

# Practical similarities and differences between Systematic Literature Reviews and Systematic Mappings: a tertiary study

Bianca M. Napoleão

Depart. of Computer

Federal Tech. Univ. of Paraná

Cornélio Procópio, PR – Brazil

[biancanapoleao@alunos.utfpr.edu.br](mailto:biancanapoleao@alunos.utfpr.edu.br)

Katia R. Felizardo, Érica F. de Souza

Depart. of Computer

Federal Tech. Univ. of Paraná

Cornélio Procópio, PR – Brazil

[katiascannavino,ericasouza@utfpr.edu.br](mailto:katiascannavino,ericasouza@utfpr.edu.br)

Nandamudi L. Vijaykumar

Lab. of Comp. and Applied Math.

Nat. Inst. for Space Research

São José dos Campos, SP – Brazil

[vijay.nl@inpe.br](mailto:vijay.nl@inpe.br)

**Abstract—Background:** Several researchers have reported their experiences in applying secondary studies in Software Engineering (SE), however, there is a lack of studies discussing the distinction between Systematic Mappings (SMs) and Systematic Literature Reviews (SLRs). **Aims:** The objective of this paper is to present the results of a tertiary study conducted to collect and evaluate evidence to better understand similarities and differences between SLRs and SMs related to four aspects: research question, search string, search strategy and quality assessment. **Method:** We identified 170 secondary studies that were reviewed to answer a set of Research Questions (RQ) related to the practical conduction of secondary studies in SE. **Results:** Results show that both SLRs and SMs have generic RQs, broad search strings, and adopt automatic search as search strategy. However, quality assessment has been more widely adopted in SLRs. **Conclusions:** In practice, only the quality assessment is conducted differently in SLRs and SMs.

**Keywords**—Systematic Literature Review, Systematic Mapping, Secondary Studies

## I. INTRODUCTION

Evidence-Based Software Engineering (EBSE) employs appropriate research methods to build a body of knowledge about when, how, and in what context methods or tools are more appropriate to be employed for Software Engineering (SE) practice. In 2004, EBSE was first introduced as a means to advance and improve the discipline of SE [1]. In this context, Systematic Literature Reviews (SLRs) [2] and Systematic Mappings (SMs) [3] have been providing methodological and structured processes to identify and aggregate research evidence. SLRs and SMs are said to be secondary studies [2]. They are studies that review primary studies related to some research questions (RQs) to integrate/synthesize evidence related to those RQs. A primary study is related to an individual publication or a study, such as an empirical study, which investigates a specific RQ. The advantages of performing secondary studies include, among others, reduced likelihood of bias in results and the potential ability to combine data from several studies. These benefits are some reasons why secondary studies have been gaining popularity over the years [4].

Kitchenham et al. [2] and Petersen and colleagues [3] recognize that currently, the distinction between SMs and conventional SLRs conduction is somewhat fuzzy. The main contribution of our paper is to establish a fair and practical discussion about the conduction of secondary studies in SE area. Several studies have reported experiences and lessons learned from researchers conducting secondary studies [3], [5], [6], [7]. However, these studies do not focus on the practical similarities and differences between secondary studies. Therefore, we performed a tertiary study in identifying practical similarities and differences between SLR and SM. The findings from this study are expected to contribute to the existing knowledge with respect to the conduction of secondary studies in SE. Some of the concrete doubts that we intend to clarify are: 1 – *What types of Research Questions (RQs) have been answered in SLRs and SMs studies? Are these RQs generic or specific?*; 2 – *How the search strings have been formulated?*; 3 – *Which search strategy has been adopted in SLRs and SMs studies?*; and 4 – *Has the quality of the studies been assessed?*

The remainder of this paper is organized as follows. Section II presents a brief overview of secondary studies. Section III discusses the research method applied to perform our tertiary study. Subsections III-A and III-B discusses the results, their implications, and limitations. Finally, Section IV concludes the paper and presents directions for future work.

## II. BACKGROUND

Secondary studies (SLR and SM) are means of finding, critically evaluating and interpreting available research papers (primary studies) on a particular set of RQs, topic area, or a phenomenon of interest. The method is intended to ensure that the literature review is unbiased, rigorous and auditable [8].

