J PROD INNOV MANAG 2019;0(0):1–30
 DOI: 10.1111/jpim.12492

State-of-the-Art: The Quality of Case Study Research in Innovation Management*

Keith Goffin, Pär Åhlström, Mattia Bianchi and Anders Richtnér

The practice of innovation management is developing fast. As new concepts emerge, exploratory studies are needed and case study research is often appropriate. To investigate the usage and quality of case study research in innovation management, all of the articles published in five top journals over 20 years (1997–2016) were reviewed. Case study research accounted for 818 of the published articles in this period (12%) and an evaluation template (termed case study evaluation template: CASET) was developed to objectively assess these articles against 10 quality criteria. It was found that the quality of case study research has often been low, although it has improved over time. Similarly, quality was found to fluctuate both within and between the different innovation journals. This indicates that the peer review process for case study research is not as robust as it should be. The assessment of individual articles using the evaluation template found significant deficiencies. Many articles: did not justify why case study research was appropriate; did not apply theoretical sampling criteria; were not transparent on how conclusions were drawn from the data; did not consider validity and reliability adequately; and did not go beyond description in their interpretation. However, the evaluation template also identified 23 "exemplary studies," which clearly addressed nearly every criterion. Such exemplary studies provide innovation management researchers with "benchmark" reading, which can help shape their own research. This article makes four contributions to the innovation management discipline. First, the evaluation template and exemplary studies can help innovation researchers improve the quality of their case study research. Second, clear recommendations are given for how reviewers can use the template to make the peer review process more consistent and robust. Third, journal editors are encouraged to consider the implications of the findings for their particular journal. Fourth, the article should stimulate a long overdue debate on methodology in innovation management research, including the use of case study research.

Introduction

The practice of managing innovation is fast-moving and new approaches are constantly being developed. For example, in recent years both open innovation and business model innovation have emerged as major streams of research. As new concepts appear, exploratory research taking a theory-building perspective is needed and, in such situations, case study research is highly appropriate. Consequently, it would be expected that case study research is a commonly used, highly developed methodology in innovation management studies, and that innovation researchers are engaged in active debate about the value of this methodology and how quality can be assured. Unfortunately, this is not the current situation in the innovation management discipline.

In contrast to innovation management, there are lively and ongoing debates about the role and quality of case study research in other management disciplines, such as operations management (e.g., Barratt, Choi, and Li, 2011), industrial marketing (e.g., Beverland and Lindgreen, 2010; Piekkari, Plakoyiannaki, and Welch, 2010), and information systems (e.g., Dubé and Paré, 2003). In several

Address correspondence to: Pär Åhlström, Stockholm School of Economics, House of Innovation, P.O. Box 6501, 11383 Stockholm, Sweden. E-mail: Par.Ahlstrom@hhs.se.

*Previous versions of this article were presented at the *Innovation and Product Development Management Conference* in 2012 and 2018, and in seminars at Cranfield University and Stockholm School of Economics. Both conference and seminar participants offered invaluable suggestions. The authors are also very grateful for the challenging and constructive comments from Editor-in-Chief Gloria Barczak and one anonymous reviewer.

disciplines, both the need for case study research *and* the need for higher quality case study research have been recognized (e.g., Gephart, 2004; Pratt, 2008).

Although other disciplines are discussing methodology, a study of the "topic landscape" in the *Journal of Product Innovation Management* (Antons, Kleer, and Salge, 2016) found that only one article has been published on methodology (and editorial staff have indicated that almost no methodology articles have been submitted nor did the editorial team consider methodological contributions). The one methodology study published, Perks and Roberts (2013), examined how innovation management researchers account for longitudinal effects. It found that case study research was the most commonly used method for longitudinal studies and, "[despite] advances in other disciplines, there is a lack of informed debate around longitudinality in innovation research" (p. 1102). The call from Perks and Roberts (2013) for informed debate is relevant not only for longitudinal studies but also for case studies (and methodology in general).

Against this background, the purpose of this article is to critically consider the state-of-the-art of case study research in the innovation management discipline. In order to evaluate the quality of case study research, 818 articles published in five top innovation journals over the last 20 years (1997–2016) were reviewed. What will be termed a case study *evaluation template* was developed covering four categories of criteria: *research design*, *data collection*, *data analysis*, and *post hoc reflection on rigor* (whether an article made a sufficient assessment of the quality of the research undertaken). These four categories relate to the main stages of designing and conducting high quality case study research.

The current study found that the quality of innovation case study research fluctuates widely, even within the same journal. This indicates that the "hurdle" for the quality of case study research is not clearly defined and the peer review process is not as robust as it needs to be. Many individual articles were found to be deficient in that they: did not justify why case study research was appropriate; did not apply theoretical sampling criteria; were not transparent about how the data were collected, analyzed, and conclusions drawn; and did not go beyond description in the interpretation of the results.

The remainder of this article begins with an explanation of the background and motivation for the study, including a description of the ongoing debates on case study research in other disciplines. Next, extant approaches to evaluating the quality of case study research are discussed. This is followed by the methodology of the current study, including how the 818 articles were identified, and how the evaluation template was developed and applied. The Results section presents statistics on the number and quality of articles published in the five innovation management journals, followed by an analysis of the relationships between the individual quality criteria. After this, the discussion moves to the ways in which researchers can improve case study research, with exemplary studies highlighted that can both guide and inspire researchers on how to design and present their own work. Finally, the Reflections section discusses the current role and future potential of case study research giving specific implications for case study researchers, reviewers, and journal editors.

BIOGRAPHICAL SKETCHES

Dr. Keith Goffin is an adjunct professor at the Stockholm School of Economics (SSE), where he teaches and conducts research in SSE's House of Innovation. He is also emeritus professor at Cranfield School of Management in the United Kingdom and an innovation management consultant. His research has been published in journals such as the *Journal of Operations Management* and the *Journal of Product Innovation Management*. His popular textbook *Innovation Management: Effective Strategy and Implementation* (with Rick Mitchell) is now in its 3rd edition (Palgrave, 2017).

Dr. Pär Åhlström is the Torsten and Ragnar Söderberg Professor of Operations Management at the House of Innovation of the Stockholm School of Economics, where he is also vice president degree programs. He is the co-author of the management best-seller *This Is Lean: Resolving the Efficiency Paradox* and has previously published in a range of journals including the *Journal of Product Innovation Management*, *International Journal of Operations and Production Management*, *Technovation*, and *International Journal of Innovation Management*.

Dr. Mattia Bianchi is a professor at the House of Innovation of the Stockholm School of Economics in Sweden. He holds a Ph.D. in management engineering from Politecnico di Milano, Italy. His main research areas include open innovation, agile, and lean product development. He has published several articles in leading journals such as the *Journal of Product Innovation Management*, the *Journal of International Business Studies*, and *Technovation* (www.mattiabianchi.com).

Dr. Anders Richtnér is an associate professor and CEO of the Stockholm School of Economics (SSE) Executive Education. He conducts his research at SSE's House of Innovation and its Center for Sports and Business. His research is conducted in close co-operation with companies with a clear aim of contributing to their competitiveness, but also scientific knowledge. He regularly publishes in leading academic journals such as the *Journal of Product Innovation Management*, *Research Policy*, and *MIT Sloan Management Review*.

Background to the Research

Intra-disciplinary Debates

The vast methodology literature on case study research, including famous, highly cited texts, such as Miles and Huberman (1994) and Yin (2009), stresses the importance of rigor. This recognition has led to debates on the role and quality of case study research in several management disciplines, including supply chain management (e.g., Ellram, 1996; Flynn, 2008), operations management (e.g., Barratt et al., 2011; Boyer and Swink, 2008; Voss, Tsikriktsis, and Frohlich, 2002), industrial marketing (Beverland and Lindgreen, 2010), and information systems (IS). In the IS discipline, the debate started early (cf. Benbasat, Goldstein, and Mead, 1987) and has continued, with Dubé and Paré (2003) concluding, "[c]ase research has commanded respect in the information systems (IS) discipline for at least a decade" (p. 597).

In the general management discipline, researchers have recognized that different methodologies are needed to progress knowledge. A highly influential article from Eisenhardt (1989) discussed how theory can be built using cases and the debate on case studies in general management research has continued (e.g., Eisenhardt and Graebner, 2007). Discussions have ensued on the role of qualitative studies in top journals (e.g., Gephart, 2004; Gibbert, Ruigrok, and Wicki, 2008; Pratt, 2009), and the challenges in publishing such studies have been identified through a survey of researchers' views (Pratt, 2008). In the general management discipline, several prestigious best article awards have been bestowed on case study articles (Gephart, 2004) and the methodology is becoming firmly embedded, leading Bansal and Corley (2011) to argue that qualitative research is "coming of age" (p. 233) in management research.

There are active debates in the operations and supply chain disciplines. In operations management, the debate around methodology and use of case studies goes back to the work of Meredith (1998). Voss et al. (2002) discussed the value of case studies in a special issue on research methodology and Boyer and Swink (2008) appealed for the use of a wider range of methodologies in operations management research. Recently Barratt et al. (2011) reviewed the quality of articles based on case study research in five top operations journals spanning 16 years and concluded that that "there is a lack of consistency in the way the case method has been applied" (p. 392). Ketokivi and Choi (2014) argued that case study research was undergoing a "renaissance" in operations, as the methodology is equally valid for theory generation, theory elaboration, and theory testing. In supply chain management, there has also been a methodological debate for many years with Ellram (1996) stressing when and how case study research should be used. Näslund (2002) called for more qualitative research because surveys have inherent limitations in the supply chain discipline (Ketchen, Craighead, and Li, 2018).

In contrast to other management disciplines, innovation management researchers have not debated methodology issues (c.f. Antons et al., 2016). Specifically, there has been no previous investigation of the usage and quality of case study research in innovation management.

Assessments of Case Study Quality

Arising from the debates in other disciplines, various tools and approaches have been developed to assess the quality of case study research. Table 1 summarizes the characteristics of a selection of these, which give useful ideas on how case study quality can be evaluated.

In a highly cited study, Dubé and Paré (2003) developed an assessment tool covering: research design (based on nine criteria), data collection (four criteria), and data analysis (nine criteria). A total of 183 articles from seven journals in the IS discipline were evaluated (case study research represented 15% of all articles published over 10 years). In applying the tool, each criterion was applied separately (assigning a "+" or "-") and no overall score was determined for individual articles. Dubé and Paré (2003) found: only 42% of IS articles had clear research questions, only 58% described data collection mechanisms, and only 23% explained the data analysis process adequately.

In reviewing general management research, Gibbert et al. (2008) developed 23 codes for evaluating an article's validity and reliability. A total of 159 articles in 10 leading general management journals were evaluated (case study research represented 6% of all articles published over six years). It was found that researchers publishing in top journals are not only aware of validity and reliability issues but also demonstrated how they mitigate for them. In addition, Gibbert et al. (2008) highlighted the problem that many articles claimed to have addressed validity

Table 1. Selected Studies of Qualitative/Case Study Research in Various Management Disciplines

#	Authors (Year)	Discipline/Focus	Number of Journals/Articles/ Years	Method and Quality Criteria Adopted	Quality Scoring Used?	Key Findings/Comments
1	Dubé and Paré (2003)	Information systems/Case study research	7 journals/183 articles/1990–1999 (10 years)	Highly developed set of criteria covering: • Research design • Data collection • Data analysis	Yes (but not reported as a total score per article)	• 15% of articles in information systems (IS) journals are based on case study research. • "Case research has commanded respect in the IS discipline for at least a decade" (p. 597). • Only 42% of articles give clear research questions; only 2% used piloting; only 58% clearly describe their data collections; and only 23% explain the data analysis process. • "the attributes presented … we caution researchers against treating this list as a cookbook recipe for how to do a rigorous positivist case study" (p. 628).
2	Gibbert et al. (2008)	General management/Case study research	10 journals/159 articles/1995–2000 (6 years)	Multiple-coder approach assessed each article on: • 23 coding dimensions related to validity (internal, construct, and external validity) plus reliability	Partly (but not reported as a total score per article)	• 6% of articles in top journals are based on case study research. • "authors in the highest ranked journals demonstrate that they not only are aware of the four validity and reliability criteria, they also demonstrate that they are aware of the relationships among them" (p. 1472). • "top-tier journals in our sample, such as the *Strategic Management Journal* do seem to be willing to publish high quality case study work" (p. 1473).
3	Beverland and Lindgreen, (2010)	Industrial marketing/Case study research	1 journal/105 articles/1971–2005 (35 years)	Two authors made notes on: • Quality-related steps reported in articles	Partly	• 5% of articles in industrial marketing are based on case study research. • "case quality and its associated practices varies widely" (p. 56). • Steady improvement in quality was found but only 49% of articles justified the use of case research; only 23% addressed validity; and only 16% considered reliability.
4	Bluhm et al. (2011)	Organizational science/ Qualitative research	5 journals/198 articles/2000–2009 (10 years)	Content analysis of each article (conducted by two researchers working separately) on three criteria: • Theoretical purpose • Research design • Transparency Results then compared to citations	Partly	• 15% of articles in organizational journals are based on case study research. • "more qualitative work has been published in top American management journals in the past 10 years than in the previous 20" (p. 2). • "reviewers evaluate qualitative research through a positivistic lens when a great deal of qualitative research is founded on an interpretivist tradition" (p. 2). • Research design and a clear theoretical purpose have a positive impact on an article's impact (citations). • Transparency enables publication in top journals, which leads to more citations.

