Information and Software Technology 94 (2018) 234–244

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

The contribution that empirical studies performed in industry make to the findings of systematic reviews: A tertiary study

David Budgen*[,a], Pearl Brereton[b], Nikki Williams[1,c], Sarah Drummond[a]

[a] Durham University, Department of Computer Science, Durham DH1 3LE, United Kingdom
[b] School of Computing & Maths, Keele University, Staffordshire ST5 5BG, United Kingdom
[c] Cranfield University, Centre for Electronic Warfare, Information & Cyber, Defence Academy of the United Kingdom, Shrivenham SN6 8LA, United Kingdom

ARTICLE INFO

Keywords:
Systematic review
Primary study
Industry study
Case study

ABSTRACT

Context: Systematic reviews can provide useful knowledge for software engineering practice, by aggregating and synthesising empirical studies related to a specific topic.
Objective: We sought to assess how far the findings of systematic reviews addressing practice-oriented topics have been derived from empirical studies that were performed in industry or that used industry data.
Method: We drew upon and augmented the data obtained from a tertiary study that performed a systematic review of systematic reviews published in the period up to the end of 2015, seeking to identify those with findings that are relevant for teaching and practice. For the supplementary analysis reported here, we then examined the profiles of the primary studies as reported in each systematic review.
Results: We identified 48 systematic reviews as candidates for further analysis. The many differences that arise between systematic reviews, together with the incompleteness of reporting for these, mean that our counts should be treated as indicative rather than definitive. However, even when allowing for problems of classification, the findings from the majority of these systematic reviews were predominantly derived from using primary studies conducted in industry. There was also an emphasis upon the use of case studies, and a number of the systematic reviews also made some use of weaker 'experience' or even 'opinion' papers.
Conclusions: Primary studies from industry play an important role as inputs to systematic reviews. Using more rigorous industry-based primary studies can give greater authority to the findings of the systematic reviews, and should help with the creation of a corpus of sound empirical data to support evidence-informed decisions.

1. Introduction

Knowledge about the effectiveness of established and emerging practices in software engineering can be derived in a number of ways, ranging from using 'expert opinion' through to conducting rigorous empirical studies. Although all have value, it has been argued that the emphasis has too often been on use of the former [1].

In the period since the idea of using secondary studies (systematic reviews) as a source of software engineering knowledge was proposed in 2004 [2], these have become a well established tool for consolidating different sources and forms of study. Terms such as 'evidence-based' or 'evidence-informed' are usually associated with their use. Because a systematic review aggregates and synthesises the findings from many 'primary' studies in an unbiased manner it can be considered as a form of *value multiplier*, in the sense that its findings should carry much greater authority than the outcomes of a single empirical study. Since empirical studies conducted in industry should themselves already carry a certain degree of authority, their use in systematic reviews is particularly important for generating findings that should carry much greater weight than expert opinion. The study described in this paper examines how far primary studies conducted in industry do actually contribute to the findings of systematic reviews.

In 2011 we undertook a tertiary study (a systematic review of systematic reviews) to identify how well the information available from published systematic reviews could be used to help inform introductory teaching about software engineering and hence, by implication, should also be suited to informing software engineering practice [3]. In this paper we refer to this as ETS1 (Education Tertiary Study 1). More recently, we have extended and refined this study, and have identified a set of 48 systematic reviews published up to the end of 2015 [4]. We

* Corresponding author.
E-mail addresses: david.budgen@durham.ac.uk (D. Budgen), o.p.brereton@keele.ac.uk (P. Brereton), nikki.williams@cranfield.ac.uk (N. Williams), sarah.drummond@durham.ac.uk (S. Drummond).
[1] The work reported in this paper was undertaken when Nikki Williams was employed by Keele University.

http://dx.doi.org/10.1016/j.infsof.2017.10.012
Received 27 April 2017; Received in revised form 11 September 2017; Accepted 16 October 2017
Available online 23 October 2017
0950-5849/

refer to this study as ETS2.

One way in which ETS2 differs from ETS1, apart from the period covered, is that for each systematic review included, we have required that its findings should not only provide knowledge about software engineering, but also that the findings should be supported by some form of *provenance* showing how they were derived, so making it possible to make some assessment of the confidence that can be placed in them. As a result, ETS2 is based upon a core set of 48 systematic reviews that address a range of software engineering practices, and provide conclusions and/or recommendations about practice that are explicitly derived from and supported by 'primary' empirical studies.