The SLR process comprises three main phases [2]: (i) *Planning*: refers to the pre-review activities, and establishes a review protocol, besides conducting a pilot test; (ii) *Conducting*: aims at searching and selecting the studies, in order to extract and synthesize data from them. This phase comprises the following activities: study searching, study selection, study quality assessment, and data extraction; (iii) *Reporting*: is the final phase, and aims at writing up the results and circulating them to potentially interested parties.

A SM is a more open form of SLR, providing an overview of a research area to assess the quantity of evidence existing

on a topic of interest [2]. In general, SM is conducted by a planning phase, which includes formulation of RQs and definition of inclusion and exclusion criteria, followed by search and screening of primary studies. The *data extraction* activity for SM is broader than the data extraction process for SLR and the analysis of a mapping does not include the use of depth analysis techniques, such as meta-analysis, but rather summaries. In a nutshell, SM deals with a broader research topic while SLR deals directly with a specific RQ narrowing the search for specific answers.

As can be observed in Table I, SLRs and SMs are different in various aspects, such as, goals, breadth, validity issues and implications. They can also vary in more basic aspects of their breadth and depth:

– **Focus of the review:** The focus of an SLR is to aggregate primary studies in terms of its results and investigate whether these results are consistent or contradictory. SLRs are performed to synthesize evidence. On the other hand, SMs provide a broader view of a research topic and identify both clusters (group of studies related to a same theme – may be suitable to undertake an SLR on this theme) and subtopics in which more primary studies are needed to be developed.

– **Research Questions:** SLRs focus on very narrow questions, while SMs focus on broad questions, which may limit the analysis, interpretation and generalization of the findings. SLRs aggregate results related to specific research questions. SMs have an objective to find and classify primary studies in subtopics.

– **Methods for searching:** Both SLRs and SMs attempt to be exhaustive in finding all relevant studies. The aim is to be able to answer the review questions by including all relevant research, however, while SLRs generally look at one type of evidence, i.e., empirical studies, SMs may include many different types of primary research, i.e., empirical studies, technical reports, theses, among others.

– **Methods for selecting:** One of the main differences between SLR and SM is related to studies selection activity. The scope of an SM is broader and the analysis and synthesis more general than in an SLR. SMs involve more studies to be selected while SLRs involve fewer studies, but they should be analyzed in a greater depth.

– **Methods for data extraction:** For undertaking an SM the data to be collected from the primary studies could include: (i) bibliographic information on the publications in which the primary studies are reported, e.g. Journal title, publication year; and (ii) basic data to describe what research has been done and how, e.g. country of the study, technique used, among others. For an SLR the data one collects from the primary studies could include: (i) detailed data on the methods and results of each study; (ii) a structured description of each study; (iii) the results (findings) of each study.

– **Synthesis:** SMs simply describe basic details about each primary study and variables can be used in coding the studies. The description may include, e.g., methods used, geographical distribution of the studies, year of the publication. SMs tabulate primary studies into categories. SMs may synthesize all, or part of the research studies described in a map. In SLRs the main interest is a full synthesis of results.

### III. TERTIARY STUDY ON PRACTICAL SIMILARITIES AND DIFFERENCES BETWEEN SLRs AND SMs

Although there are publications defining how to perform SLRs and SMs, such as [2] and [3], there is still need for studies addressing different aspects of the conduction of SLRs and SM in practice. In order to evaluate similarities and differences between SLR and SM studies we performed a tertiary study, which is presented in this section. Tertiary study is considered as a review that focuses only on secondary studies, i.e., it is a review about other secondary studies [8].

First of all, we created our search string, described as: (“*software engineering*”) AND (“*systematic literature review*” OR “*systematic review*” OR “*systematic mapping*” OR “*mapping study*” OR “*systematic literature mapping*”).

After the definition of the search string, the process of identifying relevant literature was started. We chose the databases based on criteria, such as: (i) coverage (large number of conferences proceedings and journals in different knowledge areas); (ii) content update (publications are regularly updated); (iii) availability (full text of the papers are available); (iv) quality of results (accuracy of the results obtained by the search); and (v) versatility to export (a mechanism to export the results of the search is available). The most commonly used databases in SE area that meet to the characteristics described above are: IEEE Xplore (<http://ieeexplore.ieee.org>); ACM Digital Library (<http://dl.acm.org>); ScienceDirect (<http://www.sciencedirect.com>); Scopus (<http://www.scopus.com>); and Compedex (<http://www.engineeringvillage2.org>).

