Towards Building Knowledge on Causes of Critical Requirements Engineering Problems

Marcos Kalinowski  
UFF  
Niterói, Brazil  
kalinowski@ic.uff.br

Rodrigo Oliveira Spínola  
UNIFACS/Fraunhofer  
Salvador, Brazil  
rodrigo.spinola@pro.unifacs.br

Tayana Conte  
UFAM  
Manaus, Brazil  
tayana@icomp.ufam.edu.br

Rafael Prikladnicki  
PUC-RS  
Porto Alegre, Brazil  
rafael.prikladnicki@pucrs.br

Daniel Méndez Fernández  
Technische Universität München  
München, Germany  
daniel.mendez@tum.de

Stefan Wagner  
University of Stuttgart  
Stuttgart, Germany  
stefan.wagner@informatik.uni-stuttgart.de

Abstract—[Context] Many software projects fail due to problems in requirements engineering (RE). [Objective] The goal of this paper is to gather information on relevant RE problems and to represent knowledge on their most common causes. [Method] We replicated a global family of RE surveys in Brazil and used the data to identify critical RE problems and to build probabilistic cause-effect diagrams to represent knowledge on their common causes. [Results] The survey was answered by 74 different organizations, including small, medium and very large sized companies, conducting both, plan-driven and agile development. The most critical RE problems, according to those organizations, are related to communication and to incomplete or underspecified requirements. We provide the full probabilistic cause-effect diagrams with knowledge on common causes of the most critical identified RE problems online. [Conclusion] We believe that the knowledge presented in the diagrams can be helpful to support organizations in conducting causal analysis sessions by providing an initial understanding on what usually causes critical RE problems.

Keywords—Survey; NaPiRE; Knowledge Building; Requirements Engineering; Problems; Causes; Causal Analysis.

I. INTRODUCTION

The importance of high-quality requirements engineering (RE) has been widely accepted and well documented. Pfleeger [1] states that efficient RE is one of the main factors to avoid software project failure. RE constitutes a holistic key to successful development projects [2]. However, industry is still struggling to apply high-quality RE practices [3] and getting a further understanding on common RE problems and their causes is of great interest to both industry and academy. Therefore, many researchers have addressed identifying and analyzing RE problems faced by industry [4][5].

More recently, a project called NaPiRE (Naming the Pain in Requirements Engineering) comprises the design of a family of surveys on RE practice and problems, and it is conducted in joint collaboration with various researchers from different countries [6][7]. The goal of this project is to lay an empirical foundation about the state of the practice in RE to allow steering future research in a problem-driven manner [6]. Currently, the NaPiRE survey is being conducted in several countries around the globe.

Conducting causal analysis sessions [8] is an efficient means for organizations to improve their practice to overcome problems faced during software development. In these sessions, the causes of problems are identified and addressed to prevent their recurrence in future projects.

Experience reports on conducting causal analysis sessions on RE problems can be found in [9], [10] and [11]. One of the main difficulties reported during those sessions concerns the absence of a starting point for identifying potential causes. An initial solution concept to address this problem has been proposed in [12], where an approach for integrating knowledge of successive causal analysis sessions is described. This approach introduced the concept of a probabilistic cause-effect diagram, and of using such diagrams to present accumulated knowledge on the probabilities of causes based on the organization’s prior causal analysis experiences on similar problems.

However, although this approach and the probabilistic cause-effect diagram showed to be useful to support causal analysis sessions in a proof of concept [13], an experimental study [14] and an industrial experience [9], the knowledge depicted in the diagram has to be generated based on intra-company data from previous causal analysis sessions. Thus, it has to be built gradually and from scratch for each context, as there is no general documented and empirically grounded knowledge causes of critical problems that could be used as a starting point.