(Continues)

Table 1. (Continued)

#	Authors (Year)	Discipline/Focus	Number of Journals/Articles/ Years	Method and Quality Criteria Adopted	Quality Scoring Used?	Key Findings/Comments
5	Barratt et al. (2011)	Operations management/ Case study research	5 journals/204 articles/1992–2007 (16 years)	Highly developed set of criteria covering: • Research design • Data collection • Data analysis Inter-coder agreement checked and reported	Yes (but not reported as a total score per article)	• 4% of articles in operations management journals are based on case study research. • Qualitative case studies have made contributions to the operations management discipline in terms of theory building in new areas and also from integrating existing theory with new contexts. • Details of research design, data collection and analysis were often missing. • Research protocols for doing inductive case studies are much better developed compared to the research protocols for deductive case studies.
6	Welch et al. (2012)	General management/ Qualitative research	2 journals/162 articles/1999–2011 (13 years)	Each article assessed by: • The rhetoric used to explain the theoretical purpose and contribution of qualitative studies	No	• 5% of empirical articles in top management journals are based on case study research. • Qualitative research can be used for exploratory (theory generating) research; theory development; and theory testing. • "we challenge the dominant modernist tradition that relegates qualitative research to discovery, exploration…" (p. 2).
7	Perks and Roberts (2013)	Innovation management/ Longitudinal qualitative research	10 journals/268 articles/2000–2011 (12 years)	Two authors coded the articles by: • Research method used • Data collection techniques • Approach to data analysis • Presentation of findings	No	• Percentage of articles in innovation journals based on case study research was not identified. • The most common method to study longitudinal phenomena was case studies. • "Despite such advances in other disciplines, there is a lack of informed debate around longitudinality in innovation research" (p. 1102). • "The inability to capture longitudinality in the data analysis method often led to weak presentation of … the findings" (p. 1104).

and reliability issues but failed to adequately demonstrate how this had been achieved.

In the industrial marketing discipline, Beverland and Lindgreen (2010) made notes on the quality-related steps reported in case study research articles. A total of 105 articles from one journal were reviewed (case study research represented approximately 5% of all articles published over 35 years and rising in recent years). Beverland and Lindgreen (2010) reported their results per topic and individual articles' overall quality "scores" were not determined. It was found that "case quality and its associated practices varies widely" (p. 56), although there was a steady improvement over time. Only 49% of marketing articles were found to have justified the selection of case study research, only 23% addressed validity, and only 16% considered reliability.

In the organizational science discipline, Bluhm, Harman, Lee, and Mitchell (2011) developed a tool to evaluate theoretical purpose ("generation," "elaboration," or "testing"), research design ("interview," "observation," "archival," "questionnaire," and "miscellaneous"), transparency, and citations. In their investigation, Bluhm et al. (2011) defined transparency as "whether the article reported sufficient information in both data collection and analysis for the study to be replicated to a reasonable extent" (p. 9). A total of 198 articles from five journals were evaluated (case study research represented 15% of all articles published over 10 years). It was found that research design and a clear theoretical purpose had a positive impact on an article's impact (citations), whereas transparency enabled publication in top journals, which in turn led to more citations.

In the operations management discipline, Barratt et al. (2011) based their evaluation criteria around research design, data collection, and data analysis. A total of 204 articles from five journals were evaluated (case study research represented approximately 4% of all articles published over 16 years). It was found that case study research had made contributions to the discipline but many articles did not provide sufficient detail on research design, data collection, and data analysis.

Welch, Plakoyiannaki, Piekkari, and Paavilainen-Mäntymäki (2012) studied the rhetoric used in justifying the use of qualitative research in the organizational sciences but did not look at wider quality issues. A total of 162 articles were evaluated in two journals (case study research represented 15% of all empirical articles published over 13 years). They found that qualitative research is not only useful for exploratory (theory-generating) research but it is also valid for both the development of theory and even theory-testing. This led Welch et al. (2012) to "challenge the dominant modernist tradition that relegates qualitative research to discovery, exploration…" (p. 2).

Finally, Perks and Roberts (2013) used three criteria to assess the quality of articles: the data collection techniques, the approach to data analysis and the way findings were presented. This study evaluated the highest number of articles: a total of 268 in 10 journals (unfortunately, the percentage compared to the total number of articles published over 12 years was not identified by Perks and Roberts [2013]). It was found that longitudinal studies commonly used case study research but were very often weak in both data analysis and presentation.

Comparing the different articles that have developed evaluation methods (see Table 1), a number of conclusions can be reached:

- The main categories developed for evaluating the quality of case study research are research design (including the justification for case study research and selection of cases); data collection methods; data analysis; and consideration of validity and reliability issues.
- Although scoring mechanisms (coding) have been developed, these have not been applied to evaluate the overall quality of individual articles.
- Coding was often conducted by multiple researchers and inter-coder agreement was sometimes checked (e.g., Barratt et al., 2011).
- Previous investigations have evaluated between 105 and 268 articles, accounting for periods of between 6 and 35 years, and covering between 1 and 10 journals.

Aims of This Study

The articles discussed above were all instrumental in establishing the necessity, purpose, and aims of the current research. The purpose was defined as a critical assessment of the state of case study research in the innovation management discipline. Specifically, the aims were:

1. To identify all of the case study articles published in top innovation journals from 1997–2016 (20 years), as a proportion of the total number of research articles published.
2. To develop an evaluation template for assessing the quality of case study research based on insights from the literature.
3. To apply the evaluation template to assess the quality of each of the case study articles identified, generating statistics at the journal and aggregate level, and against each of the evaluation criteria.
4. To generate guidelines on how researchers can make their case study research more rigorous, identifying best practices and exemplary studies against which innovation management researchers can benchmark their own work.
5. To identify the implications for researchers, reviewers, and journal editors.

Methodology—Identification of Journals and Articles

Identification of Relevant Journals

It was decided to focus on five leading innovation management journals, over 20 years (1997–2016). First, articles discussing innovation management journals (e.g., Linton and Thongpapanl, 2004; Thieme, 2007) were used to identify top journals. Then to select the most influential five, the 2015 five-year impact factors (IF) were used. This led to the selection of:

- *Research Policy* [designated RESPOL in some later tables] (IF = 5.118)
- *Technovation* [TECH] (IF = 3.833)
- the *Journal of Product Innovation Management* [JPIM] (IF = 3.178)
- *Technological Forecasting and Social Change* [TFSC] (IF = 3.005)
- *R&D Management* [RDM] (IF = 2.470)

Identification of Relevant Articles

Keyword searches to identify all of the case study research articles published in a particular journal are known to be unreliable (Welch et al., 2012), and were found to be unreliable in this investigation. Therefore, an encompassing approach was used, manually browsing the articles in each issue of every journal over the 1997–2016 period. When articles were identified that were possibly based on case study research, abstracts were perused for evidence of case study research. If the use of case study research could not be determined from the title and abstract, the actual article was reviewed. Having identified an initial set of case study articles, each was read thoroughly, using the inclusion and exclusion criteria set out below, generating a population of 818 articles.

Following definitions from Yin (2009), Barratt et al. (2011), and others, a broad definition of a case study was used: empirical research that primarily uses contextually rich data from bounded real-world settings to investigate a focused phenomenon. It should be noted that Yin (2009) did not regard research based on only one source of data as sufficient to be considered as case study research. Although the authors of the current article sympathize with Yin's view, this criterion was not applied as a significant portion of published articles (30%) were found to be based on a single source. Piekkari et al. (2010) also found that Yin's perspective that case study research must incorporate multiple sources of data, "proved insufficient to capture the complexity of [current] research practice" (p. 112).

As case study research covers different types of studies, criteria for which articles to "include" and which to "exclude" had to be developed. In general, the aim was to be inclusive and to include all articles where the author(s) stated they had used case study research. Thus, articles were included that were descriptive, sometimes only reporting on company practices, often without adequate discussion on theory or methodology. Articles that had developed a methodology or tool and then illustrated or applied it using one or more cases, were also included. The purpose of being inclusive was to identify the range of practices of those researchers purporting to employ the case study methodology, and thus build on a view of what the case study research means in practice.

Some articles were identified that were based on both qualitative and quantitative data. These were included when case studies were a substantial part of the research (e.g., Candi, 2010), or when the case study part was a specific stage of the research, such as developing ideas for a survey (e.g., Cristiano, Liker, and White, 2000). It is important to note, however, that in rating the quality of these types of articles, only the part that pertained to the case study research was evaluated.

Some types of articles were excluded. Articles in the track "From Experience" in the *Journal of Product Innovation Management* (e.g., Riek, 2001) were excluded because these were clearly labeled as not based on research (often written by company executives rather than researchers). Articles based on interviews, where the qualitative data were analyzed at the individual level only, were also excluded. These articles were more like survey research, only data were collected through qualitative interviews (e.g., Gemser and Leenders, 2001). A third type of article that was excluded was action research (e.g., Drejer and Gudmundsson, 2002), as this is normally viewed as distinct from case study research, being based on a different epistemology and ontology (Coughlan and Coghlan, 2002). Articles were also excluded where qualitative data only supported the quantitative data, such as modeling articles with case applications. Finally, despite being published in innovation management journals, it was decided to exclude a handful of articles that were not considered to deal with innovation management. Examples of articles that were excluded are: the implementation of employment policies in the Azores (Bettencourt, 2010), and the challenges in fighting the 2009 Victoria bushfire (Oloruntoba, 2013).

Methodology—Developing the Evaluation Template

The development of the case study evaluation template (CASET) was based on four decisions. First, although the assessment of whether case study articles have addressed particular criteria has previously been made at the aggregate level (i.e., as a percentage across all case study articles), it was decided that it was important to evaluate individual articles and generate an overall quality score for each article. Second, similar to Gibbert et al. (2008), it was decided to code articles according to whether they did or did not apply specific steps (practices) to ensure research quality. Thus, articles were coded against each of the 10 criteria in the template as a "1" (did apply practices) or "0" (did not). Third, to make CASET easy to apply, individual articles were assigned an overall quality score from "0" to "10." Fourth, it was deemed appropriate that the template layout should fit on one page (albeit with a small font).

Validity and reliability issues were carefully considered during the development of CASET. To ensure validity, the template was based on previous case study assessment criteria found in other disciplines (see Table 1), supplemented by further articles discussing case study design (e.g., Barratt et al., 2011; Benbasat et al., 1987; Eisenhardt, 1989; Eisenhardt and Graebner, 2007; Gephart, 2004; Pratt, 2008; Voss et al., 2002) and classic texts (Miles and Huberman, 1994; Yin, 2009).

High reliability was ensured by independent coding followed by inter-coder reliability checks, and developing robust anchoring statements. During its development, two different versions of CASET were presented at the International Product Development Management Conference and improvements were made based on the feedback from delegates. The template was also presented at research seminars conducted at two different academic institutions. In addition, challenging and useful feedback was received during the review process of the *Journal of Product Innovation Management*.

A positivistic philosophical perspective underlies CASET. Here, it is argued that innovation management researchers and journals, by and large, adhere to more positivistic view of the world, perhaps due to the historical roots of the discipline. Dubé and Paré (2003) argued that the IS discipline requires rigorous case study research in the vein of positivism. Some researchers do, however, reject the "positivistic" evaluation of case study research, claiming that case study research is often conducted in a constructivist paradigm and should not be measured against criteria that do not apply (Piekkari et al., 2010). Yet, in their review of case study articles published in top general management journals, Gibbert et al. (2008) found that there was a large overlap among the concrete actions taken to ensure rigor, regardless of the philosophical perspective taken by researchers.

