Abstract—Software maintainability is a very important quality attribute. Its prediction for relational database-driven software applications can help organizations improve the maintainability of these applications. The research presented herein adopts a survey-based approach where a survey was conducted with 40 software professionals aimed at identifying and ranking the important maintainability predictors for relational database-driven software applications. The survey results were analyzed using frequency analysis, and results suggest that maintainability prediction for relational database-driven applications is not the same as that of traditional software applications. The results also provide a baseline for creating maintainability prediction models for relational database-driven software applications.

Software maintainability, relational database-driven software applications; survey; predictors; frequency analysis

I. INTRODUCTION

Software maintainability is the ease with which a software system or component can be modified to correct faults, improve performance or other attributes, or adapt to a changed environment [6]. It is inherently associated with the process of software maintenance which has long been known to have the major amount of software costs associated to it [12]. In order to manage these costs, it is important to understand, predict and improve software maintainability [12]. Software maintainability prediction involves proposing and validating predictors that have a bearing on software maintainability [1] and then employing these predictors to create software maintainability prediction models [13]. With the help of a maintainability prediction model, software organizations can better manage their maintenance resources and adopt a defensive design [11].

Database-driven applications have gained much popularity in modern software development [4] with relational databases as the most used and most successful type of database [15]. Relational databases are different from and more formalized than other types of databases and persistence mechanisms [2]. Database-driven applications consist of a database, a Database Management System (DBMS), and a set of applications that interact with the database through this management system [7]. Relational database-driven applications are therefore, those database-driven applications that have a relational database backend. The change in requirements cause these applications to undergo maintenance resulting in storing increased number of data sources and relationships; increased database complexity; and increased coupling between the database and application [9]. This suggests that the maintainability of these applications is also impacted by database specific factors in addition to application specific factors [13].

Given the importance of relational database-driven software applications in modern software development, it is important to investigate the factors that impact upon their maintainability in order to enable maintainability prediction. This paper is a step towards predicting maintainability of relational database-driven application by providing a validated list of predictors that have an impact on the maintainability of relational database-driven applications. This list of predictors was initially gathered with the help of twelve interviews conducted with software professionals [13]. The interviews’ analyses resulted in the identification of 120 predictors. The focus of the work presented herein is to rank these predictors, with respect to the strength of their impact on the maintainability of relational database-driven applications, via another survey conducted with a larger sample of software professionals. The main contributions of this paper are therefore to:

- Further validate the predictors of maintainability identified by Riaz et al. [13] within the context of relational database-driven software applications.
- Rank the maintainability predictors for relational database-driven software applications in terms of their relative importance to predicting maintainability.
- Establish whether the predictors presented by Riaz et al. [13] positively or negatively impact the maintainability of relational database-driven applications.
- Present evidence on the type of applications and DBMS used the most in practice.

The remainder of the paper is organized as follows. Section II gives an account of the related work. Section III details the research methodology. Section IV presents the results. A discussion on the results and threats to validity are given in Section V, followed by conclusions in Section VI.

II. RELATED WORK

The research presented herein is informed by the results of:
1) a Systematic Review (SR) conducted on the topic of ‘Software maintainability prediction and metrics’ [12]; and (2) twelve interviews conducted with software professionals [13]...
aimed at gathering evidence relating to the maintainability prediction of relational database-driven applications.

The results of the SR revealed very little evidence on maintainability prediction. The total number of studies selected in the SR was 15. These studies were further analyzed to assess if the datasets used in these studies completely or partially comprised relational database-driven software applications. Only three of these studies [3] [5] [8] had used relational database-driven software applications for the purpose of their research evaluation. However, only one of these studies [5] presented a maintainability prediction model but did not provide its prediction accuracy; the other two studies only presented measures [3] and factors [8] that impacted upon software maintainability, without any associated maintainability prediction model. Further analysis of these studies revealed that although they used relational database-driven applications, none of their proposed predictors or factors related specifically to a back-end database or to the interaction between a back-end database and the front-end application.

