

Case Studies Synthesis: Brief Experience and Challenges for the Future

Daniela S. Cruzes
IDI-NTNU
NO-7491 Trondheim,
Norway
dcruzes@idi.ntnu.no

Tore Dybå
SINTEF
NO-7465 Trondheim,
Norway
tore.dyba@sintef.no

Per Runeson
Department Computer
Science, Lund University,
Box 118, SE-221 00 Lund,
Sweden
per.runeson@cs.lth.se

Martin Höst
Department Computer
Science, Lund University
Box 118, SE-221 00 Lund,
Sweden
martin.host@cs.lth.se

Abstract—Synthesis of case studies is different from synthesis of purely quantitative studies, for example, in that sampling and analysis in primary studies have been carried out differently, and that primary results are of a different nature. The objective of this research is to identify what challenges should be considered when choosing and using a method for synthesis of case studies. We collected experience from independent synthesis of two published case studies (on trust in outsourcing) by two teams; one team applied cross-case analysis, the other team applied thematic synthesis. The two teams reached both supporting and complimentary conclusions. Identified challenges relate to the goals and research questions of the cases to be synthesized, the number of case studies, temporal and spatial variations, and access to raw data.

Keywords—Evidence-based software engineering; systematic reviews; research synthesis; case study

I. INTRODUCTION

Case study is “an empirical enquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” [17]. Software engineering projects, processes, and artifacts are typical objects, feasible for the case study approach. Case studies are characterized by their flexible nature, evolving over the course of the study, and defined by a focal point, rather than a precise scope. Case studies are mostly informed by qualitative data, which is rich, but lack standardized structure and analysis methods, as quantitative data has. Selection of cases to study is not governed by sampling logic and representativeness; rather cases are selected for the purpose of being ‘typical’, ‘critical’, ‘revelatory’, or ‘unique’ in some respect [17]. Further, the number of study objects needed to draw conclusions is not governed by probability calculations, but by the fuzzy concept of ‘saturation’, i.e., when new study objects do not add more value to the question under study.

Research synthesis is used as a collective term for a family of methods to summarize, integrate, combine, and compare the findings of different studies on a specific topic or research question [4]. Synthesis methods are usually tailored to a particular type of evidence, for example meta-analysis aggregates and averages findings from experimental or quasi-experimental studies, whereas meta-ethnography synthesizes findings from ethnographic and qualitative studies [3]. These methods embody the idea of making a new whole out of the parts to provide novel concepts and higher-

order interpretations, novel explanatory frameworks, an argument, new or enhanced theories, or conclusions. Research synthesis is built upon the observation, that no matter how well designed and executed, empirical findings from single studies are limited in the extent to which they may be generalized [4].

The synthesis of case studies is different from conducting meta-analyses of formal experiments with quantitative data. The synthesis must take into account the flexible nature of the case study, the qualitative characteristic of the data, and the type of cases in each of the primary studies.

The empirical evidence, which such syntheses depend upon, is the data on which a conclusion or judgment may be based. Interpreting and judging such evidence, however, depends on the ‘eye of the observer’. Although there are many ways to generate evidence, case studies have a special ability to provide deep understandings of the phenomena under study from direct observation of industrial practice.

We aim to start a discussion on these challenges of synthesizing evidence from reported case studies in software engineering, and on which synthesis methods that are specifically suitable for case studies. Our main research questions is:

What challenges should be considered when choosing a method of synthesis for evidence across case studies?

Based on our studies of the literature on research synthesis, and the application of two methods to two papers, we identified the challenges and methods, presented in Section II. We split into two independent teams, applying two synthesis methods to two papers that investigated critical factors of trust in outsourced projects. The experiences and challenges from this analysis are presented in Section III. Section IV concludes and outlines further work.

II. CHALLENGES OF SYNTHESIZING CASE STUDIES

There are several factors that can impact the decision of which synthesis method to choose including: goals and research questions, types of case studies selected, number of primary studies, and knowledge of the method. Most probably no single method will offer all the required features for the synthesis, so a combination of methods may often be the best approach. Table 1 outlines some of the methods that are most relevant for synthesizing evidence across case studies (a more complete list is provided in Cruzes and Dybå [4]). In the following subsections we describe some challenges of using these methods.

TABLE 1. RELEVANT CASE STUDY SYNTHESIS METHODS.