In this paper, we aim at gathering information on relevant RE problems and to represent knowledge on their most common causes as reported by the industry. Therefore, we replicated the NaPiRE survey in Brazil. We got answers from 74 different Brazilian organizations, spread across the country. We then used the data to identify the reportedly most critical RE problems and organized knowledge on their common causes by building
probabilistic cause-effect diagrams for those RE problems and making them available online\(^1\). Therefore, we enable organizations to use the knowledge presented in these diagrams as a starting point when conducting causal analysis by providing a further understanding on common causes of RE problems.

As an initial evaluation, we interviewed an industry representative of a Brazilian CMMI-Dev level 3 company who is currently implementing causal analysis practices, showing her the probabilistic cause-effect diagrams. The feedback was positive and future work includes conducting a case study of using those diagrams in industry while conducting causal analysis sessions on RE problems.

The remainder of this paper is organized as follows. Section II describes related work. Section III describes the NaPiRE project and its replication in Brazil. Section IV presents the survey results on the most critical RE problems and the probabilistic cause-effect diagrams with knowledge on common causes of those problems. Section V discusses the obtained results and their limitations in the light of the diagrams and of an informal interview conducted with an industry representative. Section VI presents the concluding remarks and future work.

II. RELATED WORK

In this paper, we aim at identifying relevant RE problems and building knowledge on their common causes, by replicating a survey. We propose representing such knowledge using probabilistic cause-effect diagrams. The following subsections provide the related work on survey research on RE problems and on probabilistic cause-effect diagrams.

A. Survey Research on RE Problems

Well-known surveys on causes for project failure include the Chaos Report of the Standish Group on cross-company root causes for project failures. While most of these causes are related to RE, the survey has serious design flaws and the validity of its results is questionable [15]. Moreover, it exclusively investigated failed projects and general causes at the level of overall software processes. Thus, it does not directly support the investigation of RE problems in industry.

Some surveys have been focusing specifically on RE problems in industry. These surveys include the one conducted by Hall et al. [4] in twelve software organizations. Their findings, among others, suggest that most RE problems are organizational rather than technical.

Some country-specific investigations of RE problems include the surveys conducted by Solemon et al. [16] and Liu et al. [17], with Malaysian and Chinese organizations, respectively, Khankaew and Riddle [5], more recently conducted semi-structured interviews with organizations from Thailand. In the first results of the NaPiRE survey, the reported RE problems were mainly identified from German companies [6].

These investigations provide valuable insights into industrial environments. However, as each of them focuses on specific aspects in RE or on specific countries, their results are isolated and not generalizable. To address this issue, the NaPiRE survey was designed in a joint collaboration as a continuous research project with researchers from different countries [6]. The design and interpretation of the results are aligned to a theory [6]. The survey, being replicated in different countries, shall contribute to an empirical basis to allow generalizable and problem-driven research in RE [6].

Given this context, to gather data concerning RE problems, the Brazilian authors of this paper decided to join the NaPiRE team and to replicate this survey in Brazil. To facilitate the use of this knowledge on common causes of RE problems it was organized into probabilistic cause-effect diagrams [12]. More details on these diagrams follow.

B. Probabilistic Cause-Effect Diagrams

Probabilistic cause-effect diagrams were introduced in [12] to provide visual support in causal analysis sessions with knowledge on common causes of problems gathered from previous experiences. They have shown to be a useful instrument in a proof of concept [13], an experimental study [14], and an industrial experience [9].

An example of such a diagram, taken from the experience reported in [9] is shown in Figure 1. The diagram extends the traditional cause-effect diagram [18] by (a) showing the probabilities for each possible cause to lead to the analyzed problem, and (b) representing the causes using grey tones, where causes with higher probability are shown closer to the center and in darker tones. Following the suggestion of guidelines for conducting causal analysis [8], it organizes the causes of problems into five categories: Input, Method, Organization, People, and Tool. The probabilities shown in Figure 1 were calculated with data on causes gathered in successive causal analysis sessions conducted in earlier iterations of the project. Causes that happened more frequently have higher probabilities.