Studies were included in the study if they met the following inclusion criteria (IC): **IC1:** The study should be in the SE area; and **IC2:** The study should present a protocol and/or the description of its conduction.

With respect to the exclusion criteria (EC), studies were excluded if: **EC1:** The study does not have an abstract; **EC2:** The study is just published as an abstract; **EC3:** The study is not written in English; **EC4:** The study is an older version of other study already considered; **EC5:** The study is not a scientific study, such as editorials, summaries of keynotes, workshops, and tutorials; **EC6:** The study is not a secondary study; **EC7:** The study is a book chapter or a guide.

A total of 970 studies were identified during the search for evidence, including 355 duplicates. Out of these studies, we selected secondary studies by reading their titles and abstracts and applying the inclusion and exclusion criteria. As a result, a total of 490 studies were selected and 125 were excluded. Finally, the 490 papers were read in full and inclusion and exclusion criteria were again applied, resulting in 320 studies being rejected. Thus, we identified 170 relevant studies from the five sources that we searched.

With the final set of secondary studies, the data extraction and synthesis activities were carried out on all 170 papers. We planned to perform the data synthesis for our RQs using tables (totals and summaries). The RQs are answered and described in the next section.

#### A. Results

In order to answer the RQs we created a data extraction form using Microsoft Excel software to store relevant informa-

TABLE I. COMPARING SLR AND SM (ADAPTED FROM [2], [3]).

Features	SLR	SM
<b>Focus of the review</b>	<ul style="list-style-type: none"> <li>- Identify, analyze and interpret all available evidence related to a specific RQ</li> <li>- Identify best practices based on empirical evidence</li> </ul>	<ul style="list-style-type: none"> <li>- Identify and classify what evidence is available (broad review) in a specific topic of area</li> <li>- Establish the state of evidence</li> </ul>
<b>Research Questions</b>	<ul style="list-style-type: none"> <li>- Narrow RQs</li> <li>- Specific RQs</li> <li>- Consider population; intervention; comparison and outcomes (PICO)</li> </ul>	<ul style="list-style-type: none"> <li>- Broader RQs</li> <li>- Multiple RQs</li> <li>- Consider only population and intervention</li> </ul>
<b>Methods for searching</b>	- Search string highly focused	- Search string less highly focused
<b>Methods for selecting</b>	<ul style="list-style-type: none"> <li>- Generally few studies are considered</li> <li>- The studies are evaluated in details</li> </ul>	<ul style="list-style-type: none"> <li>- A large number of studies are considered (broad coverage)</li> <li>- The studies are not evaluated in details</li> </ul>
<b>Methods for data extraction</b>	<ul style="list-style-type: none"> <li>- The primary studies are assessed regarding their quality (the main goal is to establish the state of evidence)</li> <li>- Include data extraction procedures</li> <li>- It is a time-consuming task</li> </ul>	<ul style="list-style-type: none"> <li>- The primary studies are not assessed regarding their quality</li> <li>- Much broader (classification and categorization stage)</li> <li>- It is not a time-consuming task</li> </ul>
<b>Synthesis</b>	<ul style="list-style-type: none"> <li>- Include depth analysis techniques, e.g., meta-analysis and narrative synthesis</li> </ul>	<ul style="list-style-type: none"> <li>- Include no-depth analysis techniques, e.g., total and summaries</li> </ul>
<b>Dissemination of the results</b>	- Higher importance for practitioners (relevant to industry)	- May be more limited, the aim is to influence the future of the research in a specific topic

tion. The form was filled with data extracted from each study and it is available on <https://goo.gl/Uh5jck>.

The data was grouped into categories. This categorization was supported by Microsoft Excel software to filter data for analysis and results visualization. Therefore, it can be said that the adopted classification scheme emerged from the selected studies.

