effort. However, Abbes et al. [46] found that code smells can have a negative impact on code understandability. Recently, Fard et al. [4] have proposed a technique named JNOSE to detect 13 different types of code smells in JavaScript systems. The proposed technique combines static and dynamic analysis. They applied JNOSE on 11 client-web applications and found "lazy object" and "long method/function" to be the most frequent code smells in the systems. WebScent [48] is another tool that can detect client-side smells. It identifies mixing of HTML, CSS, and JavaScript, duplicate code in JavaScript, and HTML syntax errors. ESLint [12], JSLint [49] and JSHint [50] are rule based static code analysis tools that can validate source codes against a set of best coding practices. Despite this interest in JavaScript code smells and the growing popularity of JavaScript systems, to the best of our knowledge, there is no study that examined the effect of code smells on the fault-proneness of JavaScript server-side projects. This paper aims to fill this gap.

[25]https://github.com/DavidJohannesWall/smells_project

VII. Conclusion

In this study, we examine the impact of code smells on the fault-proneness of JavaScript systems. Also, we present a survival study of the smells of JavaScript systems. We present a quantitative study of fifteen JavaScript systems that compare the time until a fault occurrence in JavaScript files that contain code smells and files without code smells, with two different approaches: line grain, and line grain including dependencies approaches. This quantitative study also present some descriptive statistics about the twelve studied smells, as well as their survival by computing their lifetime. Results show that JavaScript files without code smells have hazard rates 20% lower than JavaScript files with code smells in the line grain study, and 38% lower than JavaScript files with code smells in the line grain including dependencies study. In other terms, the survival of JavaScript files against the occurrence of faults increases with time if the files do not contain code smells. We further investigated hazard rates associated with different types of code smells and found that "Variable Re-assign", "Assignment in Conditional Statements", and "Complex Code" smells have the highest hazard rates. The survival results show us that smells are introduced at the JavaScript files creation most of the time, and a big part of them still survived presently; those smells, and particularly "Variable Re-assign" which is the most proliferated into the studied systems, have a high chance of surviving a very long time. JavaScript developers should consider removing *Variable Re-assign* code smells from their systems in priority since this code smell is consistently associated with a high risk of fault, and because it is the most sizable code smell with a high chance of surviving over time. They should also prioritize *Assignment in Conditional Statements*, *Complex Code*, *This Assign*, *Nested Callbacks*, and *Long Parameter List* code smells for refactoring.

References

[1] A. Saboury, P. Musavi, F. Khomh, and G. Antoniol, "An empirical study of code smells in javascript projects," in *Software Analysis, Evolution and Reengineering (SANER), 2017 IEEE 24th International Conference on*. IEEE, 2017, pp. 294–305.

[2] Stackoverflow, "Developer survey results 2016," 2016, [Online; accessed 11-August-2016]. [Online]. Available: http://stackoverflow.com/research/developer-survey-2016

[3] Githut, "Discover languages in github," 2016, [Online; accessed 11-August-2016]. [Online]. Available: http://githut.info/

[4] A. M. Fard and A. Mesbah, "Jsnose: Detecting javascript code smells," in *Source Code Analysis and Manipulation (SCAM), 2013 IEEE 13th International Working Conference on*. IEEE, 2013, pp. 116–125.

[5] M. Fowler, "Refactoring: Improving the design of existing code," in *11th European Conference. Jyväskylä, Finland*, 1997.

[6] F. Khomh, M. D. Penta, Y.-G. Guéhéneuc, and G. Antoniol, "An exploratory study of the impact of antipatterns on class change- and fault-proneness," *Empirical Software Engineering*, vol. 17, no. 3, pp. 243–275, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10664-011-9171-y

[7] F. Jaafar, Y.-G. Guéhéneuc, S. Hamel, and F. Khomh, "Mining the relationship between anti-patterns dependencies and fault-proneness." in *WCRE*, 2013, pp. 351–360.