It should be noted that CASET does not cover the contribution of a case study article to knowledge. It was decided that to try to measure this would be too complex and would require the research team to have a detailed understanding of every topic within innovation management research. This was not realistic and so no assessment of contribution was made. That means that CASET can be used to assess the *quality* of case study research but not the *contribution* of that research to a particular innovation management topic.

The final version of CASET is shown in Table 2. A multiple-page version with additional explanatory

J PROD INNOV MANAG
2019;0(0):9–30

Table 2. CASET Evaluation Template for Scoring Case Study Articles (Rating Quality on a 0–10 Scale)

CASET - Evaluation Template for Case Study Articles

Author(s)		Article title		Journal		Year of publication	

Number of cases
Note the number of cases studied in the article

Number of data sources
Note how many sources of data were collected, based on the following five sources:
i. interviews (face to face, including unstructured/semi-structured/structured interviews and informal conversations)
ii. surveys
iii. observations (including site visits, workshops, ethnography, data feedback sessions)
iv. internal documents
v. secondary data

	Evaluation criteria	Explanation of measure	Anchoring Statements	Score (0/1)
Research Design	*Theoretical foundation*	Was a clear explanation given of why the case method was the most appropriate method to adopt?	0 = "No": no explicit argument was given for why the case method was adopted in the research. 1 = "Yes": there was an explicit argument for why the case method was adopted in the research.	
	Pilot study	Was there a pilot study proceeding the main study?	0 = "No": there was no pilot study. 1 = "Yes": a pilot study was conducted before the main study.	
	Theoretical sampling	Was an explanation provided of which case(s) were chosen and why?	0 = "No": no explicit argument was given about how the case(s) was / were selected. 1 = "Yes": case(s) were selected for theoretical purposes, example ranges from a discussion on why case(s) were chosen to a discussion on the selection of polar extremes where cases exhibited extremely high or extremely low value on the constructs of interest.	
Data Collection	*Triangulation*	Was the research based on multiple sources of data?	0 = "No": the research was based on only one source of data 1 = "Yes": the research was based on more than one source of data	
	Review and validation of evidence	Was the evidence reviewed and validated by external parties?	0 = "No": the evidence was not reviewed and validated. The article did not explicitly state if the evidence is reviewed and validated. 1 = "Yes": the evidence was reviewed and validated by the interviewee and/or the company (e.g. through data feedback-sessions); or by fellow researchers not part of the primary data collection.	
	Transparency of data collection	Was it made clear how the data collection process was conducted?	0 = "No": the data collection process was not clear and transparent because there was not sufficient information about the origin and the contents of the data collected (in terms of areas, topics, themes or constructs), which would allow replication. 1 = "Yes": the data collection process was clear and transparent because interview themes, questions and/or research instruments such as research protocols specifying data collection circumstances were reported, which would allow replication.	
Data Analysis	*Inter-coder agreement*	Were the data coded by multiple investigators?	0 = "No": the data were not coded by multiple investigators working independently, or there was no information about how inter-coder agreement was achieved. 1 = "Yes": the data were coded by multiple investigators working independently, and there was an explanation about how an acceptable inter-coder agreement was achieved.	
	Case presentation	Were findings and empirical evidence presented in a way that made it clear how the author(s) reach their conclusions?	0 = "No" - The way in which the author(s) reached their conclusions based on the case data was neither clear nor documented. Their focus was on "telling the story" and not "showing the evidence", and any quotes used were selected to support the authors' conclusions. 1 = "Yes" – The article was explicit and clear in demonstrating how the empirical data were used to arrive at the conclusions, providing a clear "trail of evidence" (through the use of approaches such as tables, data displays, coding schemes and other visual aids).	
	Case interpretation	Did the case analysis move beyond description and conceptual ordering?	0 = "No" – The results from the case analysis were mostly descriptive and/or simply consisted of condensing data into patterns and concepts. 1 = "Yes" – The interpretation moved beyond description and conceptual ordering, to the generation of meaning and of the conceptual significance of the case facts. This was achieved by, for example, developing a conceptual framework or model from the case(s), formulating propositions to be tested by future research, and/or systematically discussing results in relation to existing literature.	
Post-hoc	*Reflecting on validity and reliability*	Was there a discussion about the quality of the research?	0 = "No": there was no explicit discussion about the quality of the research. 1 = "Yes": there was an explicit discussion reflecting on the quality of the research (either in the section on research design stage or in the consideration of limitations), which covered one or more dimensions of validity and reliability, showing that authors were aware of the need to ensure rigor.	
			Overall quality score (out of 10)	

notes is available online (www.hhs.se/casestudyobservatory).

Template Layout

The top section of the template collects demographic information about an article, including the number of cases and of data sources. As regards the former, the choice of the number of cases is crucial (Drejer, Blackmon, and Voss, 2000), as it influences both the depth of observation possible (Voss et al., 2002) and the theoretical contribution of the research. Single case studies are often useful for longitudinal research (Voss et al., 2002) and can also be used if they provide extreme exemplars, or opportunities for unusual research based on particular access (Barratt et al., 2011). With multiple cases, it can be argued that external validity is higher, and the opportunities for creating more robust and testable theory are greater than for single cases (Eisenhardt and Graebner, 2007). Eisenhardt (1989) suggested that a suitable number of cases in most case study research is between 4 and 10, enabling theoretical saturation to be achieved. In CASET, the number of cases used is noted but not rated, recognizing that the number of cases chosen should match the aims of the research and the way it is conducted. More cases are not necessarily better (or worse). It is indeed possible to conduct high quality case research irrespective of number of cases.

The top section of the template also collects information on the number of data sources employed. There are many different sources of data that can be used (Yin, 2009). In general, five main categories can be identified: interviews, surveys, observations, internal documents, and secondary data. The number of data sources used is noted but not rated in CASET. The rationale is that that the number of data sources employed depends on the particular nature of the case study setting, of the subject being investigated and on data availability. What is important from a quality standpoint is that more data collected from multiple sources allows triangulation to be conducted. This criterion was, similar to all others, operationalized as a dichotomous variable and included in the template under "triangulation" (explained later).

The core section of CASET is structured according to the four categories: research design, data collection, data analysis, and post hoc reflection on rigor. These categories reflect the generic steps taken when conducting research: designing the research, collecting the data, analyzing the data, and finally reflecting on the research quality. Although the way research is conducted in practice is seldom linear, it was decided to divide the template up into the main (iterative) steps identified.

Aligned to the four categories are a total of 10 subcategories (criteria), against which articles are rated as a "0" or "1." A set of anchoring statements was developed so that articles could be reliably assessed against each criterion. The full evaluation of each article against all 10 criteria led to an overall quality score of between "0" (extremely low quality) and "10" (exceptional quality) for each article. Each of the criteria will be explained in the following sections, which should be read in conjunction with Table 2.

Category #1—Research Design

This category covered three criteria:

1. Whether there was a clear explanation of why it was appropriate to adopt the case study methodology.
2. Whether a pilot study was conducted.
3. Whether theoretical sampling was used in selecting case(s).

Theoretical foundation. Following the notion of methodological fit (Edmondson and McManus, 2007), it is crucial for researchers to articulate explicitly why case study research was appropriate (Eisenhardt and Graebner, 2007), and why it was chosen in preference to other methodologies (Barratt et al., 2011). For example, it might be argued that the research is exploratory (Yin, 2009) but it could also be that the context is unique and so the research can generate novel insights (Benbasat et al., 1987). Furthermore, the phenomena(on) being examined might not have been adequately explained by existing theory (Eisenhardt and Graebner, 2007). It should be noted that case study research is not only useful for exploratory research but also theory-building, theory-elaboration, and theory-testing research (Ketokivi and Choi, 2014; Welch et al., 2012).

Pilot study. Conducting a pilot study prior to the main case studies can be an important way of testing, revising, and sharpening research protocols, interviews, observation guides, and the like (Dubé and Paré, 2003). Even single cases can benefit from "piloting," for example, data collection tools can be tested in the first interviews with managers. Miles and Huberman (1994,

p. 38) pointed out that data collection "instrumentation can be revised—in fact, should be revised," to ensure valid data collection. In the IS discipline pilot studies were found to be very rare (only 2% of articles; Dubé and Paré, 2003) and most previous investigations of case study quality overlooked this criterion (cf. Table 1).

Theoretical sampling. When conducting case study research, researchers should provide an explanation of how and why particular case(s) were chosen. Benbasat et al. (1987) stressed that cases should be chosen carefully and not opportunistically. In contrast to the statistical sampling used in surveys, case study research chooses cases for theoretical reasons (Eisenhardt, 1989; Glaser and Strauss, 1967; Meredith, 1998). For example, cases can be chosen to demonstrate similar results, or to generate contrary results (Yin, 2009), or to include cases which exhibit extremely high or low values on the constructs of interest.

Category #2—Data Collection

This category covered three criteria:

1. Whether data from multiple sources were collected to enable triangulation.
2. Whether the evidence was reviewed and validated externally.
3. Whether the data collection process was transparent.

Triangulation. Through utilizing multiple sources of data, researchers have an opportunity for triangulation. Yin (2009) considered multiple sources of data as essential to case study research. The use of multiple data sources increases the internal and construct validity of research (Benbasat et al., 1987), as the "triangulation made possible by multiple data collection methods provides stronger substantiation of constructs and hypotheses" (Eisenhardt, 1989, p. 538). In addition to enabling triangulation, different sources of data, due to their distinct nature, can generate different insights (Gibbert and Ruigrok, 2010). It should be noted that the availability of multiple sources of data enables triangulation without guaranteeing that it was conducted.

Review and validation of evidence. It is important to have a review and validation of the evidence that case study research generates. The review should be conducted by external parties, and examples include validation by the interviewee(s), the company (e.g., through data feedback sessions), or fellow researchers not part of the primary data collection. Such reviews can help avoid researcher bias and subsequent misinterpretations. They can also allow researchers to identify deeper findings (Miles, Huberman, and Saldana, 2014). Having independent reviews of the evidence can also lead to unexpected findings (Gioia, Corley, and Hamilton, 2012) and has been called "giving voice to participants" (Bluhm et al., 2011, p. 1870).

Transparency of data collection. This criterion is concerned with whether the data collection process was conducted in a clear and transparent way (including whether instruments such as interview questions and research protocols were included in an article, its appendices, or were available online). In addition, it is essential that the circumstances around data collection (e.g., information on the characteristics of interviewees and of data gathering sessions) are specified (Yin, 2009). Careful documentation and clarification of the data collection procedures is important to ensure the reliability of case study research (Gibbert and Ruigrok, 2010). It should be possible from the information provided in a study to both understand the logic and purpose of the research actions taken, assess the type of data collected, and replicate data collection (Leonard-Barton, 1990).

Category #3—Data Analysis

This category covered three criteria:

1. Whether inter-coder checks were applied.
2. Whether the case study evidence was clearly presented.
3. Whether the case study results were appropriately interpreted.

Inter-coder agreement. In order to increase rigor, it is beneficial to have multiple researchers code the data independently, as then inter-coder agreement can be determined (Barratt et al., 2011). Involving multiple investigators is also a form of triangulation as it helps handle the richness of contextual data and lends more confidence to the findings of the research (Benbasat et al., 1987; Eisenhardt, 1989). When multiple investigators are involved, it is necessary to explicitly describe the process by which an acceptable level of

inter-coder agreement was achieved. According to Miles et al. (2014), an acceptable level of inter-coder agreement is within the range 85% to 90%.

Case presentation. One of the biggest challenges of case study research is to demonstrate how data were analyzed and conclusions reached (Eisenhardt, 1989; Miles and Huberman, 1994). If the process of analysis and reporting is not systematic, then the results are prone to be weak (Gephart, 2004). Presentation includes documenting the coding scheme(s) and process(es), documenting each step, and describing the analysis techniques used (e.g., summarizing field notes, coding of raw data, data displays, etc.). The way case study data are analyzed and presented has a fundamental impact on whether research will be accepted by top journals (Gibbert et al., 2008).