This representation can be easily interpreted by causal analysis teams and highlights causes with greater probabilities of creating the analyzed problem. It allows the teams to efficiently answer questions during causal analysis sessions, such as: “Given similar past projects within my organizational context, with which probability does a certain cause lead to a specific problem?”. During the causal analysis sessions, the team can use the probabilistic cause-effect diagram, together with data on the problem, as input to help building a new cause-effect diagram with the causes identified in the current session. The newly identified causes can then be used to update the probabilities for

\(^1\) http://www.ic.uff.br/~kalinowski/seke15
the next session. This support has shown to be useful to support efficient cause identification [9][14].

![Figure 1. A probabilistic cause-effect diagram based on intra-company industry data for incorrect facts in functional specifications, taken from [9].](image)

However, the main shortcoming of using these diagrams is that the knowledge on causes of problems is generated intra-company and has to be built gradually. We believe that cross-company data taken from a sufficiently wide range of data from industry, with knowledge on causes of problems provides useful initial input when analyzing those problems. The problems have to be calibrated later with intra-company data containing the specific causes identified in new causal analysis sessions.

To build the necessary inter-company knowledge on common causes of critical RE problems, initially based on data from Brazilian companies, we replicated the NaPiRE survey in Brazil. Information on the NaPiRE project and on the conducted replication in Brazil is described next.

III. NAPIRE BRAZIL

A. The NaPiRE Project

The NaPiRE project resulted in the design of a global family of surveys to overcome the problem of isolated investigations in RE that are not representative [6]. Thus, a long-term goal of the project is to establish an empirically sound basis for understanding trends and problems in RE [7].

The design of the survey and its instruments have been extensively reviewed by several researchers [6]. In summary, the NaPiRE survey contains 35 questions gathering the following type of data from the responding organizations: (a) general information, (b) RE status quo, (c) RE improvement status quo, (d) RE problems faced in practice, and (e) RE problem manifestation (e.g., causes, impact).

The family of surveys is currently being conducted in several countries. Further information on the project, including the countries in which the survey is being replicated and a sample of the questionnaire can be found online[2]. Concerning its results, so far initial results from Germany have already been published [6].

B. NaPiRE Survey Replication in Brazil

When we decided to replicate the NaPiRE survey in Brazil, it was already designed and all the instruments were available. Therefore, in this section we focus on the details of how we planned and operated the replication in Brazil. Further information on the design of the surveys can be found in [6].

To plan the survey replication in Brazil, we held a couple of meetings with the NaPiRE general organizers[2]. During these meetings, the online environment (EFS survey tool[3]) was presented and some general guidelines for conducting the survey were provided. We decided to translate all instruments to Portuguese, the participants' native language.

Given the geographic dimensions of Brazil, to reach organizations from different regions and to gather representative data, the first author assembled a team of industry-focused researchers spread across the country. The strategy consisted of having researchers from the four main industry intensive regions of the country involved. The resulting NaPiRE Brazil team[2] comprises a researcher from the South of the country, one from the Southeast, one from the North and one from the Northeast.

Additionally, we contacted Softex, the organization responsible for the most widely adopted software reference model in Brazil, the MPS-SW, with over 600 assessments in all Brazilian regions [19]. They promptly trusted us contacts of 254 organizations with currently valid MPS-SW assessments so that they could be invited to take part in the survey.

Including a set of 80 additional relevant industry contacts from the authors (20 contacts per author on average), we created a list with contacts of representatives from 334 software organizations. We believe this set to be representative for the Brazilian software industry. Given the size of this industry (thousands of software organizations [20]), an extensive survey to reach all of them would be almost impossible.

We then configured the environment and sent the invitations with a link and password to the online survey to the list of contacts by e-mail. The survey was sent in December 2014, with reminders in January 2015 and February 2015. In total, 118 of the 334 invited organization representatives logged in to answer the survey. Out of these, 74 representatives answered the questionnaire completely (9 only read the initial instructions, 18 dropped at the first page of the questionnaire, and 17 dropped the survey in the middle). The median time to answer the survey completely was 29 minutes.