[8] "npm-coding-style," 2016, [Online; accessed 17-October-2016]. [Online]. Available: https://docs.npmjs.com/misc/coding-style

[9] "Node.js style guide," 2016, [Online; accessed 17-October-2016]. [Online]. Available: https://github.com/felixge/node-style-guide

[10] "Airbnb javascript style guide," 2016, [Online; accessed 17-October-2016]. [Online]. Available: https://github.com/airbnb/javascript

[11] "jquery javascript style guide," 2016, [Online; accessed 17-October-2016]. [Online]. Available: https://contribute.jquery.org/style-guide/js/

[12] "Eslint: The pluggable linting utility for javascript and jsx. http://eslint.org/."

[13] E. Brodu, S. Frénot, and F. Oblé, "Toward automatic update from callbacks to promises," in *Proceedings of the 1st Workshop on All-Web Real-Time Systems*. ACM, 2015, p. 1.

[14] K. Gallaba, A. Mesbah, and I. Beschastnikh, "Don't call us, we'll call you: Characterizing callbacks in javascript," in *2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 2015, pp. 1–10.

[15] T. J. McCabe, "A complexity measure," *IEEE Transactions on software Engineering*, no. 4, pp. 308–320, 1976.

[16] R. Marinescu and M. Lanza, "Object-oriented metrics in practice," 2006.

[17] F. A. Fontana, P. Braione, and M. Zanoni, "Automatic detection of bad smells in code: An experimental assessment." *Journal of Object Technology*, vol. 11, no. 2, pp. 5–1, 2012.

[18] A. Mardan, *Express. js Guide: The Comprehensive Book on Express. js*. Azat Mardan, 2014.

[19] "About bower," 2016, [Online; accessed 4-October-2016]. [Online]. Available: https://bower.io/docs/about/

[20] "Who uses grunt," 2016, [Online; accessed 4-October-2016]. [Online]. Available: http://gruntjs.com/who-uses-grunt

[21] J. Śliwerski, T. Zimmermann, and A. Zeller, "When do changes induce fixes?" in *ACM sigsoft software engineering notes*, vol. 30, no. 4. ACM, 2005, pp. 1–5.

[22] M. Fischer, M. Pinzger, and H. Gall, "Populating a release history database from version control and bug tracking systems," in *Software Maintenance, 2003. ICSM 2003. Proceedings. International Conference on*. IEEE, 2003, pp. 23–32.

[23] I. Neamtiu, J. S. Foster, and M. Hicks, "Understanding source code evolution using abstract syntax tree matching," *ACM SIGSOFT Software Engineering Notes*, vol. 30, no. 4, pp. 1–5, 2005.

[24] I. D. Baxter, A. Yahin, L. Moura, M. Sant'Anna, and L. Bier, "Clone detection using abstract syntax trees," in *Software Maintenance, 1998. Proceedings., International Conference on*. IEEE, 1998, pp. 368–377.

[25] F. Pfenning and C. Elliott, "Higher-order abstract syntax," in *ACM SIGPLAN Notices*, vol. 23, no. 7. ACM, 1988, pp. 199–208.

[26] R. Marinescu, "Detection strategies: Metrics-based rules for detecting design flaws," in *Software Maintenance, 2004. Proceedings. 20th IEEE International Conference on*. IEEE, 2004, pp. 350–359.

[27] D. Mazinanian and N. Tsantalis, "Migrating cascading style sheets to preprocessors by introducing mixins," in *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering*. ACM, 2016, pp. 672–683.

[28] J. Fox and S. Weisberg, *An R companion to applied regression*. Sage, 2010.

[29] A. G. Koru, K. El Emam, D. Zhang, H. Liu, and D. Mathew, "Theory of relative defect proneness," *Empirical Software Engineering*, vol. 13, no. 5, pp. 473–498, 2008.

[30] J. D. Singer and J. B. Willett, *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press, 2003.

[31] G. M. Selim, L. Barbour, W. Shang, B. Adams, A. E. Hassan, and Y. Zou, "Studying the impact of clones on software defects," in *2010 17th Working Conference on Reverse Engineering*. IEEE, 2010, pp. 13–21.