To further investigate maintainability prediction in the context of relational database-driven applications, interviews with software professionals were conducted to gather evidence from practice [13]. The results from the interviews revealed that a formal prediction model, or approach to predicting the maintainability of relational database-driven applications is not used in practice. The practitioners rely on expert judgement when assessing maintainability of relational database-driven applications. The analysis of these interviews also resulted in a list of factors that impact upon the maintainability of relational database-driven applications and may be used as their maintainability predictors.

### III. RESEARCH METHODOLOGY

#### A. Survey-based Questionnaire

In order to determine which maintainability predictors are the most relevant and stand out the most from the list of predictors compiled with the help of SR and interviews’ results, a survey was conducted with software practitioners. The list used in the survey comprised a total of 120 predictors [14], and the survey was informed by the results from the SR and the conducted interviews. It aimed at ranking these predictors in terms of their relative impact on the maintainability of relational database-driven software applications.

The survey questionnaire had 3 parts: The first part related to the respondents’ characteristics, such as job function at the company, highest qualification, and total experience in years; and companies’ characteristics, such as the year it was established, whether it is involved in developing relational database-driven applications, and the DBMS used most often.

The second part related to maintainability predictors. There were a total of 120 predictors, which belonged to one of the 10 categories below:

1. Database design
2. Relational database-driven applications
3. Design aspects of relational database-driven applications
4. Size, scope, and complexity of relational database-driven applications
5. Quality, quality assurance, and testing of relational database-driven applications
6. Change and maintenance of relational database-driven applications
7. Development and deployment environment of relational database-driven applications
8. Web applications with relational database back-end
9. Organizational culture and policies
10. Team communication and project management.

The possible responses against each of the factors, in relation to their impact on the maintainability of relational database-driven applications, were recorded on a 7-point bipolar scale. The responses could be one of ‘high decrease’, ‘medium decrease’, ‘low decrease’, ‘no impact’, ‘low increase’, ‘medium increase’, and ‘high increase’. The scale was designed such that both the direction of the factors’ relationship (positive or negative i.e., whether a given factors increase maintainability or decreases maintainability, respectively) and the strength of the relationship with maintainability could be determined.

The third part of the questionnaire recorded opinions on which aspect of a relational database-driven application, from interface, data model, or algorithm objects/computation intensive modules, was hardest to maintain; asked of any additional predictors of maintainability; and recorded information on how a judgment is made on the maintainability of relational database-driven applications in a company (for which the option supplied was the amount of time to carry out a maintenance task). Refer to Riaz [14] for the complete survey questionnaire.

The survey was tested with a pilot run of four practitioners. Based on the feedback from the pilot, the survey was improved where the predictors from the SR, especially those that corresponded to various software metrics, were described using a simpler language in order to be clearly understandable by the practitioners.

#### B. Survey process

In order to conduct the survey, approval from the University of Auckland Human Participants Ethics Committee was obtained (Ref. 2009/527). The process prescribed by UAHPEC was to contact the competent authorities within the companies which would then extend the invitation to their employees. The invitations were sent to more than 50 companies in New Zealand and more than 25 companies in Pakistan through email. The email addresses were obtained through companies’ listings on the Internet. In addition, a mailing list of the University of Auckland’s Center of Software Innovation and authors’ own contacts were also used to make contact with the companies. The email invitation explained the purpose and nature of the survey, provided the Web link to the survey, and asked the companies’ competent authorities to request participation by those employees within their companies that had experience with developing and/or maintaining relational database-driven software applications.
The survey and the relevant UAHPEC documents were made available online and were filled out via a Web interface [14]. Some respondents also downloaded the questionnaire and returned an e-copy. The survey was conducted between June 2010 and September 2010.

C. Data Source

The data source used herein comprised responses of a total of 40 software professionals - 13 from 8 different software companies in New Zealand and 27 from 15 different software companies in Pakistan, where one respondent was self-employed. The roles of the participants varied from ‘Software Developer’ to ‘Head of Development’. The participants had at least 2 years of experience with software development and/or maintenance, and had an average experience of 7.8 years with a minimum of 2 and maximum of 28 years of experience.