---

2 www.re-survey.org
3 www.unipark.com/en
IV. TOWARDS BUILDING KNOWLEDGE ON CAUSES OF CRITICAL RE PROBLEMS

In this section, we provide the initial survey results concerning the identified critical RE problems and their common causes as reported by industry. We also explain how the gathered information was organized into probabilistic cause-effect diagrams to provide a further understanding on common causes of RE problems. We start by presenting the characterization of the responding organizations as this information is crucial to enable a correct interpretation of the results.

A. Characterization of the Responding Organizations

To provide a summary of the characterization of the responding organizations, we will present information on their size and the used process models and RE standards. We will also present the roles of the participants within the organizations and their experience in this role.

Concerning size, in Table I we can observe that the survey included both extremes, small and medium-sized and very large-sized organizations.

<table>
<thead>
<tr>
<th>Size*</th>
<th>No. of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 Employees</td>
<td>11 (15.49%)</td>
</tr>
<tr>
<td>11-50 Employees</td>
<td>15 (21.13%)</td>
</tr>
<tr>
<td>51-250 Employees</td>
<td>17 (23.94%)</td>
</tr>
<tr>
<td>251-500 Employees</td>
<td>5 (7.04%)</td>
</tr>
<tr>
<td>501-1000 Employees</td>
<td>3 (4.23%)</td>
</tr>
<tr>
<td>1001-2000 Employees</td>
<td>5 (7.04%)</td>
</tr>
<tr>
<td>More than 2000</td>
<td>15 (21.13%)</td>
</tr>
<tr>
<td>Invalid (missing)</td>
<td>3 (N/A)</td>
</tr>
</tbody>
</table>

* Size including software and other areas.

Regarding the process model, Table II shows that most of the surveyed organization adopt agile (mainly Scrum-based) process models, followed by iterative and incremental process models and the traditional waterfall model. It is noteworthy that some organizations informed to use more than one process model to handle different types of projects. One explanation for changing process models is that organizations might have to follow a waterfall model during a bidding procedure while adopting scrum once the project is formally assigned.

<table>
<thead>
<tr>
<th>Process Model</th>
<th>No. of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scrum</td>
<td>45 (60.81%)</td>
</tr>
<tr>
<td>Waterfall</td>
<td>22 (29.73%)</td>
</tr>
<tr>
<td>Rational Unified Process (RUP)</td>
<td>19 (25.68%)</td>
</tr>
<tr>
<td>Extreme Programming (XP)</td>
<td>7 (9.46%)</td>
</tr>
<tr>
<td>V Model</td>
<td>4 (5.41%)</td>
</tr>
<tr>
<td>Others*</td>
<td>11 (14.86%)</td>
</tr>
</tbody>
</table>

* Others includes self-adapted process models (4), other iterative and incremental development process models (4) and other process models based on agile methods (3).

In Table III, we can observe that most of the surveyed organizations follow reference-model-based standards, such as MPS-SW and CMMI-Dev. This, of course, may have been influenced by the strategy of also distributing the survey to the organizations with valid MPS-SW assessments. Nevertheless, many organizations answered that they follow the standards of the adopted development process and their own standards.

<table>
<thead>
<tr>
<th>RE Standard (or Reference Model)</th>
<th>No. of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW reference model (e.g., CMMI-Dev, MPS-SW)</td>
<td>39 (52.70%)</td>
</tr>
<tr>
<td>Adopted development process (e.g., RUP, Scrum)</td>
<td>25 (33.78%)</td>
</tr>
<tr>
<td>Self-defined (including a process with deliverables, milestones and phases)</td>
<td>19 (25.68%)</td>
</tr>
<tr>
<td>Self-defined (including artefacts and templates)</td>
<td>18 (24.32%)</td>
</tr>
<tr>
<td>None</td>
<td>1 (1.35%)</td>
</tr>
</tbody>
</table>

To characterize the participants, their roles in the organization are shown in Table IV and their experience in these roles is shown in Table V. It can be seen that participants are mainly project managers and highly experienced.