The following is a brief description of the results. First of all, we identified the secondary studies types. The results are showed in Table II. The majority of studies are SLRs (approximately 70%) and around 30% are SMs. We identified three updated SLRs, however they were not separately categorized, i.e., they were similar to the other studies and were categorized according to their type: SLR or SM.

TABLE II. SECONDARY STUDIES: TYPE

Study Type	Quantity	Percentage
SLR	118	69.41%
SM	52	30.59%

Moreover, we verified the publication type to know what are the most common targets in which these studies are published. The majority of the studies were published in Conferences (53.53%) followed by Journals (33.53%). More detailed findings can be observed in Table III.

TABLE III. SECONDARY STUDIES: LOCAL OF PUBLICATION

Publication Type	SLR (%)	SM (%)	Total (%)
Conference	62 (52.54%)	29 (55.77%)	91 (53.53%)
Journal	40 (33.90%)	17 (32.69%)	57 (33.53%)
Congress	2 (1.69%)	0 (0%)	2 (1.18%)
Symposium	11 (9.32%)	5 (9.61%)	16 (9.41%)
Workshop	3 (2.54%)	3 (2.54%)	4 (2.35%)

With an objective to show an overview of studies area, we followed “The Guide to the Software Engineering Body of Knowledge (SWEBOK)”, which describes generally accepted knowledge about software engineering [9], to classify the studies according to SE subareas. We created 11 subareas based on SWEBOK and another one labeled as “others”. In particular, this category includes studies not classified in other available areas, such as, studies about specific SE tools and experiments. The defined categories and the number/percentage of studies classified in each category can be visualized in Table IV.

As shown in Table IV the majority of SLRs and SMs are related to Software Requirements (17.65%) and Software Engineering Models and Methods (17.06%), followed by Software Construction (10.59%) and Software Engineering Professional Practice (10.00%). Considering only SLRs, the Software Requirements area (19.49%) remaining as the main area of publication. On the other hand, in SM context, the Software Engineering Models and Methods is the area of major interest (19.23%). Only a few of SLRs and SMs are on Software Maintenance, adding three SLRs and two SMs, totaling 2.94% of the included studies.

In the sequence, the RQs will be answered in details.

**RQ1:** *What types of research questions (RQs) have been answered in SLRs and SMs studies?*

We assessed the RQs of SLRs and SMs considering two main aspects: (i) RQ granularity: generality or specificities; and (ii) RQ formulation: the use of PICO (Population, Intervention, Comparison, Outcome) to structure RQs.

A RQ describing the main objective of a research is a generic question. Jia and Yu [10] suggest that the 5W+1H model can be used to formulate RQs in secondary studies, consequently generic questions are initiated with W-words and H-words, such as, “Why”, “Who”, “What”, “Why”, “Where”, “When”, “How”, “How many”. For example, in the SM performed by Paz et al. [11] the following generic RQs were created: “*What are the most commonly used methods to evaluate the usability of software applications in the context of a development process?*” and “*What types of applications are frequently reported in the literature as part of a usability evaluation in software developments?*”.

On the other hand, a specific RQ highlights particularities of an investigated research topic. For example, the SLR performed by Engström et al., [12] presents one generic RQ (“Which techniques for regression test selection in the literature have been evaluated empirically?”) and three specific RQs: (i) “Can these techniques be classified, and if so, how?”; (ii) “Are there significant differences between these techniques that can be established using empirical evidence?”; and (iii) “Can technique A be shown to be superior to technique B, based on empirical evidence?”. It can be observed that the specific RQs are related to the generic one.