Synthesis method	Description
Case survey [12][16]	Formal process for systematically coding relevant data from a large number of case studies for quantitative analysis, allowing statistical comparisons across studies. Study findings and attributes are extracted using closed-form questions for increased reliability, while survey analysis methods are used on the extracted data. The resulting dataset is used to construct cross-case matrices or summary tables.
Qualitative comparative analysis (QCA) [13]	The qualitative comparative analysis method is a mixed synthesis method that analyzes complex causal connections using Boolean logic to explain pathways to a particular outcome based on a truth table. The Boolean analysis of necessary and sufficient conditions for particular outcomes is based on the presence/absence of independent variables and outcomes in each primary study
Cross-case analysis [9]	Includes a variety of devices, such as tabular displays and graphs, to manage and present qualitative data. It includes meta-matrices for partitioning and clustering data in various ways. Evidence from each primary study is summarized and coded under broad thematic headings, and then summarized within themes across studies with a brief citation of primary evidence. Commonalities and differences between the studies are noted.
Thematic Synthesis [5][15]	A method for identifying, analyzing, and reporting patterns (themes) within data. It organizes and describes the data set in rich detail and interprets various aspects of the research topic. It can be used within different theoretical frameworks, and it can be an essentialist or realist method that reports experience, meanings, and the reality of participants. It can also be a constructionist method, which examines the ways in which events, realities, meanings, experience, and other aspects affect the range of discourses.

A. Goals and Research Questions

Several methods have a broad application to a variety of different questions. It is necessary therefore, to select a synthesis method that is applicable to the underlying study aim and question.

Typically, a synthesis focuses on a well-defined question and aims to provide an answer by synthesizing the findings from a relatively narrow range of quality-assessed studies. A fundamental distinction regarding the objective of such syntheses is whether they attempt to provide *knowledge support* or *decision support* [12]. A synthesis directed to knowledge support will typically bring together and synthesize evidence on a particular topic, while a synthesis aimed at decision support will be more specific and include analytical tasks to help make a decision within a particular context [8].

Although these are the two ultimate goals, the synthesis goal may vary from the need of pure factual knowledge to attainment of judgment and decision [1]. These two are not polar opposites, but fall at opposite ends of a question spectrum (impacts of objects of study, comparison of objects of study, feasibility of objects of study, impacts of context on the object of study, etc.). An object of study can be a technique, a method, an approach, or a phenomenon.

Knowledge of facts such as whether a specific object of study is important or not, can be suitably answered by a *thematic synthesis* which can bring broad conclusions and is flexible to the buildup of knowledge [1]. Contextualizing an object by comparing different usage contexts can be performed with a *cross-case analysis*, for example. Impacts of an object of study on software development as well as the feasibility of the object can be synthesized by *case survey*, *comparative analysis* or *cross-case analysis*. Some more specific techniques and a more interpretative approach would be needed to provide guidelines for decision support.

B. Number of Case Studies

Some synthesis methods require more studies than others to be effectively applied. However, it is not possible to say for sure how many studies are needed to answer a specific research question. For qualitative studies, the notion of ‘saturation’ must be taken into account, i.e., judging whether new studies add knowledge on the research question. The number of studies needed is very dependent of how broad the research question is and how many independent variables and factors affect the results of the object of study. For example, a *case survey* cannot be meaningfully performed with a small number of cases, as the goal is to have statistically significant results.

In the case that the synthesis comprises many studies, then the synthesis will probably be more quantitative than qualitative. This is so because whenever one is attempting to incorporate a large number of cases into a single synthesis, it will be necessary to reduce the evidence to a smaller number of dimensions [6].

There is always a trade-off between the ability to generalize and the ability to understand fully all the nuances of individual cases. The use of different methods may result in different conclusions. This is a general issue on studies based on qualitative data and is an effect of the richness and lack of precision of such data. A measure to increase the validity of analysis is to maintain a clear chain of evidence from the primary studies to the synthesized evidence.

C. Temporal and Spatial Variation

Gerring identifies two possible styles of covariational evidence in a case study synthesis: *temporal* and *spatial* [6]. Spatial variation refers to case studies that were run by different research groups/authors but with similar objectives and instruments of data collection. A central aspect of the spatial variation is the case context. Petersen and Wohlin,

for example, proposed a scheme comprising several context facets [11]: product, processes, practices, tools, techniques, people, organization, and market. Besides, if different groups perform the studies, one challenge is that they may have different measurement procedures or definitions etc.