The two countries, Pakistan and New Zealand, were chosen for the following reasons: i) the interviews of which this research is informed were also conducted in New Zealand and Pakistan; ii) the results of the interviews established that the context of this research was not culture-sensitive as the focus was on software applications in general and not specific aspects related to people or their way of carrying out their work; and iii) the first author is originally from Pakistan and research is being carried out at the University of Auckland, New Zealand to which all the three authors are affiliated.

In terms of the participating companies, only 3 companies, from which only one respondent each participated, were created 1, 2 and 5 years ago. All the other companies were at least 10 years old. In addition, all companies except two were engaged in development or maintenance of applications where 90% or more of their applications had relational database backend. The other two also were involved in developing or maintaining relational database-driven applications; however, most of their work did not involve this type of development. In relation to the use of DBMS, all the participating companies used either Oracle or Microsoft SQL Server.

D. Data Analysis

Prior to analyzing the data, it was consolidated from two different sources, a database residing on Microsoft SQL Server used for the online survey and e-copies of the survey forms filled by some respondents for which the data was manually entered. Frequency analysis was the technique used for data analysis. It was well suited to the purpose as the intention was to see how many respondents thought of a predictor as having an impact 1 to 7 (high decrease to high increase) on the maintainability of relational database-driven software application.

During the frequency analysis, initially only the predictors that were reported to have an impact of ‘high decrease’ or ‘high increase’ on the maintainability of relational database-driven applications were considered. However, the frequency values did not stand out when only the extreme ends of the bi-polar scale were considered. For instance, for most of the factors the proportion of the respondents (out of 40) that chose one of the extreme ends of the scale (1 and 2 OR 6 and 7) were half or less than half i.e., between 15 and 20 (out of 40 respondents) and the proportion of the respondents that chose other options ‘medium decrease’ to ‘medium increase’ was equal or higher to those that selected one of the values at the extreme ends of the scale. Therefore, for final analysis, the sum of the frequencies for ‘high decrease’ and ‘medium decrease’ (1 and 2 on the bi-polar scale), and ‘medium increase’ and ‘high increase’ (6 and 7 on the bi-polar scale) were considered. The sum of the frequencies for the two ends of the bi-polar scale presented a better choice and represented better insights into the data.

IV. Results

A. Maintainability Predictors

The results of the survey data analysis corresponding to the second part of the survey are presented in Figure 1, and show only those maintainability predictors for relational database-driven applications that had frequency values of 21 or higher i.e., those predictors that were reported by more than 50% of the respondents as having a high impact on the maintainability of the mentioned applications. Note that, the predictors presented in the table are all those predictors that had an impact of ‘medium increase’ or ‘high increase’ on the maintainability of relational database-driven applications. None of the predictors that had an impact of ‘medium decrease’ or ‘high decrease’ on maintainability were selected as a result of the frequency analysis. This shows that all the predictors listed in Figure 1 have a positive relationship with the maintainability of relational database-driven software applications i.e., they all result in an increase in the maintainability of relational database-driven applications whenever they also increase.

The number of selected predictors for each category (as defined in the survey questionnaire and mentioned in Section III-A) against the total in that category are given in Table I. Note that predictors belonging to categories ‘Size, Scope, and Complexity of Relational Database-Driven Software Applications’ and ‘Web Applications with Relational Database Backend’ were not selected during frequency analysis and are therefore not presented in the results discussed further.

It is also worth noting that the best predictor relates to the relational database schema. In addition, other predictors, such as good database design, optimal use of DBMS features, are also among the top predictors of maintainability for relational database-driven software applications. Note that these predictors were not reported in the studies selected in the SR yet these rank higher than the predictors that were identified both by SR and interviews.
The percentages of predictors selected for each category presented in Table I.