TABLE IV. SECONDARY STUDIES: STUDIES AREA

Study Area	SLR (%)	SM (%)	Total (%)
<b>1- Software Requirements</b>	23 (19.49%)	7 (13.46%)	30 (17.65%)
<b>2- Software Design</b>	8 (6.78%)	8 (15.38%)	16 (9.41%)
<b>3- Software Construction</b>	11 (9.32%)	7 (13.46%)	18 (10.59%)
<b>4- Software Testing</b>	4 (3.39%)	4 (7.69%)	8 (4.71%)
<b>5- Software Maintenance</b>	3 (2.54%)	2 (3.85%)	5 (2.94%)
<b>6- Software Management</b>	8 (6.78%)	2 (3.85%)	10 (5.88%)
<b>7- Software Engineering Process</b>	9 (7.63%)	0 (0%)	9 (5.29%)
<b>8- Software Engineering Models and Methods</b>	19 (16.10%)	10 (19.23%)	29 (17.06%)
<b>9- Software Quality</b>	5 (4.24%)	2 (3.85%)	7 (4.12%)
<b>10- Software Engineering Professional Practice</b>	14 (11.86%)	3 (5.77%)	17 (10.00%)
<b>11- Software Engineering Economics</b>	4 (3.39%)	7 (13.46%)	11 (6.47%)
<b>12- Others</b>	7 (5.93%)	3 (5.77%)	10 (5.88%)

Overall, the results (shown in Table V) indicate that both SLRs and SMs presented more generic RQs (SLR 83.05% and SM 86,54%) than specific RQs. Nevertheless SLRs presented a slightly larger number of specific RQs (16,95%) than SMs (13,46%).

TABLE V. SECONDARY STUDIES: GENERIC AND SPECIFIC RQs

Study Type	Generic RQs (%)	Specific RQs (%)
SLR	98 (83.05%)	20 (16.95%)
SM	45 (86.54%)	7 (13.46%)

According to [2], PICO can help the RQs structuring. In summary, our results showed that PICO is used only in 26.27% of SLRs and in 32.69% of SMs (see Table VI). We can conclude that most of the SE researchers are not using PICO to structure their RQs.

TABLE VI. SECONDARY STUDIES: USE OF PICO

Study Type	Yes (%)	No (%)
SLR	31 (26.27%)	87 (73.73%)
SM	17 (32.69%)	35 (67.31%)

#### RQ2: How the search strings have been formulated?

In order to evaluate how search strings have been formulated, we considered three aspects, described following: (i) the description of the search string in the paper; (ii) the quantity of logical operator AND used to compose the search string; and (iii) the validation of the search string through the conduction of a pilot test.

The search string is fundamental to find relevant studies about a specific area or topic [2]. Approximately in 74% of SLRs and 77% of SMs (see Table VII) the search string used was not described, not even an external link containing this information was provided. In some secondary studies, such as, [13] and [14], only generic terms used for searches were described, however the formalization of the search string containing logical operators and the list of synonyms is not shown.

TABLE VII. SECONDARY STUDIES: SEARCH STRING DESCRIPTION

Study Type	Yes (%)	No (%)
SLR	87 (73.73%)	31 (26.27%)
SM	40 (76.92%)	12 (23.08%)

During the definition of the search string, the focus is the identification of terms related to the research topic that are commonly used in primary studies. A well accepted practice for formulating the search string is to identify related terms that can be considered synonyms concatenating them using

the logical operator OR. Subsequently, each group of terms is concatenated with the other groups using the logical operator AND.

As can be noted in Table VIII, 43.68% of SLRs search strings were formed by only one AND logical operator and 27.59% were formed by two ANDs. Similarly, 32.50% of SMs search strings were also formed by only one AND logical operator, and a total of 37.50% were formed by two ANDs.

The use of the logical operator AND tends to restrict the number of studies returned, since the result of the operation is TRUE only if all rows of the truth table contains values TRUE (e.g. term A=true AND term B=true THEN study retrieved). As shown in line two of Table VIII in seven SLRs their search strings were formed only by terms and their synonyms, without any AND operator, characterizing a fairly generic string.

TABLE VIII. SECONDARY STUDIES: SEARCH STRING AND LOGICAL OPERATORS

Quantity	SLR	MS
0	7 (8.05%)	1 (2.50%)
1	38 (43.68%)	13 (32.50%)
2	24 (27.59%)	15 (37.50%)
3	6 (6.90%)	5 (12.50%)
4	6 (6.90%)	4 (10.00%)
≥5	6 (6.90%)	2 (5.00%)

The pilot test, according to [2], aims to validate each one of the items defined in the protocol. The test verifies the feasibility and execution of the study allowing the identification of necessary modifications. During the execution of the pilot test, it is common to note, for example, the absence of terms in the search string that may lead to the non-identification of relevant studies. Our results, presented in Table IX, revealed that the vast majority of both SLRs and SMs did not conduct the pilot test. Only 20.34% of SLRs and 11.54% of SMs did it. The pilot test was conducted more times in SLRs than SMs.