Since these systematic reviews address topics relevant to practice, rather than research, an obvious question to ask is how far their findings are based upon using primary studies that have been conducted in industry, or have used industry data? In this paper we describe a supplementary analysis of these studies, aimed at addressing the following research question:

> *"For those systematic reviews that address topics relevant to practice and teaching, to what degree are the findings derived from the use of primary studies that have been conducted in an industry context?"*

To answer this, we have interpreted 'derived' as being the proportion of primary studies that have been conducted in an industry context. Ideally, what we would really like to know is in what way these primary studies contribute to the individual findings of a systematic review. However, as systematic reviews rarely report upon their analysis or synthesis processes in sufficient detail to determine this, we have had to use proportion as a surrogate measure.

We also need to explain what is meant by 'industry context'. For this study, we consider this to be where an empirical study (such as a case study) is either performed in an industry setting and/or with participants who are employed in industry; or where the study makes use of industry artifacts in some way.

Inevitably, since the systematic reviews rarely report the characteristics of the primary studies in detail, there are some limitations upon the confidence that we can place upon the counts of primary studies obtained from our analysis.

Despite these limitations, what does emerge very clearly is that, taken as a whole, the findings of this set of 48 systematic reviews are substantially derived from primary studies that have been conducted in an industrial setting, to an extent that we were not really expecting. This highlights the important role that such studies can play in providing well-founded software engineering knowledge, and hence the importance of finding ways to improve their quality. We are also able to make some observations about the forms of empirical studies that have been used as the primary studies.

The rest of this paper is structured as follows. The next section provides a brief background about the roles and use of systematic reviews in software engineering, as well as the role performed by the primary studies. We then describe our research method—and since much of the detail of this is reported elsewhere, we confine our detailed description to the elements specific to this study. Similarly we provide only an outline of the way that the study was *conducted*, placing our main emphasis upon the findings. We then discuss the findings and make observations about how far these appear to have been influenced by empirical studies in industry.

2. Background

The systematic review is now a well-established tool of empirical software engineering, and the book by Kitchenham, Budgen and Brereton describes their use in software engineering, as well as providing an updated set of guidelines for conducting and reporting them [5]. However, although systematic reviewers often comment on the poor quality of reporting provided by the authors of the primary studies, the processes and findings of systematic reviews are not always reported particularly well either [6].

This section provides a brief summary of the forms that systematic reviews can take; followed by a discussion about the sort of knowledge they can provide; and finally outlines some relevant characteristics of the context for primary studies used in software engineering.

2.1. Forms of systematic review

A systematic review is classified as a *secondary study*, since it aims to identify all empirical studies relevant to the chosen topic (referred to as the *primary studies*) and to synthesise their results in order to produce its findings. As such therefore, a systematic review does not involve making any direct measurements related to the topic, its role is entirely concerned with aggregation and synthesis of the findings from other studies.

The degree and form of synthesis can vary. Many systematic reviews are less concerned with synthesising the findings of the primary studies and more with categorising their characteristics (such as the type of research question they address), usually using some model or framework. Such studies are referred to as *mapping studies*, and while they can perform a useful role in terms of identifying what aspects of a topic have or have not been studied, the lack of findings means that they do not contribute to the analysis described in this paper. *Tertiary studies* are usually a form of mapping study performed to categorise secondary studies. The underlying study for this paper (ETS2) is a tertiary study, identifying and categorising the secondary studies that address software engineering topics of relevance to teaching and practice.

An obvious question is why systematic reviews are viewed as an important form of empirical study. And in the context of this paper, we might also ask what contribution can they make to improving the practice of conducting studies performed in industry?

To answer the first question, one reason why they are viewed as important is that they are *systematic*, conducted according to a predefined plan (the *research protocol*) that is designed to minimise possible bias arising from different factors, including any pre-conceived ideas of the researchers or 'cherry-picking' among primary studies [5]. Another reason is that the process of synthesis should help avoid an over-reliance upon specific studies. All human-centric studies (and most software engineering studies are of this form) can be expected to demonstrate a degree of *variation* in their outcomes, especially (as in software engineering) where the participants may need to be selected on the basis of their skills and experience [7].

For studies performed in industry there are additional sources of possible bias, such as the culture of any organisations concerned. So, synthesising the outcomes from a set of such studies can help with distinguishing those effects that arise from the 'intervention' being studied (such as the use of a test-first strategy) from the effects that are produced by the practices and culture of the host organisation.

The second question is essentially one of motivation, and partly relates to the role of a tertiary study as a mapping study. Identifying how extensively industry-based studies are used in systematic reviews, and the types of study commonly used, can help determine where improvements in the conduct of such primary studies could make a particularly valuable contribution.

2.2. Knowledge provided by systematic reviews

The findings of a systematic review can take a range of forms. In the case of mapping studies, the findings are usually concerned with *categorisation* of the primary studies, and so concentrate upon the research issues addressed by the primary studies, although they may report on other characteristics of these such as the date and venue of publication (to identify trends).

Systematic reviews may also report on other aspects of the primary studies that they have identified, some of which may be related to the *provenance* of the findings. Many perform a quality analysis of the