There is a need to present data carefully and systematically, for example, through synoptic tables and exhibits, designed to provide a trail of evidence, and ensure that the reader understands how the researchers reached their conclusion (Miles and Huberman, 1994). An appropriate balance needs to be struck between data and interpretation (Pratt, 2009). An engaging story is not enough; it needs to be supported by data, demonstrating how the evidence led to the case study findings.

Case interpretation. Case study research needs to be written in a solid yet engaging way but it is also necessary to complement the careful presentation of data with a "drive toward some new concept development and theoretical discovery" (Gioia et al., 2012, p. 23). This entails moving beyond either describing the data or condensing them into patterns and concepts (what an insightful reviewer termed "conceptual ordering"), toward theorizing (Yin, 2009). Conceptual ordering is typically achieved through a host of tactics for drawing conclusions from case study data, including clustering, counting, partitioning variables, and subsuming particulars into the general (Miles and Huberman, 1994). Theorizing needs to go further and requires activities such as abstracting, generalizing, relating, selecting, explaining, synthesizing, and idealizing (Weick, 1995). In the iterative process of theorizing (Eisenhardt, 1989), it is important to look for relationships between variables, finding intervening variables, and building a logical chain of evidence. Theorizing should also lead to concrete outputs, such as the development of a conceptual framework(s) or model(s) from the case(s), or the formulation of propositions to be tested in future research. Finally, it is imperative to discuss the outcomes in relation to the extant literature.

Category #4—Post hoc Reflection on Rigor

Reflecting on validity and reliability. The final criterion in CASET was whether the author(s) of a case study article had reflected on the rigor of their research. Such reflection is imperative in case study research. The reflection needs to specifically consider the measures taken to ensure *validity* (that the results are correct for the cases investigated) and *reliability* (that the study could be replicated). Following the ideas of Gibbert et al. (2008), the criteria in the CASET which mainly address validity are theoretical foundation, pilot study, theoretical sampling, triangulation, review and validation of evidence, case presentation, and case interpretation. Similarly, CASET assesses reliability through the transparency of data collection and inter-coder agreement. In published studies, reflections on rigor (if included at all) are typically reported in the section describing methodology and/or in the final section outlining limitations.

Validity is a broad term and it includes several elements. First, it covers *construct validity*—whether a study investigated what it claimed to investigate using appropriate operational measures. Next, it includes *internal validity*—whether the findings on causal relationships between variables identified through triangulation and *pattern-matching* are sound. (Pattern matching refers to the practice of comparing empirically observed patterns with either predicted ones, or patterns established in prior studies and in different contexts [Eisenhardt, 1989; Gibbert and Ruigrok, 2010].) The third aspect is *external validity*—whether the findings from the specific cases are applicable elsewhere and whether there was literal or theoretical replication. Reliability is whether a study is replicable in that other researchers following the same steps would arrive at the same insights.

Gibbert and Ruigrok (2010) stressed that all three types of validity plus reliability must be considered. However, they found that general management studies based on case study research did not do this. In fact, external validity was only discussed in 20% of studies, reliability in 10%, internal validity in 5%, and construct validity in 2%. These low figures indicate

very low awareness among case study researchers on the value of reflecting on rigor.

Honing the Evaluation Template

To ensure that CASET was robust, it was subjected to five stages of development:

1. *Calibration:* This stage aimed to calibrate the template and to familiarize the research team (the four authors of this article) with the process of evaluating articles (by reading and coding them against the 10 criteria). Here, 40 articles were rated by the research team and their results compared. This led to a number of clarifications and enhancements to the template.
2. *Inter-coder Reliability Checks:* Next, 25% of all case study articles from the *Journal of Product Innovation Management* were coded by two authors working alone and the results compared. The inter-coder reliability of this process was 85% (defined as: 100 × the total number of agreements/ [total number of agreements + disagreements]). All disagreements were then discussed and resolved by the complete research team, which also led to further enhancements to the template. Then, the third author read and rated five randomly selected articles out of the pool of articles using the updated version of CASET. The result showed a 91% correspondence with the other two authors' coding and established that the template could be reliably applied by others.
3. *Evaluating Articles:* The third step involved dividing the remaining articles up between the four authors in the research team. Each author received over 200 articles: an equal number from each of the five journals, chronologically spread across the 20 years considered. Each author then rated the articles but consulted the others where problems were found. For example, it was found that many articles did not clearly state their data sources and often only implicitly indicated that documentary evidence was used. For such articles, it was agreed that a single source of data would be assumed, if the author(s) did not explicitly identify additional sources. Although shared by four authors, the evaluation of 818 articles was a massive task. However, it enabled a comprehensive view of case study research in innovation management to be derived, and the scope of the current study is more than comparable with previous investigations in other disciplines (cf. Table 1). It should be noted that there were a number of case study articles where members of the research team were the authors or co-authors of the article to be evaluated. To ensure objectivity, the coding process for these articles excluded the authors and co-authors.
4. *Creation of the data set:* As each article was coded against each criterion, the values were entered into an Excel spreadsheet, to enable the calculation of overall quality scores and further quantitative analysis. The research team also added notes on articles which addressed individual criteria in a rigorous and noteworthy way.
5. *Template Enhancements and Subsequent Recoding of Articles.* Feedback on CASET was received from conference attendees, colleagues, a journal editor, and an anonymous reviewer. Their views led to significant enhancements of the template. In particular, they led to a reconceptualization of the criteria case presentation and case interpretation (and a sharper differentiation between them). These changes made it necessary to recode all of the 818 articles. The same procedure for checking inter-coder reliability (described in point 2) was followed to test the new version of CASET. Once the reliability of the template was established, the remaining articles were divided between the four authors and rated. As a consequence of the recoding, nearly every article was assessed by at least two authors.

Results

The results, based on the rich data from coding 818 case study research articles, will be presented under the following six headings:

1. Number of case study articles
2. Overall quality of case study articles
3. Demographics and quality
4. Chronological development of quality
5. Individual quality criteria and lessons for researchers
6. Relationships among quality criteria and identification of exemplary studies

Number of Case Study Articles

Table 3 gives the number and proportion of case study articles in the five journals investigated. In calculating the proportion of case study articles, all special issues were included but editorials and book reviews were excluded.

It can be seen that, across all five journals, the average proportion of case study research articles was 12%. *R&D Management* published the highest percentage of cases (28%); *Technological Forecasting and Social Change* the lowest (5%); and in the *Journal of Product Innovation Management,* 13% of articles were based on case study research. It should be noted that *Technological Forecasting and Social Change* had four years (1998, 2002, 2003, and 2005) where it did not publish any case study articles.

Figure 1 shows how the proportion of case study research articles has changed over the 20 years 1997–2016. Overall, the proportion has been fairly stable during this period, with a range normally from 10% to 15% (across the five journals), peaking in 2006 at 17%. *R&D Management* shows large fluctuations from year to year. A downward trend in the years 2015 and 2016 can be noted, with the percentage of case study articles falling below 10% (for the *Journal of Product Innovation Management*, this trend was pronounced). These results suggest that case study research is not the prevalent methodology in articles published in the innovation management discipline.

Overall Quality of Case Study Articles

Table 4 gives statistics for the overall quality score of case study articles. It also shows the proportion of articles across all different quality levels, ranging from a minimum of "0" to a maximum of "10," for the five journals. Over the whole period 1997–2016, articles scored on average 3.05 but the most common quality score for individual articles in the sample was only "1." This indicates that the quality of many innovation management case study research articles was low, as the maximum score achievable by an article is "10" (although no articles in the population achieved the maximum-possible quality score—see Table 4). It is unfortunate that the majority of case study research in the innovation management discipline appears to be poorly designed and conducted.

Looking at the quality scores of the different journals, it was found that articles in the *Journal of Product Innovation Management* scored highest (4.75 on average), followed by *R&D Management* (3.14), *Research Policy* (2.96), *Technological Forecasting & Social Change* (2.59), and *Technovation* (2.44). To determine whether the quality differences across the journals were statistically significant, the Kruskal–Wallis Test was applied (a rank-based, nonparametric, and one-way analysis of variance), which was appropriate given the ordinal nature of the variable for overall quality score. Due to the nature of this test, it only revealed that at least one of the five journals differed significantly from at least one other journal in terms of overall quality. In order to determine whether there were significantly differences in pairwise quality across the five journals, Dunn's post hoc multiple comparison test was performed (see bottom of Table 4). Articles in the *Journal of Product Innovation Management* were found to have significantly higher quality than those in the other four journals. Articles in *R&D Management* scored significantly higher than those in both *Technological Forecasting & Social Change* and *Technovation*, but there was no significant difference in quality to articles published in *Research Policy*. Similarly, there was no significant difference in quality between articles in *Technological Forecasting & Social Change* and *Technovation*. Overall, it was surprising that statistically significant differences in overall quality scores were found between the five top-rated, high-impact innovation management journals.

From Table 4 it can be seen that high-scoring articles, for example, those scoring "8" and above, are very rare—only 3% of the population. From these, the *Journal of Product Innovation Management* was found to have the highest share, 10%; *R&D Management* was found to have 3%; *Research Policy* and *Technovation* both had only 2%; while *Technological Forecasting & Social Change* was found to have no articles with a score of "8" or over.

Demographics and Quality

It was also interesting to examine how the overall quality score was related to the demographic attributes; the number of cases and number of data sources. The results for the number of cases are reported in Table 5 and it can be seen that single case studies were by far the most common type of

Table 3. Number of Case Study Articles Published by Journal from 1997–2016

	JPIM	RDM	RESPOL	TECH	TFSC	Total
Total number of research articles	898	695	2057	1338	1983	6971
Number of case-based articles	115	192	164	246	101	818
Proportion of case study research (%)	13%	28%	8%	18%	5%	12%

Abbreviations: JPIM is the *Journal of Product Innovation Management;* RDM is *R&D Management;* RESPOL is *Research Policy*; TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*

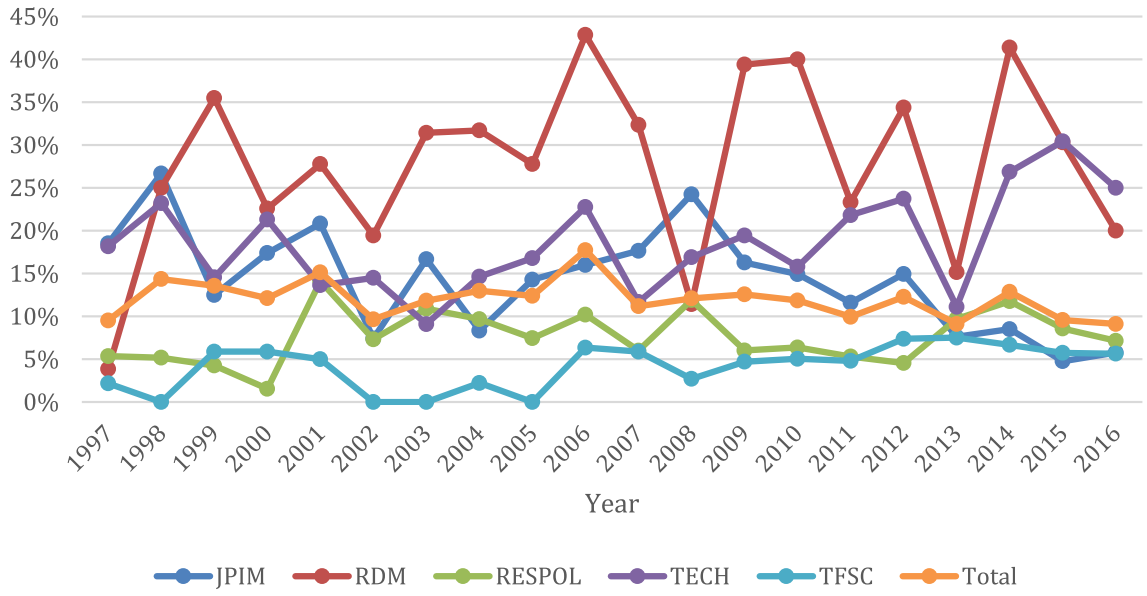

Figure 1. Proportion of Case Study Articles in Top Innovation Management Journals (1997–2016)
Abbreviations: JPIM is the *Journal of Product Innovation Management*; RDM is *R&D Management*; RESPOL is *Research Policy*; TECH is *Technovation*; TFSC is *Technological Forecasting and Social Change*.

Table 4. Means, Medians, and Proportion of Articles across the Different Levels of the Overall Quality Score.