TABLE IX. SECONDARY STUDIES: SEARCH STRING PILOT TEST CONDUCTION

Study Type	Yes (%)	No (%)
SLR	24 (20.34%)	94 (79.66%)
SM	6 (11.54%)	46 (88.46%)

#### RQ3: Which search strategy has been adopted in SLRs and SMs studies?

With respect to RQ3 it is possible to note that the most employed search strategy in conducting secondary studies in SE is the automatic search (see Table X). A total of 70 of the 118 SLRs (59.32%) and 24 of the 52 SMs (46.15%) used the automatic search as exclusive search strategy.

Although the automatic search is the most adopted strategy, it presents difficulties in its use, for example, a challenge is the definition of the search string [15]. Only the use of the automatic search may not be sufficient for identification of all relevant studies. This leads researchers to explore other strategies such as complementary mechanisms to extend the identification of relevant studies. The use of automatic search combined with other strategies is adopted in 36.44% of SLRs and in 48.07% of SMs (see lines four, five and six in Table X).

Only 5 SLRs (4.24%) and 3 SMs (5.77 %) adopted exclusively manual search. The snowballing strategy, which does not use search string, was not exclusively used. Some authors argue that snowballing is a complementary search strategy [16]. However, there are authors, such as, [17], [18] and [19] who have analyzed the possibility of using snowballing as main search strategy in secondary studies; they advocate the use of snowballing specially in SLRs updates scenario.

TABLE X. SECONDARY STUDIES: SEARCH STRATEGY

Search Strategy	SLR	MS
Automated	70 (59.32%)	24 (46.15%)
Manual	5 (4.24%)	3 (5.77%)
Automated + Manual	32 (27.12%)	13 (25.00%)
Automated + Manual +Snowballing	6 (5.08%)	4 (7.69%)
Automated + Snowballing	5 (4.24%)	8 (15.38%)

#### RQ4: Has the quality of the studies been assessed?

A significant difference between SLRs and SMs refers to study quality evaluation. In SMs the quality assessment is not mandatory, although it may be useful to ensure that there is sufficient information for data extraction. According to [3], performing quality assessment is highly recommended for SLRs, but optional in SMs. Our results, shown in Table XI, confirm this trend, since 62.71% of SLRs had the quality of their studies evaluated and 80.77% of SMs did not evaluate the quality of their included studies.

TABLE XI. SECONDARY STUDIES: QUALITY ASSESSMENT

Study Type	Yes (%)	No (%)
SLR	74 (62.71%)	44 (37.29%)
SM	10 (19.23%)	42 (80.77%)

## B. Discussions

In general, the purpose of a secondary study is to provide an overview of a research area and identify research gaps in this area [20]. SLRs and SMs, as secondary studies, enable the identification and aggregation of available evidence to answer research questions [21]. We can affirm that SLRs and SMs support the decision making related to a research to be developed and that there are many similarities between them. However, theoretically, there are also differences, especially regarding the objectives, research questions, search strategy, selection and evaluation of the quality of primary studies, as well as the analysis of data and results obtained. In this study we evaluated if the theoretical differences pointed out in the literature, presented in Section II, reflect in the practical conduction of secondary studies in ES.

Regarding the types of secondary studies that have been conducted by the SE community, SLRs are predominant. One possible explanation is that the first guidelines [1] for

conducting SLRs focused on describing this type of study. The SM was superficially presented for the first time, only in 2008, as a mapping process defined by Petersen et al. [22] and later it was updated in 2015 [3]. Therefore, the greater knowledge of the community is on SLRs.

Although secondary studies have been conducted for more than 10 years, only three included studies were updates. Currently, there are processes focused on SLR updates [23], as well as proposals for using visual techniques to support selection of new evidence [24]. However, there is a lack of guidelines on what is updated and how long to update SLRs, explaining the low number of updated studies. It is noteworthy that an outdated secondary study loses its relevance.

Most of secondary studies are published in conferences. One concern with this practice is that there is a limitation of the number of pages available to document the revision. Therefore, it is important that researchers provide additional external information on the conduct of the study in external web pages.