TABLE I. PERCENTAGE OF MAINTAINABILITY PREDICTORS PER CATEGORY

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
<th>Total in Category</th>
<th>Selection Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Schema</td>
<td>4</td>
<td>19</td>
<td>21%</td>
</tr>
<tr>
<td>Relational Database-Driven Applications</td>
<td>7</td>
<td>15</td>
<td>47%</td>
</tr>
<tr>
<td>Design Aspects of Database-Driven Applications</td>
<td>4</td>
<td>9</td>
<td>44%</td>
</tr>
<tr>
<td>Quality, QA, and Testing</td>
<td>17</td>
<td>20</td>
<td>85%</td>
</tr>
<tr>
<td>Change and Maintenance</td>
<td>4</td>
<td>10</td>
<td>40%</td>
</tr>
<tr>
<td>Development and Deployment Environment</td>
<td>4</td>
<td>7</td>
<td>57%</td>
</tr>
<tr>
<td>Organizational Culture and Policies</td>
<td>2</td>
<td>4</td>
<td>50%</td>
</tr>
<tr>
<td>Team, Team Communication, and Project Management</td>
<td>2</td>
<td>7</td>
<td>29%</td>
</tr>
</tbody>
</table>

The results given in Table I show that the most important category of predictors was ‘Quality, Quality Assurance, and Testing’, both in terms of total number of predictors selected for overall analysis and number of predictors selected from the total number of predictors within the category. The second important category, perhaps not in terms of the percentage of factors selected from the category, but the number of factors contributing to maintainability prediction is ‘Relational Database-Driven Applications’. Other important categories relate to ‘Development and Deployment Environment’ and ‘Organizational Culture and Policies’. The results clearly indicate that the factors related to quality, database schema, relational database-driven application play an important role in predicting maintainability. However, the results also indicate that the factors related to development environment and organizational culture are also considered important by
practitioners in order to determine the maintainability of relational database-driven applications.

B. Ranking of the Aspect of Relational Database-Driven Applications for Ease of Maintenance

The first section of the third part of the survey asked the respondents to assign a ranking of 1(easiest) to 3(hardest) to each of three listed aspects of the relational database-driven applications i.e., application’s interface, application’s data model, and application’s algorithmic objects or computation intensive modules. The results (see Figure 3) show that the application’s interface was considered the easiest to maintain by 33 respondents, followed by the application’s data model (24 respondents) and algorithmic objects (23 respondents). As such, there was not too much difference between the number of respondents for second and third rankings, which suggest that the data model of a relational database-driven software application can be as difficult to maintain as algorithmic or computation intensive objects.

In addition to the ranking against the application’s aspects, this part of the survey also asked if the respondents wanted to report any other factors they believed were important for predicting maintainability of relational database-driven applications. Only three respondents added their responses, which included software development process level used in an organization; non-technical aspects related to customer behavior which drives the timelines for performing maintenance tasks; and conformance of DBMS to standards, reliability of WANs, and impact of other software installed on the system such as anti-virus, operating system etc. As such, these factors were not considered significant as part of the analysis as each of these were reported by only one respondent and would have ultimately been eliminated from the list of top maintainability predictors.

V. DISCUSSION

This paper presents the results of a survey conducted with software professionals. The survey was informed by a previously conducted SR [12] and the results of interviews conducted with software professionals [13]. It was aimed at identifying maintainability predictors that had the highest impact on the maintainability of relational database-driven software applications. A list of 120 predictors belonging to 10 different categories was used on which expert opinions from software professionals were recorded. The survey results suggest that the top 5 predictors related to 'Database Design', 'Design Aspects of Relational Database-Driven Applications', 'Team, Team Communication and Project Management', and 'Change and Maintenance'. The number of predictors selected by more than 50% respondents was highest for the category 'Quality, Quality Assurance, and Testing' followed by the predictors related to 'Relational Database-Driven Application'.