Overall Quality Score	JPIM (*n* = 115)	RDM (*n* = 192)	RESPOL (*n* = 164)	TECH (*n* = 246)	TFSC (*n* = 101)	Total (*n* = 818)
0	3%	14%	12%	17%	16%	13%
1	5%	16%	18%	23%	17%	17%
2	8%	13%	19%	16%	21%	16%
3	12%	14%	15%	16%	16%	15%
4	18%	15%	13%	12%	10%	13%
5	15%	11%	8%	8%	10%	10%
6	14%	6%	8%	4%	5%	7%
7	15%	8%	5%	2%	5%	6%
8	7%	3%	2%	1%	0%	2%
9	3%	0%	0%	1%	0%	1%
10	0%	0%	0%	0%	0%	0%
Median	5	3	3	2	2	3
Mean	4.75	3.14	2.96	2.44	2.59	3.05
Kruskal–Wallis test χ^2			82.771***			
Statistically significant mean comparisons			JPIM versus RDM***; JPIM versus RESPOL***; JPIM versus TECH***; JPIM versus TFSC***; RDM versus TECH***; RDM versus TFSC*; RP versus TECH**			

Abbreviations: JPIM is the *Journal of Product Innovation Management;* RDM is *R&D Management;* RESPOL is *Research Policy*; TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*
***$p < .001$; **$p < .01$; *$p < .05$; $^{\dagger} < .1$.

Table 5. Proportion of Articles per Number of Cases

Number of Cases	JPIM	RDM	RESPOL	TECH	TFSC	Total	Overall Quality Score Mean	Jonckheere–Terpstra Test z	Statistically Significant Mean Comparisons
1	26%	41%	39%	38%	54%	39%	2.43	8.041***	1 versus 4***; 1 versus 5**; 1 versus 6 to 8***; 1 versus 9 or more***
2	9%	11%	20%	14%	9%	13%	2.66		2 versus 4**; 1 versus 5**; 1 versus 6 to 8***; 1 versus 9 or more***
3	3%	9%	9%	10%	8%	8%	2.70		3 versus 4**; 1 versus 5**; 1 versus 6 to 8***; 1 versus 9 or more***
4	8%	7%	7%	8%	10%	8%	3.62		
5	7%	2%	2%	3%	4%	4%	3.93		
6 to 8	16%	16%	10%	11%	5%	12%	3.90		
9 or more	31%	14%	13%	16%	10%	16%	3.92		

Abbreviations: JPIM is the *Journal of Product Innovation Management;* RDM is *R&D Management;* RESPOL is *Research Policy*; TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*
***$p < .001$; **$p < .01$; *$p < .05$; $^{\dagger} < .1$.

research, accounting for 39% of all articles. This holds true for all journals, except for the *Journal of Product Innovation Management*, where 31% of articles were based on nine cases or more. Multiple case studies are more common in that journal and it is the only one of the five journals where articles based on five or more cases accounted for the majority of published articles.

Table 5 also shows that articles based on more cases appear to be associated with higher overall quality scores with, for example, articles based on two cases scoring 2.66 and those based on four cases scoring 3.62 on average. To test this observation, a Jonckheere–Terpstra test was used and this found that articles based on more cases *are* associated with statistically higher-quality case study research. Stated in another way, most case study research (39% of the published articles) was based on single cases and many of these were poorly designed and conducted, achieving an overall quality score of 2.43 on average.

Interestingly, Dunn's test for multiple comparison showed that there was no significant difference in overall quality between articles based on one, two, and three case studies. Similarly, no significant differences were found between articles based on four, five, six to eight, and more than eight case studies. However, the group of studies based on up to three cases showed significantly lower quality scores than the group with more than three. Here, the findings are surprisingly consistent with the recommendations of Eisenhardt (1989), who argued that a suitable number of cases in most research contexts is four or more. It is, of course, perfectly feasible to conduct high quality case study research with only a few cases. However, it seems that researchers who are electing to conduct studies based on single or few cases, are less aware of the issues that need to be considered in conducting high quality case study research. There is a particular danger that single case studies will attract a poor reputation, whereas it is the way they are conducted not the fact they are based on a single case that is the problem (although some management journals may have a preference for multiple cases rather than single ones).

In Table 6, the distribution of articles according to the number of data sources used is given. This shows that 30% of the articles were based on a single source of data. The overwhelming majority of these studies were based on interviews (with just a few studies based on historical data). It can also be seen that 94% of articles used between one and three sources of data, 5% of articles used four sources of data, while only 1% (three articles) used five sources. When studies were based on multiple sources of data, interviews were, with few exceptions, one of the sources of data used.

Applying Jonckheere–Terpstra's and Dunn's tests showed that higher overall quality scores were associated with a higher number of data sources. Only the 1% of articles based on five sources were not significantly different in terms of quality compared to

Table 6. Proportion of Articles per Number of Data Sources

Number of Sources	JPIM	RDM	RESPOL	TECH	TFSC	Total	Overall Quality Score Mean	Jonckheere–Terpstra Test z	Statistically Significant Mean Comparisons
1	25%	35%	23%	34%	30%	30%	1.42	15.097***	1 versus 2***; 1 versus 3***; 1 versus 4***; 1 versus 5*
2	40%	33%	39%	39%	35%	37%	3.39		2 versus 3***; 2 versus 4**
3	30%	26%	29%	23%	30%	27%	4.03		3 versus 4*
4	5%	5%	8%	3%	5%	5%	5.02		
5	0%	1%	1%	1%	0%	1%	3.67		

Abbreviations: JPIM is the *Journal of Product Innovation Management;* RDM is *R&D Management;* RESPOL is *Research Policy*; TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*
***$p < .001$; **$p < .01$; *$p < .05$; †$< .1$.

those based on fewer sources of data. (This result was most likely due to the extremely low number of observations in this category.) So, overall it appears that the researchers electing to use multiple sources of data are those researchers that are more aware of quality issues in case study research (and take actions to address these issues).

Chronological Development of Quality

Figure 2 depicts how the quality of case study research has changed over the period 1997–2016. A positive trend can be observed, with the average quality ("total") having increased from a score around "2" at the end of 1990s to above "3" since 2007, although the pace of improvement seems to have slowed in the last five years. The quality of articles in the *Journal of Product Innovation Management* has typically tracked higher than the other journals with peaks of about "6" in 2005, 2011, and 2015. Worth noting is that, while Figure 2 shows the average quality score in each year for each journal, there is substantial variation in the overall quality of case study articles published in the same year by the

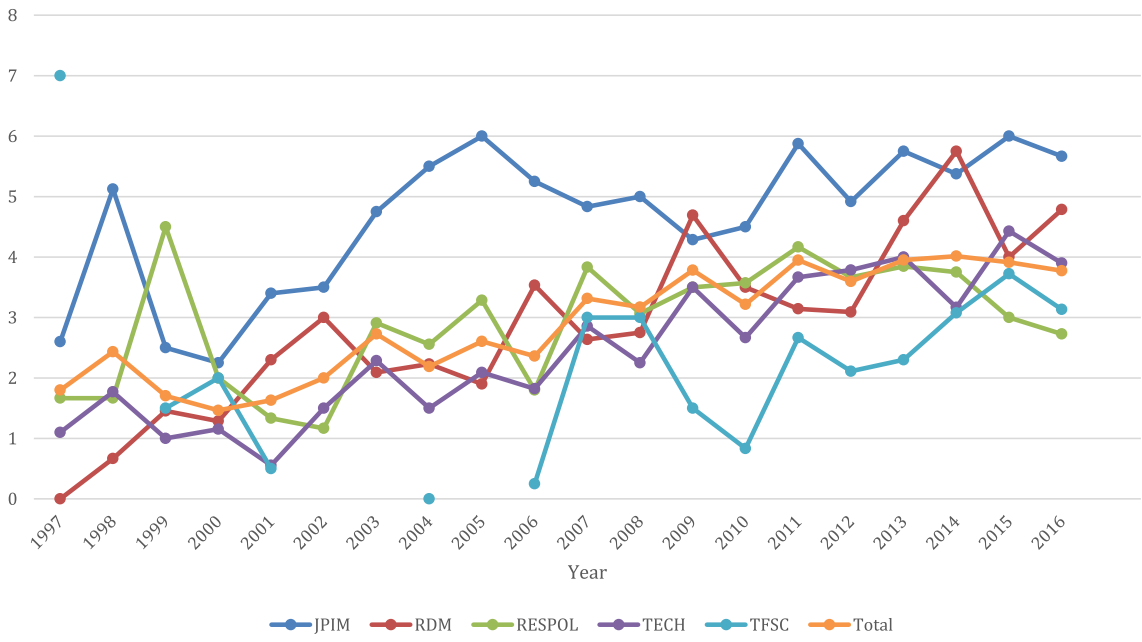

Figure 2. Overall Quality Scores of Case Study Articles in Top Innovation Management Journals (1997–2016)
Note: Single data points not connected by the line are due to the lack of case study articles published in *Technological Forecasting and Social Change* in 1998, 2002, 2003, and 2005. Abbreviations: JPIM is the *Journal of Product Innovation Management*; RDM is *R&D Management;* RESPOL is *Research Policy*; TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*

Table 7. Ordered Logistic Regression Estimates for Overall Quality Score

DV: Overall Quality Score	Model 1		Model 2	
Time	.145***	(.130)		
Year dummies				
1997			–	
1998			.699	(.506)
1999			−.113	(.540)
2000			−.244	(.546)
2001			−.131	(.512)
2002			−.464	(.532)
2003			1.122*	(.511)
2004			.675	(.504)
2005			1.104	(.487)
2006			.902†	(.465)
2007			1.654***	(.504)
2008			1.471**	(.490)
2009			2.024***	(.486)
2010			1.410**	(.486)
2011			2.181***	(.500)
2012			1.887***	(.476)
2013			2.262***	(.492)
2014			2.478***	(.469)
2015			2.542***	(.483)
2016			2.381***	(.485)
Log-likelihood	−1614.223***		−1602.385***	
Pseudo R^2	.066		.073	

N = 818; Robust standard errors are presented in parentheses;
The models control for journal level effects by including journal dummies.
***p < .001; **p < .01; *p < .05; †< .1; Year 1997 serves as reference year.

same journal. The within-year variations are higher than the between-year variation for the *Journal of Product Innovation Management* (1.88 versus 1.19), *R&D Management* (1.89 versus 1.46), *Research Policy* (1.94 versus .99), and *Technovation* (1.62 versus 1.16). For *Technological Forecasting and Social Change,* the within-year and between-year variations are nearly equal (1.69 versus 1.68).

To offer more systematic temporal analysis, an ordered logistic regression, with the article's overall quality score as dependent variable was performed. The results are presented in Table 7. In Model 1, time was operationalized as an ordinal variable going from "0," if the article was published in 1997 to "19," if the article was published in 2016. The coefficient for time suggested the existence of a significant positive time trend for overall case quality. This was also supported by a Jonckheere–Terpstra nonparametric test for ordered alternatives that rejected the null hypothesis of random ordering (z = 10.144, p < .001). Model 2 provided a more fine-grained examination of time dynamics using year dummies. It can be noticed that during the first 10 years of the observation period, only articles published in 2003 and 2006 had quality significantly higher from the articles published in 1997, while those in the remaining years are not significantly different from 1997. It should be noted that, from 2007 onward, the overall quality was significantly higher, peaking in the year 2014. One possible reason for this change was that, in 2007 and shortly after, a number of influential editorials, articles, and special issues on case study methodology were published (e.g., Beverland and Lindgreen, 2010; Eisenhardt and Graebner, 2007; Pratt, 2008, 2009). These articles called for more attention to methodological rigor and provided insights on how researchers can enhance the quality of case study research. Potentially, these articles will have influenced innovation management researchers.

Following the observation that from 2007 onward, quality appeared to be improving faster, the data were considered in two equal time periods: 1997–2006 and 2007–2016. This enabled statistical checks to be made on whether there were significant differences between the two time periods, in terms of overall quality and for each quality criterion. To determine whether there are statistically significant differences in quality between the two periods, a Kruskal–Wallis test was run, for overall quality, and chi-square tests for each of the 10 binary quality criteria. Table 8 shows that articles in the two time periods differed significantly along all quality variables, except for the criterion capturing the usage of a pilot study (for a discussion on this result, see the text below dedicated to this criterion). Similar tests for case demographics were performed. While articles in 2007–2016 used a significantly higher number of data sources than articles in 1997–2006, there is no statistically significant difference between the two periods in terms of the number of cases.