Under circumstances of extreme case-heterogeneity, the researcher may decide that it is better to focus on a single case or a small number of relatively homogeneous cases [6]. Cross-case evidence drawn from a handful of most-similar cases may be more useful than cross-case evidence of many studies, even though the ultimate interest of investment is in a broader population of cases. The issue of population heterogeneity/homogeneity may be understood, therefore, as a trade-off between the number of cases and the number of variables.

Temporal variation refers to development over time. If a research group is running a series of case studies successively, the synthesis must consider context variations over time in the studies that may explain the change. Clearly, cases must be similar to each other in whatever respects might affect the causal relationship that the researcher is investigating, or such differences must be controlled for [6]. Uncontrolled heterogeneity means that cases are “apples and oranges”, and that one cannot learn anything about underlying causal processes by comparing their histories.

D. Limited access to raw data

Synthesis of evidence published in journal and conference proceedings involve a key challenge in the limited access to raw data. Only syntheses and quotations are generally available, and the raw data is rarely openly published for several reasons. Getting access to or working with other researchers’ (probably disclosed) data is not easy. The synthesis of several studies may have to be a joint venture between the involved researchers. Or perhaps a joint effort to disclose raw data of the primary studies may be a way to overcome this challenge.

Terminology and definitions may also be different between studies. In some cases they are quite well defined, which helps, but does not solve the problem. Well-established software engineering terminology may help addressing this challenge. The challenge here is that the underlying factors of interest have different meanings in different contexts (conceptual stretching) or the causal relationships are different in different contexts.

III. EXPERIENCE WITH CASE STUDY SYNTHESIS

To explore some ideas of this paper, we defined a synthesis goal as an example. We decided to run an independent synthesis of two papers. One synthesis of the two papers would be performed in Sweden and the other in Norway. The goal of the synthesis was to:

Understand factors of trust in outsourcing relationships.

This is a knowledge support goal, and at first we could use any of the syntheses methods described in Table 1.

We identified two papers that could help us to fulfill our goal [2][10]. The first study was based on interviews of 18 software development practitioners in India [10], while the second study was based on interviews of 12 Vietnamese practitioners developing software for Far Eastern, European, and American clients [2]. As these papers do not describe causal relationships between the factors and the trust relationship, we excluded the comparative qualitative analysis from our options of methods. Having only two papers to synthesize, we also disregarded the case survey method, because it requires a larger amount of papers for a meaningful quantitative synthesis. Of the two remaining methods, we decided that the team in Sweden would perform a synthesis using the *cross-case analysis* method and the team in Norway the *thematic synthesis* method.

The *thematic synthesis* method produced a graph showing the relations between the concepts identified, with legends showing which ‘trust’ factors originate from one study, the other, or both. The *cross-case analysis* method produced tables, comparing the characteristics of the two cases, and comparing the ‘trust’ factors originating from the two studies. Further, the analysis identified factors from one study, which is defined as higher-level factor in the other.

In our example, the primary studies had the same goals and methodological framework. The main variations were the target culture (India vs. Vietnam) and the research groups. There was a temporal variation in the sense that Babar et al.’s study was run based on Oza et al.’s previous paper and results. There is hence a threat that Babar’s results may be very much influenced by Oza et al.’s results. But they also added two important spatial variations: definitions and target cultures. The terminology and definitions are partly different; e.g., the factor “performance” was defined by Babar et al. as: “*How performance (productivity/effectiveness) of staff in carrying out the projects help to maintain trusts with clients*”, while in Oza et al. the same term was defined as: “*You have to perform the work to gain the trust, it is based on performance*”.

However, in both cases important terms were well defined, which helped with understanding the differences between them. Both teams were conscious about the definitions. In the cross-case analysis the results table also include pairing of the definitions across the two studies. In the thematic synthesis, the definitions were kept in the thematic network so the researchers could always see and compare the different definitions.

While synthesizing the results from the two papers, both teams found that the quotes inserted in the papers were not enough to be totally confident that we were synthesizing the papers in the right level of abstraction and granularity. Further, in both papers only the off-shore side of the relationship was interviewed, so the synthesis cannot reflect the perceptions of the clients on the trust relationship.