The subareas Software Requirements and Software Engineering Models and Methods concentrated the largest number of secondary studies. As a result it would be interesting the SE community to conduct tertiary studies on these topics.

Petersen et al. [3] affirm that RQs of SLR are specific and RQs of SMs are generic. However, our results showed that both studies presented a predominance of generic RQs. We observed that in this context there is a difference between theory and practice. The community does not distinguish the two studies based on the type of RQ addressed. There is no framework for defining RQs. We can affirm that PICO, although recommended in the guidelines for conducting secondary studies [8], has not been adopted in practice. The set of PICO criteria derived from Medicine [8], therefore, we argue that the main challenge to adopt PICO in an area different from Medicine, is to adapt the set of PICO criteria. For example, Kitchenham and Charters [8] advocate that population could be: (i) a specific role of ES (e.g. tester, manager); (ii) a specific category of software engineer (e.g., novice, experienced); (iii) an application area (e.g., IT systems, control systems); Or (iv) an industry group (e.g. telecom company, small business). Conversely, Biolchini et al. [25] report that population is the group that will be observed by intervention (e.g., publications on the subject investigated).

Surprisingly, in more than 70% of SLRs and SMs researchers do not formally present the search string adopted. As an alert, we can say that the absence of the search string description negatively impacts the reproduction of the secondary studies. In addition to that, both studies (71.27 % SLRs and 70.0 % SMs) have adopted generic strings, containing one or two ANDs (logical operator). Considering the theory, we expected that the SLRs presented strings containing a greater number of ANDs, i.e., more specific strings than SMs. The pilot test to validate the search string is usually not applied, not even in SLRs.

The automatic search, with the support of search strings, has been the main strategy used in both SLRs and SMs.

Corroborating with the theory of Kitchenham and Charters [8] and Petersen et al. [3], most of SLRs (62.71%) evaluate

the quality of their studies, while only 19.23% of SMs adopt the evaluation practice.

As a threat to the validity of this work we highlight:

– Recovery of studies: to avoid missing out relevant studies, one specialist on SLR was asked to create our set of quasi-gold standard. We used this set of studies for filtering the retrieved data set, hence increasing the precision rate of our tertiary study;

– Selection of relevant studies: to validate the selection of studies, an external reviewer independently applied the selection criteria to a subset of randomly studies. Then, we used the inter-raters agreement rate, which was determined using the Cohen's kappa statistic, to measure the reliability of our selection process; – Data extraction: to validate the data extraction activity, an external reviewer independently extracted data from a subset of randomly studies. We intend to further mitigate this threat by making our data set externally available so that our results can be replicated and validated by external researchers.

#### IV. CONCLUSIONS

The main contribution of this paper was to present a practical panorama about the current state-of-the-art on the conduction of secondary studies in SE. To build such panorama, we performed a tertiary study. Based on this panorama, it is possible to affirm that in practice, only the quality assessment is differently conducted in SLRs and SMs.

As future work we intend to: (i) analyze in depth other items to distinguish SLRs from SMs, such as, selection of studies and extraction of data; and (ii) define a decision tree to support SE researchers to decide if conducting an SLR or an SM.