The main contribution of this paper is that it identifies the top predictors of maintainability for relational database-driven applications. Other contributions include establishing and supporting by evidence that the type of applications most developed in software organizations have a relational database backend, as discussed in Section III-C, interface of a relational database-driven application is considered easiest to change by software practitioners whereas data model and computation intensive modules are considered equally difficult to maintain in comparison to interface; and maintainability in practice is measured using expert judgment.

The results of the survey presented some interesting findings in terms of the factors considered important by software practitioners. From the results of the SR, the most frequently mentioned predictors related to coupling, complexity and size related measures. However, the survey results suggested that the factors related to ‘Size, Scope, and Complexity of Relational Database-Driven Applications’ were not ranked higher than predictors belonging to other categories by the surveyed software practitioners. In fact, the list of predictors ranked higher by more than 50% of the respondents does not include any factor from this category. This is a very important finding suggesting a huge gap between research and practice with regards to the perception of maintainability predictors in the context of relational database-driven applications. The possible reasons for this can be that most research experiments reported in literature are not performed in an industrial setting and the applications used in research experiments are generally small in scope. This makes it easier to gather metrics and experiment on these applications whereas
in practice there is lesser room for experimentation due to tight deadlines. In addition, in practice no sophisticated techniques for maintainability prediction are used and therefore, practitioners rely more on the visible factors on which they can more easily formulate a judgment on maintainability. Also, the sample size for this research is not large so there is also the possibility of lack of external validity. We believe that irrespective of the reasons, there is a need to further investigate maintainability predictors of for relational database-driven applications. In this regard, our work in progress involves gathering project specific data on the predictors identified herein and creating maintainability prediction models based on the gathered data. This, in addition to providing prediction models, will also be a means to validating the predictors and will suggest a need to consider predictors other than the ones selected herein.

The maintainability predictors for software applications suggested in the literature involved measures related to coupling, complexity and size of the application. The results of the survey suggest that size related measures are not important predictors whereas predictors related to database design are very important for maintainability prediction of relational database-driven applications. This suggests that the maintainability of relational database-driven applications is different from that of the software applications.

It may appear that the data analysis method used is not very sophisticated and rigorous. However, the purpose of this survey was to identify top predictors from the list of 120 predictors such that these could be used for future work involved in this research. The analysis technique used perfectly suited the purpose and the use for a more sophisticated technique was hence, not required. Moreover, there are examples in the literature on software prediction where frequency analysis has been used to rank the predictors to be used further for the creation of prediction models [10].

The possible limitations of this research are related to unequal number of respondents from the two countries and smaller sample size. While equal number of respondents from both countries would have provided better opportunity for data analysis and comparison of results, the research itself was context free as established by the results from the interviews [13]. Therefore, we believe that the unequal number of respondent in this case is not a threat to the validity of the research. In regards to small sample size, we believe that it poses threats to the external validity of the research as the chances of obtaining different results from a larger sample size cannot be ignored. However, considering the number of companies to which invitations to participate were emailed, the length of the survey, and the time frame within which the survey had to be completed; we believe that the sample data is of a considerable size to derive valuable conclusions.

VI. CONCLUSIONS

This paper presents the results of a survey conducted with software professionals. The survey was informed by the results of a previously conducted SR and interviews conducted with software professionals. The aim of the survey was to rank the 120 maintainability predictors of relational database-driven applications belonging to 10 different categories for their strength in predicting maintainability. Out of 120 total factors, 44 factors were found to have an impact of either 'high increase' or 'medium increase' on maintainability, as per more than 50% of the respondents. The most important category of predictors was 'Quality, Quality Assurance' and 'Testing' whereas the top predictor 'Correct definition of entities' belonged to the category 'Database Design'. Our results also suggest that the practitioners measure maintainability using expert judgment and consider computation intensive modules of relational database-driven applications as easy or difficult to maintain as the data models.

Future work involves creating maintainability prediction models for relational database-driven software applications. In this regard, the work on gathering project related data on the predictors is already in progress.

REFERENCES