The final conclusions of the synthesis reached by the two teams were not the same in all aspects, but give

different views of the synthesis of the two papers. So the factors the two teams derived as the most important factors for trust in outsourcing relationships were sometimes complimentary and sometimes grouped in different perspectives. But overall, the two teams reached similar conclusions. Additionally, Babar et al. included a *narrative synthesis* (which the teams in this study did not read until after their synthesis) focusing on hypothesized differences between the Indian and Vietnamese contexts, which were not part of the original studies. Having access to at least the raw data from one of the studies gave them the opportunity to go deeper in their synthesis.

Another challenge we faced in our example, was related to the fact that only the off-shore people were interviewed. In searching for the next study to synthesize, a study from the outsourcing point of view would be more relevant than adding a third study from the off-shore perspective. Hence, evidence selection is more complex than just searching for available studies.

IV. CONCLUSIONS AND FUTURE WORK

Choosing a suitable method for synthesizing evidence across a set of case studies is not straightforward. Based on our studies of the literature on research synthesis, and the application of two methods to two papers, we identified the following key challenges: First, the selection of synthesis method depends on the aims of the synthesis. Better understanding of the methods and how suitable they are to specific research questions is essential. Second, some synthesis methods require more studies than others to be effectively applied. Third, temporal and spatial variations may limit the use of a method or the conclusions of the synthesis. Finally, limited access to raw data may limit the ability to fully understand and synthesize studies.

Further research includes obtaining more experience from synthesizing evidence from multiple case studies, both across time and across space, exploring different methods of synthesis. A goal would be to better understand which synthesis methods that fit to a given situation and to tailor methods of synthesis to the specific challenges of case studies in software engineering.

REFERENCES

- [1] Ashrafiyan, H., Darsi A., and Athanasios, T., *Evidence Synthesis: Evolving Methodologies to Optimise Patient Care and Enhance Policy Decisions in Evidence Synthesis in Healthcare*, Springer-Verlag, London, 2011.
- [2] Babar, M.A., Verner, J.M., and Nguyen, P.T., Establishing and maintaining trust in software outsourcing relationships: An empirical investigation. *J. Syst. Softw.* 80(9): 1438-1449, 2007.
- [3] Bethel, E.C., and Bernard, R.M., "Developments and trends in synthesizing diverse forms of evidence: beyond comparisons between distance education and classroom instruction," *Distance Education*, 31(3): 231-256, 2010.
- [4] Cruzes, D.S., and Dybå, T., "Research synthesis in software engineering: A tertiary study", *IST*, 53(5): 440-455, 2011.
- [5] Cruzes, D.S., and Dybå, T., "Recommended Steps for Thematic Synthesis in Software Engineering", *ESEM*, 2011.
- [6] Gerring, J. *Case Study Research: Principles and Practices*, New York: Cambridge Univ. Press, 2007.
- [7] Kitchenham, B.A. and Charters, S., *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Version 2.3, Keele University, EBSE Technical Report, EBSE-2007-01, 2007.
- [8] Mays, N., Pope, C. and Popay J., Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field, *J. Health Serv. Res. Policy*, 10: 6-20, 2005.
- [9] Miles, M.B. and Huberman, A.M. *Qualitative Data Analysis: An Expanded Source Book*, Sage, 1994.
- [10] Oza, N., Hall, T., Rainer, A., Grey, S., Trust in software outsourcing relationships: an empirical investigation of Indian software companies. *IST*, 48: 345-354, 2006.
- [11] Petersen, K. and Wohlin, C., "Context in industrial software engineering research". *ESEM* 2009, 401-404..
- [12] Pope, C., Mays, N., and Popay, J. *Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods*, Open University Press, 2007.
- [13] Ragin, C.C. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, Univ. of California Press, 1987.
- [14] Runeson, P and Höst, M., "Guidelines for conducting and reporting case study research in software engineering". *Empirical Software Engineering*, 14(2):131-164, 2009.
- [15] Thomas, J., and Harden, A., "Methods for the thematic synthesis of qualitative research in systematic reviews", *BMC Medical Research Methodology*, 8:45, 2008.
- [16] Yin, R.K. and Heald, K.A. "Using the Case Survey Method to Analyze Policy Studies," *Adm. Science Quarterly*, 20: 371-381, 1975.
- [17] Yin R.K. *Case study research. Design and methods*, 4th edn. London, Sage, 2008.