Individual Quality Criteria and Lessons for Researchers

Table 9 shows which of the individual quality criteria in CASET were met by articles in the five journals (i.e., that scored "1" on that criterion). The results indicate the areas where most articles are stronger or weaker and so where there is an opportunity for improving the way innovation management case study

Table 8. Comparison of Quality Between Articles Published in Two Time Periods, 1997–2006 and 2007–2016

	Means		Kruskal–Wallis Test	Chi-Square Test
Dependent variables	1997–2006 (***n*** = 342)	2007–2016 (***n*** = 476)	χ^2	Pearson χ^2
Overall Quality Score	2.148	3.695	101.587***	
Single quality criteria				
Theoretical foundation	.351	.634		64.080***
Pilot study	.061	.078		.805
Theoretical sampling	.310	.569		53.890***
Triangulation	.585	.777		39.913***
Review and validation of evidence	.181	.282		10.972**
Transparency of data collection	.135	.271		22.052***
Inter-coder agreement	.050	.107		8.613**
Case presentation	.178	.405		47.941***
Case interpretation	.187	.292		11.733**
Reflecting on validity and reliability	.111	.279		34.092***
Case demographics				
Number of cases	5.061	4.834	.231	
Number of data sources	1.819	2.263	48.523***	

***$p < .001$; **$p < .01$; *$p < .05$; $^{\dagger} < .1$.

research is conducted and reported. In the following sections, in addition to commenting on the results for each quality criterion, examples of ways in which specific articles met the criteria will be given. Thus, these next sections juxtapose results with lessons for researchers, with the aim of identifying *best practices* "on" each criterion.

On theoretical foundation. Surprisingly, only 52% of all articles adequately reported the reasons for why the case methodology was adopted. It can be seen that 72% of articles in the *Journal of Product Innovation Management* addressed "theoretical foundation" and 42% in *Research Policy.* Given the importance of choosing a research methodology appropriate for the topic(s) being investigated and accounting for the current state of current knowledge on the particular phenomena(on) under investigation, this was perceived as a serious limitation in much case study research. Perhaps the explanation is that researchers are taking it for granted why case study research is used. However, this brings the risk that case study research is not being chosen because it is the most appropriate methodology, rather because it is the most-favored methodology for the researchers, or the "methodology of choice" of a particular institution. It was observed that although all articles included a review of the literature, many did not discuss the methodology adopted by previous studies and the implications of this for their own research.

A very good discussion on the need for case studies (in combination with a survey) can be found in Candi (2010), who based her argument for a multi-method design on the need for methodological fit (Edmondson and McManus, 2007). In addition, the relationship between the case study and survey research was neatly explained using a diagram. A structured approach, using the extant literature to show the need for exploratory case study research, is found in Micheli, Jaina, Goffin, Lemke, and Verganti (2012), who used a table to synthesize their findings from two streams of literature.

On pilot study. Conducting pilot studies in order to sharpen research instruments seems very rare. Only 7% of articles conducted a pilot study, making this quality criterion the lowest scoring one among the 10 criteria evaluated in the CASET. The *Journal of Product Innovation Management* has a higher percentage (17%) than other innovation management journals, whose figures do not exceed 8%. Thus, there is an opportunity for improvement here. However, as case studies can and often do evolve over the time period where empirical data are being collected, it perhaps means that piloting is being used but not reported. This might explain the absence of

Table 9. Proportion of Articles with Score "1" Across Single Quality Criteria

Journal	Theoretical Foundation (%)	Pilot Study (%)	Theoretical Sampling (%)	Triangulation (%)	Review and Validation of Evidence (%)	Transparency of Data Collection (%)	Inter-coder Agreement (%)	Case Presentation (%)	Case Interpretation (%)	Reflecting on Validity and Reliability (%)
JPIM	72	17	66	75	48	43	30	45	43	34
RDM	60	8	41	65	29	25	7	29	27	24
RESPOL	42	7	48	77	20	18	5	35	26	18
TECH	45	4	40	66	15	14	2	24	17	16
TFSC	44	2	45	70	18	12	5	30	18	17
Total	52	7	46	70	24	21	8	31	25	21

Abbreviations: JPIM is the *Journal of Product Innovation Management;* RDM is *R&D Management;* RESPOL is *Research Policy;* TECH is *Technovation;* TFSC is *Technological Forecasting and Social Change.*

any improvement on this quality criterion from the first 10 years in the observation period (1997–2006) to the second 10 years (2007–2016) (see Table 8). Nevertheless, there is an opportunity to more clearly articulate these aspects in manuscripts.

A good example of the effective use of a pilot is found in Jassawalla and Sashittal (1998), who interviewed six managers in four high-technology firms. They used the pilot data together with the literature to develop research questions and an interview protocol, which was used in the subsequent main study. Similarly, Kok and Biemans (2009) used five pilot case studies to pretest their interview format, conceptual framework, and operationalization of their variables, and to refine their methodology for the second stage of the study, where four in-depth case studies were conducted.

On theoretical sampling. Looking at the important issue of sampling, only 46% of all articles discussed how the case(s) was/were selected. In the *Journal of Product Innovation Management*, this discussion was more common (66%). In the other innovation management journals, the figures are 48% or below. This result is disquieting as it suggests that most case study research is based on rather opportunistic sampling. High quality case study research needs to be based on cases chosen for appropriate theoretical reasons. This is particularly important for articles that rely on evidence from a single case study. From a temporal perspective, up until and including 2006, only 31% of articles addressed sampling but from 2007 the figure was 57%. However, there is still room for improvement.

Van Echtelt, Wynstra, Van Weele, and Duysters (2008) clearly articulated the reasons for selecting their eight embedded case studies, defined as the collaboration between the case firm and a supplier. Two theoretically derived criteria were used to select cases: first, the degree of innovation of the collaboration project and; second, the degree of technical complexity. Schmickl and Kieser (2008) used a very rigorous approach to selecting eight embedded case studies (development projects), based on high and low innovativeness. To do this, they constructed a scale of innovativeness (based on four dimensions) and devised a short survey that was completed by company experts. A Mann–Whitney test was used to differentiate between cases and the quantitative data was complemented with indirect qualitative assessments by developers in the projects.

On triangulation. The majority of articles (70%) used multiple sources of data. This is the quality criterion where articles scored the highest, particularly in *Research Policy* (77%). Collecting data from multiple data sources is a prerequisite for triangulation. The majority of case articles use interview data and triangulate these against other sources. This is critical in order to address potential self-reporting and retrospective bias inherent in interview evidence (Gino and Pisano, 2008). Yet, it is troubling to recognize that a large share of the remaining 30% of articles relied on manager-reported verbal data only, without any triangulation being made.

Having collected data from multiple sources, the next step is, of course, using the different data sources in the analysis. The details of how this is being done are normally very hard to discern from reading articles. The overwhelming majority of articles mention nothing about how they have used the multiple sources of data in the analysis. In this respect, Van de Kaa and de Vries (2015, p. 225) are unusual as they explicitly state: "in other words, we collected information from multiple data sources and tried to corroborate that a particular factor led to the success or failure of a particular format." Going even further, Goffin and Koners (2011), were explicit on how triangulation was achieved and used tables to illustrate how the analysis triangulated different data sources.

On the review and validation of evidence. Only 24% of all articles took the time to have their evidence formally reviewed and validated by people other than the researchers. Such reviews can be conducted by interviewees, the company, or fellow researchers. On this criterion, articles in the *Journal of Product Innovation Management* stood out, as 48% scored "1," whereas only 15% of articles in *Technovation* were found to have had their evidence reviewed. The analysis also revealed that having a review and validation of evidence was seldom used before 2007 (18%), and that there is a positive trend during the last five years to include a review and validation of the evidence from the case study (more than 28%). A possible reason for these low scores is the need to involve actors outside the research team to meet this quality criterion. Researchers might be hesitant to ask for additional input from time-constrained interviewees and company representatives. Researchers might also be confident in their evidence and not see the necessity for checks and the further round(s) of analysis that follow data feedback sessions. While review and validation of evidence involves extra effort, it has a strong impact on the validity and reliability of research, hence, there is a clear potential for improvement on this criterion.

Sköld and Karlsson (2007) used a steering group consisting of managers at their case study company to regularly review their results. Schweisfurth and Herstatt (2016) used a number of practices to validate their data, including sending draft text to informants for comments and discussing findings with industry experts.

On transparency of data collection. An often mentioned criterion for high quality case study research is a transparent data collection process, so that others can replicate the study or, at least, have a detailed understanding of the type of data that were collected. Here, an astonishingly low number of articles—21%—included research instruments, such as interview questions and research protocols, in the article or in an appendix (or in a web-based file). On this criterion, 43% of articles in the *Journal of Product Innovation Management* scored "1," whereas only 12% of articles in *Technology Forecasting and Social Change* were transparent on their data collection. Such low scores are surprising, as meeting this criterion would simply require reporting in the article information and instruments that have already been created and used to collect data. Journals' restrictions on the length of articles is not a reason for not reporting data collection adequately, this detail can be provided in appendices or online. It is simple to rectify this limitation. Therefore, the innovation management community should make sure that data collection mechanisms are transparent as this will also enable researchers to build on existing research designs.

Ettlie and Subramaniam (2004) used a combination of open-ended and structured interviews to investigate how companies change the way they originate and develop new products. They used a table to give a very clear overview of the questions asked, how they were coded and selected quotes. Stadler (2011) gave a very clear description of how the different forms of data were collected, including publicly available data, internal data, and interviews. Sample questions for the interviews were also documented.

On inter-coder agreement. In only 8% of articles could evidence be found of data being coded independently by multiple investigators, where the process of arriving at an acceptable inter-coder agreement was unambiguously described. Of course,

inter-coder checks require multiple investigators, but it was found that articles with more than one author often did not employ independent and transparent coding. Independent coding is a time consuming and, some would argue, laborious process. However, it can be rewarding, in that it can produce both more reliable and more interesting results. It can also mitigate the risk that the person who has conducted the fieldwork exaggerates the representativeness of certain observations and, particularly, individual quotes.

O'Connor and DeMartino (2006) independently used two different ways of coding, a computer-aided text analysis software and Microsoft Word. Any discrepancies in coding between the two authors were discussed until agreement was reached. Interestingly, despite being a single-authored article, Perrons (2009) used an independent reviewer who had no prior exposure to the case study or the research topic. This person interpreted the case study data and mapped the evidence onto the framework employed in the research. From a comparison of the interpretations, inter-coder agreement was calculated.

On case presentation. The importance of providing an explicit, comprehensive trail of evidence, from raw data to empirical results, is central to high quality case study research. However, only 31% of articles were found to have provided comprehensive evidence. Evidence can be provided in the form of tables, exhibits, and quotes, with documentation on the coding and pattern-matching processes. Articles in the *Journal of Product Innovation Management* score the highest (45%) on this criterion, followed by *Research Policy* (35%). Of particular note was the large number of articles which were found to rely on selected (supporting) quotes, as their sole presentation of evidence. The selection of quotes is very susceptible to *confirmation bias*—humans are prone to identify evidence that supports their views, rather than what contradicts their views (Gino and Pisano, 2008). Researchers should take steps to mitigate this by presenting not only confirmatory evidence but also checking for the presence (or absence) of *contradictory evidence.*

Haefliger, Jäger, and Von Krogh (2010) employed several practices to demonstrate how data led to conclusions. They started with a detailed description of how the data were collected; then explained their coding process and coding scheme; and showed how the propositions they developed were grounded in the data (summarized this in a table). Stevens (2014) used tables to provide examples of the link between the categories that were developed and a detailed account of respondents' perspectives and compared key variables across the projects studied.

On case interpretation. A good case study article needs to go beyond the presentation of results and evidence. It also needs to interpret the results in the light of existing concepts, models, and findings from the extant literature, in order to generate meanings and insights that can be tested by researchers in future research. Only 25% of articles were coded as providing substantial interpretation and adequate theorizing from the case study findings. The gap in this quality criterion between the highest scoring journal (the *Journal of Product Innovation Management*) and the lowest (*Technovation*) is rather large (43% versus 17%). Researchers seem to "skimp on theory," as Pratt (2009, p. 857) phrased it. A possible reason why many researchers limit their interpretation to simply describing their findings, is that theorizing requires advanced knowledge of the current conversations around theory and an ability to push the frontier further by drawing potentially complex and nontrivial implications. When an article fails to articulate such implications in, for example, research propositions or conceptual framework(s), it also fails to make its utmost contribution. Leading innovation management journals should follow the example of top-tier general management journals like *Academy of Management Journal,* which require authors to clearly indicate the theoretical conversation they want to participate in and what their substantial contribution to this conversation is (Pratt, 2009).