[32] H. Westergaard, *Contributions to the History of Statistics*. P.S. King, London, 1932.

[33] A. G. Koru, D. Zhang, and H. Liu, "Modeling the effect of size on defect proneness for open-source software," in *Proceedings of the Third International Workshop on Predictor Models in Software Engineering*. IEEE Computer Society, 2007, p. 10.

[34] T. M. Therneau and P. M. Grambsch, *Modeling survival data: extending the Cox model*. Springer Science & Business Media, 2000.

[35] T. Therneau, "R survival package," 2000.

[36] R. K. Yin, *Case Study Research: Design and Methods - Third Edition*, 3rd ed. SAGE Publications, 2002.

[37] E. Shihab, A. Ihara, Y. Kamei, W. M. Ibrahim, M. Ohira, B. Adams, A. E. Hassan, and K.-i. Matsumoto, "Studying re-opened bugs in open source software," *Empirical Software Engineering*, vol. 18, no. 5, pp. 1005–1042, 2013.

[38] E. Gamma, R. Helm, R.Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object Oriented Software*, 1995.

[39] F. Khomh, S. Vaucher, Y.-G. Guéhéneuc, and H. Sahraoui, "Bdtex: A gqm-based bayesian approach for the detection of antipatterns," *J. Syst. Softw.*, vol. 84, no. 4, pp. 559–572, Apr. 2011.

[40] A. Chatzigeorgiou and A. Manakos, "Investigating the evolution of bad smells in object-oriented code," in *Quality of Information and Communications Technology (QUATIC), 2010 7th Int'l Conf. on the*. IEEE, 2010, pp. 106–115.

[41] S. Olbrich, D. S. Cruzes, V. Basili, and N. Zazworka, "The evolution and impact of code smells: A case study of two open source systems," in *3rd Int'l Symposium on Empirical Software Engineering and Measurement, ESEM 2009,*, 2009, pp. 390–400.

[42] R. Peters and A. Zaidman, "Evaluating the lifespan of code smells using software repository mining," in *Software Maintenance and Reengineering (CSMR), 2012 16th European Conf. on*. IEEE, 2012, pp. 411–416.

[43] M. Tufano, F. Palomba, G. Bavota, R. Oliveto, M. Di Penta, A. De Lucia, and D. Poshyvanyk, "When and why your code starts to smell bad," in *Proceedings of the 37th International Conference on Software Engineering-Volume 1*. IEEE Press, 2015, pp. 403–414.

[44] R. Shatnawi and W. Li, "An investigation of bad smells in object-oriented design," in *Information Technology: New Generations, 2006. ITNG 2006. 3rd Int'l Conf. on*. IEEE, 2006, pp. 161–165.

[45] F. Khomh, M. Di Penta, Y.-G. Guéhéneuc, and G. Antoniol, "An exploratory study of the impact of antipatterns on class change-and fault-proneness," *Empirical Software Engineering*, vol. 17, no. 3, pp. 243–275, 2012.

[46] M. Abbes, F. Khomh, Y.-G. Gueheneuc, and G. Antoniol, "An empirical study of the impact of two antipatterns, blob and spaghetti code, on program comprehension," in *Software Maintenance and Reengineering (CSMR), 2011 15th European Conf. on*, March 2011, pp. 181–190.

[47] D. I. K. Sjoberg, A. Yamashita, B. Anda, A. Mockus, and T. Dyba, "Quantifying the effect of code smells on maintenance effort," *IEEE Trans. Softw. Eng.*, vol. 39, no. 8, pp. 1144–1156, Aug. 2013.

[48] H. V. Nguyen, H. A. Nguyen, T. T. Nguyen, A. T. Nguyen, and T. N. Nguyen, "Detection of embedded code smells in dynamic web applications," in *Automated Software Engineering (ASE), 2012 Proceedings of the 27th IEEE/ACM International Conference on*. IEEE, 2012, pp. 282–285.

[49] "Jslint: The javascript code quality tool. http://www.jslint.com."

[50] "Jshint: A static code analysis tool for javascript. http://jshint.com/."