#### REFERENCES

- [1] B. Kitchenham, "Procedures for performing systematic reviews," Software Engineering Group - Department of Computer Science - Keele University and Empirical SE - National ICT Australia Ltd, Joint Technical Report TR/SE-0401 (Keele) - 0400011T.1 (NICTA), 2004.
- [2] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*, ser. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series. Chapman & Hall/CRC, 2015.
- [3] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: an update," *Information and Software Technology*, vol. 64, no. 1, pp. 1–18, 2015.
- [4] H. Zhang and A. Muhammad, "Systematic reviews in software engineering: An empirical investigation," *Information and Software Technology*, vol. 55, no. 7, pp. 1341–1354, 2013.
- [5] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, no. 12, pp. 2049–2075, 2013.
- [6] P. Brereton, B. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, April 2007.
- [7] S. Imtiaz, M. Bano, N. Ikram, and M. Niazi, "A tertiary study: Experiences of conducting systematic literature reviews in software engineering," in *17<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering (EASE)*. ACM, 2013, pp. 177–182.
- [8] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele University and Durham University, UK, Tech. Rep. EBSE 2007-001, 2007.
- [9] I. C. Society, P. Bourque, and R. E. Fairley, *Guide to the Software Engineering Body of Knowledge (SWEBOK(R)): Version 3.0*, 3rd ed. Los Alamitos, CA, USA: IEEE Computer Society Press, 2014.
- [10] C. Jia and Y. T. Yu, "Using the 5w+1h model in reporting systematic literature review: A case study on software testing for cloud computing," in *2013 13<sup>th</sup> International Conference on Quality Software*, July 2013, pp. 222–229.
- [11] F. Paz and J. A. Pow-Sang, "Usability evaluation methods for software development: A systematic mapping review," in *2015 8<sup>th</sup> International Conference on Advanced Software Engineering Its Applications (ASEA)*, Nov 2015, pp. 1–4.
- [12] E. Engström, M. Skoglund, and P. Runeson, "Empirical evaluations of regression test selection techniques: A systematic review," in *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM '08. New York, NY, USA: ACM, 2008, pp. 22–31. [Online]. Available: <http://doi.acm.org/10.1145/1414004.1414011>
- [13] M. M. Hassan, W. Afzal, M. Blom, B. Lindström, S. F. Andler, and S. Eldh, "Testability and software robustness: A systematic literature review," in *2015 41<sup>st</sup> Euromicro Conference on Software Engineering and Advanced Applications*, Aug 2015, pp. 341–348.
- [14] M. I. Babar, M. Ghazali, and D. N. A. Jawawi, "Systematic reviews in requirements engineering: A systematic review," in *2014 8<sup>th</sup>. Malaysian Software Engineering Conference (MySEC)*, Sept 2014, pp. 43–48.
- [15] M. Riaz, N. Sulayman, M. Salleh, and E. Mendes, "Experiences conducting systematic reviews from novices' perspective," in *14<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering (EASE)*. BCS-eWiC, 2010, pp. 1–10.
- [16] B. Kitchenham, P. Brereton, M. Turner, M. Niazi, S. Linkman, R. Pretorius, and D. Budgen, "Refining the systematic literature review process—two participant-observer case studies," *Empirical Software Engineering*, vol. 15, no. 6, pp. 618–653, 2010.
- [17] S. Jalali and C. Wohlin, "Systematic literature studies: Database searches vs. backward snowballing," in *6<sup>th</sup> International Symposium on Empirical Software Engineering and Measurement (ESEM)*. ACM, 2012, pp. 29–38.
- [18] C. Wohlin, "Second-generation systematic literature studies using snowballing," in *Proceedings of the 20<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering (EASE)*. ACM, 2016, pp. 15:1–15:6.
- [19] K. Felizardo, E. Mendes, M. Kalinowski, r. F. Souza, and N. Vijaykumar, "Using forward snowballing to update systematic reviews in software engineering," in *10<sup>th</sup> International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2016.
- [20] C. Wohlin, P. Runeson, P. Silveira Neto, E. Engstrom, I. Machado, and E. de Almeida, "On the reliability of mapping studies in software engineering," *Journal of Systems and Software*, vol. 86, no. 10, pp. 2594–2610, 2013.
- [21] B. Kitchenham, R. Pretorius, D. Budgen, P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering - a tertiary study," *Information and Software Technology*, vol. 52, no. 8, pp. 792–805, 2010.
- [22] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *12<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering (EASE)*. BCS-eWiC, June 2008, pp. 1–10.
- [23] F. Ferrari and J. Maldonado, "Experimenting with a multi-iteration systematic review in software engineering," in *5<sup>th</sup> Experimental Software Engineering Latin America Workshop (ESELAW'08)*, 2008, pp. 1–10.
- [24] K. Felizardo, E. Nakwgawa, S. MacDonell, and J. Maldonado, "A visual analysis approach to update systematic reviews," in *18<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2014, pp. 1–10.
- [25] J. Biolchini, P. Mian, A. Natali, and G. Travassos, "Systematic review in software engineering: Relevance and utility," PESC/COPPE/UFRJ, Tech. Rep., 2005.