Kester, Griffin, Hultink, and Lauche (2011) provided an excellent example of how case data can be interpreted to generate meaning beyond the case facts. They developed a number of propositions regarding portfolio decision-making processes. Interestingly, the link between the data and developed constructs was clearly articulated in tables that defined the constructs, illustrated them with quotes, and also linked the constructs to the extant literature. Levén, Holmström, and Mathiassen (2014) used the results of an in-depth case study of a large-scale government program to

increase regional competitiveness and they developed a model of managing research and innovation networks. The model was explained using a figure and its key constructs and propositions are clearly articulated and summarized using tables.

On reflecting on validity and reliability. Only 21% of articles were found to include a meaningful reflection on the quality achieved in the research conducted, covering one or more of the following dimensions: construct validity, internal validity, external validity, and reliability. Therefore, 79% of all articles had no discussion of research quality at all—there is significant room for improvement here. The majority of innovation management case study articles failed to employ specific practices to enhance validity and reliability (as suggested by the analysis of the previous nine criteria) and failed to reflect on the quality achieved. This means that members of the innovation management research community do not appear to be as aware of validity and reliability issues as they need to be. If researchers start to reflect on rigor, it is likely to raise the trustworthiness of their findings and improve the quality of their subsequent case study research.

Frishammar, Ericsson, and Patel (2015) discussed the practices they employed to enhance validity and reliability in the section of their article on research design, and then later reflected on the limitations of their study. Hienerth (2006) critically discussed the limitations of case study research in his explanation of research design and then went on to explain the practices adopted to address these limitations. His reflection on rigor continued later in the article with discussions on the remaining limitations of the study.

Relationships among Quality Criteria and Identification of Exemplary Studies

To investigate the existence and strength of relationships among the different quality criteria in CASET, a Spearman's correlation analysis was performed, including case demographics in the analysis. It was expected that if a study was rigorous regarding some of the quality criteria, it would also be rigorous in others. This analysis followed the notion of a "researcher effect," introduced by Aguinis and Solarino (2019), that is, informed researchers would be aware of multiple factors that influence rigor and so a given study would be consistently rigorous (or not, in the case of uninformed researchers) across multiple criteria. The results are reported in Table 10. From the table, it can be seen that nearly all rated quality criteria are positively and significantly correlated. The exceptions are between pilot studies and both inter-coder agreement and case interpretation, which might derive from the heavily skewed distribution of the "pilot study" variable.

While the analysis suggests that articles that scored high on one quality criterion tend to score high on others, the correlation coefficients reported in Table 10 are relatively low, reaching a maximum of .371. The mean value of the correlations among the 10 quality criteria is only .2. Following Sperman's rho interpretation, this means that on average, only 4% (i.e., .2 × .2) of variance in a given criterion is explained by any other criterion. So, the results indicate that meeting one quality criterion does not necessarily mean meeting another one criterion, and this is true even for criteria that refer to the same phase of the research process (e.g., design, data collection, and analysis). This seems to indicate a rather idiosyncratic approach to quality and very little "researcher effect" (i.e., researchers aware of one factor impacting rigor are not necessarily aware of another). This finding is in line with the recent results of Aguinis and Solarino (2019), who investigated the transparency of qualitative research in the strategic management discipline. Looking at the highest correlations between quality criteria in the current study, theoretical foundation and sampling share about 14% of variance. These criteria might be inter-related to larger extent than others. This is because researchers with a good knowledge of the extant research on a topic can clearly articulate why case study methodology is appropriate for investigating certain phenomena, as well as giving theoretical arguments for the choice of suitable cases. Similarly, there is 11% of variance overlap between transparency of data collection and case presentation. Researchers who offer a "window" on their instruments and the specifics of their data collection seem more likely to offer a similar "window" on how the data were analyzed and how conclusions were reached. As regards case demographics, there is no significant correlation between number of cases and number of data sources.

Drawing on the results of the correlation analysis, the data were further perused to identify whether there were articles in the investigated population (of 818 articles) which consistently and systematically

Table 10. Correlation matrix for quality criteria and case demographics

Spearman's rho	Theoretical Foundation	Pilot Study	Theoretical Sampling	Triangulation	Review and Validation of Evidence	Transparency of Data Collection	Inter-coder Agreement	Case Presentation	Case Interpretation	Reflecting on Validity and Reliability	Number of Cases	Number of Data Sources
Theoretical foundation	1.000											
Pilot study	.125*	1.000										
Theoretical sampling	.371*	.098*	1.000									
Triangulation	.186*	.120*	.284*	1.000								
Review and validation of evidence	.183*	.180*	.176*	.208*	1.000							
Transparency of data collection	.249*	.170*	.271*	.124*	.140*	1.000						
Inter-coder agreement	.194*	.055	.219*	.073*	.277*	.167*	1.000					
Case presentation	.259*	.144*	.312*	.150*	.143*	.326*	.200*	1.000				
Case interpretation	.273*	.051	.247*	.114*	.168*	.177*	.155*	.208*	1.000			
Reflecting on validity and reliability	.318*	.127*	.236*	.169*	.219*	.267*	.205*	.305*	.227*	1.000		
Number of cases	.182*	.155*	.230*	.059	.079*	.194*	.116*	.182*	.168*	.114*	1.000	
Number of sources	.198*	.128*	.303*	.839*	.253*	.145*	.108*	.198*	.140*	.198*	.048	1.000

*$p < .05$.

addressed the majority of criteria in CASET (i.e., the issues connected to research design, data collection, and data analysis). An arbitrary overall quality score of "8" or above was set as the threshold and this identified 23 "exemplary studies," corresponding to only 3% of the population (see Table 4). These articles, which are listed in the Appendix Table A1, represent benchmarks in terms of the execution and reporting of the case study methodology and are thus "must-read" studies for innovation management researchers. Of the exemplary studies, 11 were published in the *Journal of Product Innovation Management*, six in *R&D Management*, four in *Research Policy*, and two in *Technovation*. Interestingly, two exemplary articles were based on single case studies (demonstrating that it is indeed possible to conduct excellent case study research based on a single case).

Reflections on the State-of-the-Art of Case Study Research

Key Findings

There is a vast body of methodology literature to support case study researchers and there have been ongoing debates in several management disciplines about the role and quality of case study research. As the innovation management community has not previously reflected on the role and quality of case study research, the current article aimed to assess case study quality, to stimulate debate in the innovation management discipline, and contribute to the improvement of the quality of case study research. To do this, the quality of 818 case study articles published in five top innovation management journals over 20 years (1997–2016) was assessed using the specially developed evaluation template CASET.

The most striking finding of the current study is that the quality of case study research in the innovation management discipline is relatively poor, with the articles scoring an average of only 3.05 against the 10 criteria in the template. However, the results give hope for the future, as the quality of case study research has increased over time. In the past, articles addressed quality issues in a somewhat haphazard way and overlooked many issues, as shown by the low but improving scores. Considering the body of generic case study methodology literature available for the last 30 years, it appears that the lessons from this literature are not being applied (or have not been understood?) by researchers in the innovation management discipline. The questions are: Why is this the case? And: Is this something specific to the innovation management discipline or is it a more general phenomenon? Studies in other disciplines some years ago found similar aggregate results (e.g., only 23% of articles in IS explained their data collection process; in industrial marketing only 23% considered validity) and prompted their communities of researchers to improve their work. Here, the message for innovation management researchers has to be clear—articles should not, on average, address only 3 out of 10 criteria that have a fundamental impact on the quality of the case study research (especially since the factors which impact case study research quality have been fully documented in the methodology literature for many years).

The four most commonly overlooked issues in articles in the innovation management discipline are the use of a pilot study (overlooked by 93%), coding with inter-coder reliability checks (overlooked by 92%), the transparency of data collection (overlooked by 79%), and reflecting on the validity and reliability (overlooked by 79%). It should not be hard for researchers to learn how to address these issues. Similarly, theoretical sampling, learning to present cases more effectively, to interpret them more deeply, and to have evidence validated are priorities. The only criteria addressed by the majority of articles were triangulation (70%) and theoretical foundation (52%) but even those results indicate that they are overlooked by many articles.

The process of conducting case study research is complex, iterative, and involves the challenging task of writing up the analysis and results. CASET should allow researchers to check that they have adequately addressed the main points when carrying out their research. However, similar to Dubé and Puré (2003), it is important to stress that the evaluation template should not be treated as a recipe; it raises generic issues but all of these need to be considered in the specific context of the actual research project. This does not mean that researchers should address quality in an idiosyncratic way, disregarding certain CASET criteria because they are less applicable, require too much work, or are too complex to address in the context of their research. Rather, there is value in regarding the combination of the 10 criteria as the "pillars" on which good quality case study research can be based, in using as many pillars as possible, and

in using different pillars in combination to give maximum support to quality (cf. Aguinis and Solarino, 2019). From this standpoint, the identification of 23 exemplary studies, which scored "8" or "9" against CASET, provides case study researchers with a rich source of learning on how the pillars can be used.

This study set out with the underlying idea that case study research should be highly suited for the innovation management discipline, with its constant flow of new concepts, making exploration and theory building necessary. However, a different conclusion has been reached. It is not the proportion of case study research that is important in itself. Rather, by raising quality, the proportion of published articles based on case study research can be increased. However, for more articles based on case study research to be published in the top innovation management journals, the quality of the case study research is a necessary but not sufficient condition. A significant contribution to theory and practice is of course crucial. As mentioned earlier, CASET can be used to check the quality of case study research but it does not check whether an article makes a real contribution to theory, or to practice.

Implications for Researchers

Researchers who conduct case study research can apply CASET at the design stage, in carefully considering the theoretical purpose of their work and sample; in designing robust and transparent data collection and analysis. Later, they can apply it in reflecting on how they will theorize based on the results and how they will ensure reliability and validity. During the current study process, it became clear that many articles did not consider how previous case studies had been conducted. Researchers need to study not only the content of extant research in the literature but also how it was conducted. If case study research has already been conducted in the area of interest, then researchers can learn from previous case study designs.

Writing up case study research for publication is challenging. Particularly challenging is documenting the results in such a way that reviewers are convinced of the rigor of the analysis process. It has been recognized that if a sufficient *trail* of evidence is not provided in qualitative research, then reviewers will conduct a *trial* of the evidence (Goffin, Raja, Claes, Szwejczewski, and Martinez, 2012). As with any endeavor, gaining experience and learning will help in writing up case study research.

Implications for Reviewers

Reviewers are implicitly aware of the quality requirements of the particular journals they review for. Now, the results discussed in this article make quality expectations for case study research explicit. This should enable reviewers to take actions to improve the consistency of reviewing and to give more structured feedback—suggesting specific steps to improve the quality of a submission (or how to improve future research, if the current submission is judged to be substantially flawed and so to be rejected). CASET is not a panacea but reviewers can use it to quickly and explicitly check key points of the design and execution of case study research, which should help to make the review process more consistent. Also, the template can to a certain extent free up reviewers' time, allowing them to concentrate on assessing the contribution to theory and contribution to practice of a submission.

Implications for Editors

The results also have strong ramifications for the editors of the top innovation management journals. It was found that the quality of case study research is low and it varies widely, even within the same journal and even on the same year, indicating that the peer review process is not as robust and reliable as it could be. These data provide editors with facts on which they can base decisions about how to treat case study submissions. For instance: What level of quality do they require from case study articles to be published in their journal? What are the "must-have" quality attributes for an article? Do submitted articles exceed the required quality threshold and exhibit the "must-haves"?

According to the analysis, it could be argued that the reason why case studies have often been treated as "children of a lesser god" compared to quantitative methods, is not due to a built-in limitation of the methodology per se, but due to its often flawed application by researchers. There is significant potential for innovation researchers to improve the way case study research is conducted and reported. Editors need to decide if and how they want to promote this improvement through, for example, editorials and special issues.