<table>
<thead>
<tr>
<th>Role</th>
<th>No. of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Manager</td>
<td>32 (45.07%)</td>
</tr>
<tr>
<td>Business Analyst</td>
<td>8 (11.27%)</td>
</tr>
<tr>
<td>Developer</td>
<td>4 (5.63%)</td>
</tr>
<tr>
<td>Software Architect</td>
<td>4 (5.63%)</td>
</tr>
<tr>
<td>Test Manager / Tester</td>
<td>3 (4.23%)</td>
</tr>
<tr>
<td>Requirements Engineer</td>
<td>2 (2.82%)</td>
</tr>
<tr>
<td>Others*</td>
<td>18 (25.35%)</td>
</tr>
<tr>
<td>Invalid (missing)</td>
<td>3 (NA)</td>
</tr>
</tbody>
</table>

* Other informed values include development directors, program managers and portfolio managers (7), quality assurance analysts (7), and people from the software engineering process group (4).

<table>
<thead>
<tr>
<th>Experience</th>
<th>No. of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialist (more than 3 years)</td>
<td>52 (73.24%)</td>
</tr>
<tr>
<td>Experienced (1 to 3 years)</td>
<td>15 (21.13%)</td>
</tr>
<tr>
<td>Newbie (up to 1 year)</td>
<td>04 (5.63%)</td>
</tr>
</tbody>
</table>

While we had no control over which organizations and representatives would answer the survey, we were happy to obtain such a representative characterization, including small, medium and very large-sized organizations enrolled in both, plan-driven and agile development methods, and to have our answers provided mainly by highly experienced professionals.

B. Critical RE Problems

Based on a set of 21 precompiled general RE problems listed in the NaPiRE questionnaire [6], the participants were asked to rank the five most critical ones.
The most critical RE problems, as ranked by the survey participants, are shown in Table VI. This table shows the 8 RE problems that were cited between the five most critical ones by more than 20% of the respondents. The table also shows how often each of these problems was ranked as being the most critical problem of all. We observe that communication problems were often cited (problems #1 and #4), as well as incomplete and underspecified requirements (problems #2 and #3).

<table>
<thead>
<tr>
<th>#</th>
<th>RE Problems</th>
<th>Cited*</th>
<th>Ranked #1*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Communication flaws between the project team and the customer</td>
<td>32 (43.24%)</td>
<td>9 (12.16%)</td>
</tr>
<tr>
<td>2</td>
<td>Incomplete and/or hidden requirements</td>
<td>31 (41.89%)</td>
<td>12 (16.22%)</td>
</tr>
<tr>
<td>3</td>
<td>Underspecified requirements that are too abstract and allow for various interpretations</td>
<td>31 (41.89%)</td>
<td>3 (4.05%)</td>
</tr>
<tr>
<td>4</td>
<td>Communication flaws within the project team</td>
<td>26 (35.14%)</td>
<td>5 (6.67%)</td>
</tr>
<tr>
<td>5</td>
<td>Insufficient support by customer</td>
<td>21 (28.38%)</td>
<td>5 (6.76%)</td>
</tr>
<tr>
<td>6</td>
<td>Inconsistent requirements</td>
<td>18 (24.32%)</td>
<td>2 (2.70%)</td>
</tr>
<tr>
<td>7</td>
<td>Time boxing / Not enough time in general</td>
<td>17 (22.97%)</td>
<td>1 (1.35%)</td>
</tr>
<tr>
<td>8</td>
<td>Moving targets (changing goals, business processes and/or req.)</td>
<td>15 (20.27%)</td>
<td>5 (6.67%)</td>
</tr>
</tbody>
</table>

* The probabilities were calculated based on the overall amount of 74 participants although some of them (9) did not inform any of the problems. We decided to keep the total amount as basis because we were not sure if they did not find the problems relevant or if they did not want to think about it.

C. Causes of Critical RE Problems

After selecting the five most critical RE problems, we asked our respondents to provide what they believe of being the main causes for each of the problems. They provided the causes in an open question format, with one open question for each of the previously selected RE problems.

We analyzed the provided qualitative data and aggregated similar causes even when textual descriptions differed; always counting the number of times each cause was reported for a given problem. Therefore, we used the constant comparative method [21] to compare each textual cause description against our already catalogued list of causes.

Thereafter, to build the probabilistic cause-effect diagrams for each RE problem, we categorized the causes as suggested in [8] and generated probabilities based on frequency counting. Figure 2 shows the probabilistic cause-effect diagram for the problem ‘Incomplete and/or hidden requirements’ (#2 in Table VI). The 31 organizations that ranked this problem among the most critical provided 34 instances of causes for it, which could be mapped to a list of 20 catalogued causes with different frequencies that were used as input to generate the diagram. It can be seen that, according to the survey participants, the most common causes for this problem are related to the people category and the lack of skills in RE (29.41%), although the method and input categories each were responsible for more than 20% of the causes informed for this problem.

Due to space limitations, the remaining probabilistic cause-effect diagrams for the five top ranked RE problems are available online1, where the user can select the problem and then look at the respective diagram to obtain a further understanding on common causes.

V. DISCUSSION

Given our previous causal analysis experiences, the wide range of organizations that participated in the survey and their representative characterization (including, for instance, differently sized plan-driven and agile-oriented organizations), we believe that the resulting probabilistic cause-effect diagrams generated from cross-company data should provide useful additional input into causal analysis sessions for companies working in same or similar project settings. This is especially true for organizations that do not have any accumulated knowledge on common causes of RE problems.

It is noteworthy that organization may also need to calibrate the probabilities of the diagrams with the causes identified in their own causal analysis sessions, by using an approach similar to the one detailed in the experience described in [9]. Still, we believe that the knowledge on common causes generated as a contribution of this paper constitutes a useful starting point.

To provide some preliminary support for these claims, we interviewed an industry representative of the software engineering process group of a Brazilian CMMI-Dev level 3 company that is currently implementing causal analysis practices. We showed her the probabilistic cause-effect diagrams and she found the contained knowledge useful and was promptly willing to use the diagrams to support causal analysis sessions in her environment. This strengthens our confidence in the suitability of our results to establish an intra-company knowledge base on common RE problems and their causes.
VI. CONCLUDING REMARKS

In this paper, we gathered data on critical RE problems and their common causes by replicating the NaPiRE survey in Brazil. We presented the results concerning the most critical RE problems and represented the knowledge on common causes of these problems by building probabilistic cause-effect diagrams.

The chosen dissemination strategy enabled us to get answers from a wide range (74) of Brazilian organizations. The characterization showed a large diversity of the responding organizations including differently sized plan-driven and agile-oriented organizations.

The survey results allowed us to identify the most critical RE problems according to the responding organizations (Table VI) and to observe that they are mainly related to communication problems and incomplete or underspecified requirements. In addition, the probabilistic cause-effect diagrams (see Figure 2) showed to be suitable for the presentation of knowledge on common causes of the RE problems in an easily understandable way. The probabilistic cause-effect diagrams for the RE problems identified as the five most critical ones are available online1. We believe that these diagrams provide useful input into causal analysis sessions at specific organizations, especially for organizations that do not have accumulated knowledge on common causes of their RE problems. Case studies on this matter form a high-priority scope of our future work.

Future work also comprises: (a) considering more NaPiRE data (from different countries) to build the knowledge on common causes based on a larger cross-company dataset, (b) further exploring other data gathered as part of the NaPiRE Brazil survey, such as RE problem mitigation actions, and (c) integrating the overall survey results into an empirical software engineering body of knowledge [22].

ACKNOWLEDGMENT

The authors would like to thank Softex and each of the 74 responding organizations. Thanks also to CNPq for financial support (project #460627/2014-7).

REFERENCES