Editors can also take steps to improve the consistency of the review process. The evaluation template developed in this article can play a role in improving the quality and consistency of submissions and reviews. Table 8 shows editors on which criteria the articles published in their journal were particularly weak and enables them to take steps to ensure that future articles address these weaknesses. Editors can also benchmark their journal against other journals in the same discipline and set improvement targets. This study also highlighted exemplary articles, the authors of which could be invited to join the pool of researchers who review case study research. Finally, the results can hopefully stimulate an active debate among innovation management researchers regarding the methods that they use. And this debate should be broader than the discussion of case study research presented in this article.

Reflections on the Future of Case Study Research

Having evaluated the quality of 20 years of case study research and having found an inflection point around 2007, it is interesting to reflect on the future. Can there be a step change going forward? The authors of this current article hope that, in a modest way, they will have contributed to further improvements in the quality of case study research, similar to the influence of Dubé and Paré (2003) on IS research and Beverland and Lindgreen (2010) on marketing research. (And, in 10- or 20-years' time, perhaps other researchers will extend the current study and determine whether quality did improve.)

The quality requirements for journals are likely to climb and it is likely that the evidence presented in case study research will be examined more strictly. In the past, it has often been enough for articles to claim to have addressed rigor—it could be said that researchers' work was "presumed innocent, until proven guilty." In the future, case study researchers' work may be "presumed guilty, until proven innocent"—articles will need to provide explicit evidence of rigor (showing that quality was considered at all stages of the research process).

As a consequence, CASET might need modification to reflect stricter demands. Anchoring statements might need reformulating, to raise the hurdle to achieve a score of "1." For instance, triangulation is currently defined as having collected data from multiple sources. This criterion was operationalized in an inclusive fashion, reflecting the fact that very few articles in extant literature actually showed how data from different data sources were juxtapositioned. In the future, it is reasonable to expect that articles will need to be explicit in explaining how multiple sources of data were used for triangulation purposes, moving beyond stating that data were collected from multiple sources.

As the practice of case study research in innovation management evolves, it might be that the influence of each quality criterion on overall rigor of the research might be perceived differently. While the current version of CASET treats all the criteria similarly, it is reasonable to expect that some criteria will be considered hygiene factors going forward (equivalent to a minimum overall score being required). For instance, it will be necessary to provide an explicit explanation of why the case study methodology was the most appropriate and to have evidence validated by respondents or other external experts. For these criteria, the binary coding used in the current version of CASET will still be appropriate, although the anchoring statements might change.

Some criteria in the evaluation template may in the future be considered "performance factors"; that is, if they are addressed in more sophisticated, effective ways, there will be a bigger positive impact on case study research quality. For example, using multiple coders in analyzing the data and achieving high inter-coder reliability scores will lend more confidence to the findings of the research. Furthermore, presenting findings and empirical evidence in a way that achieves an effective balance between "telling the story" and "presenting the evidence" can differentiate a brilliant case study article from an average one. Researchers can, to different degrees, provide a "window" on their data, which influences the confidence readers can have that the data support the story. Similarly, it might be possible to rate the proficiency with which researchers theorize from case study findings by capturing, defining, and completely explaining a full set of constructs about the phenomena investigated, including the complex relationships between them. The coding of these criteria may need to be enhanced, using ordinal scales to capture different levels.

Meeting the future standards for high quality case study research will be a challenging task. However, with all that has been written on how to conduct high quality case study research, there should be few excuses for *not* doing high quality case study research even today. Innovation management researchers need to be both more rigorous and more disciplined in the way they apply case study research. What currently appears to be an art needs to become more of a science.

References

Aguinis, H., and A. M. Solarino. 2019. Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*. https://doi.org/10.1002/smj.3015.

Antons, D., R. Kleer, and T. O. Salge. 2016. Mapping the topic landscape of JPIM, 1984–2013. In search of hidden structures and development trajectories. *Journal of Product Innovation Management* 33 (6): 726–49.

Bansal, P., and K. Corley. 2011. The coming of age for qualitative research: Embracing the diversity of qualitative methods. *Academy of Management Journal* 54: 233–37.

Barratt, M., T. Y. Choi, and M. Li. 2011. Qualitative case studies in operations management: Trends, research outcomes, and future research implications. *Journal of Operations Management* 29 (4): 329–42.

Benbasat, I., D. K. Goldstein, and M. Mead. 1987. The case research strategy in studies of information systems. *MIS Quarterly* 11 (3): 368–86.

Bettencourt, R. 2010. Strategic prospective for the implementation of employment policies in the Azores. *Technological Forecasting and Social Change* 77: 1566–74.

Beverland, M., and A. Lindgreen. 2010. What makes a good case study? A positivist review of qualitative case research published in Industrial Marketing Management, 1971–2006. *Industrial Marketing Management* 39: 56–63.

Bluhm, D. J., W. Harman, T. W. Lee, and T. R. Mitchell. 2011. Qualitative research in management: A decade of progress. *Journal of Management Studies* 48 (8): 1866–91.

Boyer, K. K., and M. L. Swink. 2008. Empirical elephants – Why multiple methods are essential to quality research in operation and supply chain management. *Journal of Operations Management* 26 (3): 337–48.

Candi, M. 2010. Benefits of aesthetic design as an element of new service development. *Journal of Product Innovation Management* 27 (7): 1047–64.

Coughlan, P., and D. Coghlan. 2002. Action research for operations management. *International Journal of Operations and Production Management* 22 (2): 220–40.

Cristiano, J. J., J. K. Liker, and C. C. White, III. 2000. Customer-driven product development through quality function deployment in the U.S. and Japan. *Journal of Product Innovation Management* 17: 286–308.

Drejer, A., K. Blackmon, and C. Voss. 2000. Worlds apart – A look at the operations management area in the US, UK and Scandinavia. *Scandinavian Journal of Management* 16: 46–66.

Drejer, A., and A. Gudmundsson. 2002. Towards multiple product development. *Technovation* 22: 733–45.

Dubé, L., and G. Paré. 2003. Rigor in information systems positivist case research: Current practices, trends, and recommendations. *MIS Quarterly: Management Information Systems* 27 (4): 597–635.

Edmondson, A. C., and S. E. McManus. 2007. Methodological fit in management field research. *Academy of Management Review* 32 (4): 1246–64.

Eisenhardt, K. M. 1989. Building theories from case study research. *Academy of Management Review* 14 (4): 532–50.

Eisenhardt, K. M., and M. E. Graebner. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal* 50 (1): 25–32.

Ellram, L. M. 1996. The use of the case study method in logistics research. *Journal of Business Logistics* 15 (2): 93–138.

Ettlie, J. E., and M. Subramaniam. 2004. Changing strategies and tactics for new product development. *Journal of Product Innovation Management* 21 (2): 95–109.

Flynn, B. B. 2008. Having it all: Rigor versus relevance in supply chain management research. *Journal of Supply Chain Management* 44 (2): 63–67.

Frishammar, J., K. Ericsson, and P. C. Patel. 2015. The dark side of knowledge transfer: Exploring knowledge leakage in joint R&D projects. *Technovation* 41: 75.

Gemser, G., and M. Leenders. 2001. How integrating industrial design in the product development process impacts company performance. *Journal of Product Innovation Management* 18: 28–38.

Gephart, R. P. Jr. 2004. Qualitative research and the *Academy of Management Journal*. *Academy of Management Journal* 47 (4): 454–62.

Gibbert, M., and W. Ruigrok. 2010. The "what" and "how" of case study rigor: Three strategies based on published work. *Organizational Research Methods* 13 (4): 710–37.

Gibbert, M., W. Ruigrok, and B. Wicki. 2008. What passes as a rigorous case study? *Strategic Management Journal* 29: 1465–74.

Gino, F. and G. Pisano. 2008. Toward a theory of behavioural operations. *Manufacturing & Service Operations Management* 10 (4): 676–91.

Gioia, D. A., K. G. Corley, and A. L. Hamilton. 2012. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods* 16 (1): 15–31.

Glaser, B. G., and A. L. Strauss. 1967. *The Discovery of grounded theory: Strategies for qualitative research*. Hawthorn, NY: Aldine de Gruyter.

Goffin, K., and U. Koners. 2011. Tacit knowledge, lessons learnt, and new product development. *Journal of Product Innovation Management* 28 (2): 300–18.

Goffin, K., J. Raja, B. Claes, M. Szwejczewski, and V. Martinez. 2012. Rigor in qualitative supply chain research—Lessons from applying repertory grid technique. *International Journal of Physical Distribution and Logistics Management* 42 (8/9): 804–27.

Haefliger, S., P. Jäger, and G. Von Krogh. 2010. Under the radar: Industry entry by user entrepreneurs. *Research Policy* 39 (9): 1198–213.

Hienerth, C. 2006. The commercialization of user innovations: The development of the rodeo kayak industry. *R&D Management* 36 (3): 273–94.

Jassawalla, A. R., and H. C. Sashittal. 1998. An examination of collaboration in high-technology new product development processes. *Journal of Product Innovation Management* 15 (3): 237–54.

Kester, L., A. Griffin, E. J. Hultink, and K. Lauche. 2011. Exploring portfolio decision-making processes. *Journal of Product Innovation Management* 28 (5): 641–61.

Ketchen, D. J., C. W. Craighead, and C. Li. 2018. Achieving research design excellence through the pursuit of perfection:

Toward strong theoretical calibration. *Journal of Supply Chain Management* 54 (1): 16–22.

Ketokivi, M., and T. Choi. 2014. Renaissance of case research as a scientific method. *Journal of Operations Management* 32: 232–40.

Kok, R. A. W., and W. G. Biemans. 2009. Creating a market-oriented product innovation process: A contingency approach. *Technovation* 29 (8): 517.

Leonard-Barton, D. 1990. A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science* 1 (3): 248–66.

Levén, P., J. Holmström, and L. Mathiassen. 2014. Managing research and innovation networks: Evidence from a government sponsored cross-industry program. *Research Policy* 43 (1): 156–68.

Linton, J. D., and N. Thongpapanl. 2004. Perspective: Ranking the technology innovation management journals. *Journal of Product Innovation Management* 21 (2): 123–39.

Meredith, J. 1998. Building operations management theory through case and field research. *Journal of Operations Management* 16 (4): 441–54.

Micheli, P., J. Jaina, K. Goffin, F. Lemke, and R. Verganti. 2012. Perceptions of industrial design: The "means" and the "ends". *Journal of Product Innovation Management* 29 (5): 687–704.

Miles, M. B., and M. Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. London, UK: Sage.

Miles, M. B., A. M. Huberman, and J. Saldana. 2014. *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Näslund, D. 2002. Logistics needs qualitative research – Especially action research. *International Journal of Physical Distribution & Logistics Management* 32 (5): 321–38.

O'Connor, G. C., and R. DeMartino. 2006. Organizing for radical innovation: An exploratory study of the structural aspects of RI management systems in large established firms. *Journal of Product Innovation Management* 23 (6): 475–97.

Oloruntoba, R. 2013. Plans never go according to plan: An empirical analysis of challenges to plans during the 2009 Victoria bushfires. *Technological Forecasting and Social Change* 80: 1674–702.

Perks, H., and D. Roberts. 2013. A review of longitudinal research in the product innovation field, with discussion of utility and conduct of sequence analysis. *Journal of Product Innovation Management* 30 (6): 1099–111.

Perrons, R. K. 2009. The open kimono: How Intel balances trust and power to maintain platform leadership. *Research Policy* 38 (8): 1300–12.

Piekkari, R., E. Plakoyiannaki, and C. Welch. 2010. "Good" case research in industrial marketing: Insights from research practice. *Industrial Marketing Management* 39 (1): 109–17.

Pratt, M. G. 2008. Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods* 11 (3): 481–509.

Pratt, M. G. 2009. From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal* 52 (5): 856–62.

Riek, R. 2001. From experience: Capturing hard-won NPD lessons in checklists. *Journal of Product Innovation Management* 18: 301–13.

Schmickl, C., and A. Kieser. 2008. How much do specialists have to learn from each other when they jointly develop radical product innovations? *Research Policy* 37 (3): 473.

Schweisfurth, T. G., and C. Herstatt. 2016. How internal users contribute to corporate product innovation: The case of embedded users. *R&D Management* 46 (S1): 107–26.

Sköld, M., and C. Karlsson. 2007. Multibranded platform development: A corporate strategy with multimanagerial challenges. *Journal of Product Innovation Management* 24 (6): 554–66.

Stadler, C. 2011. Process innovation and integration in process-oriented settings: The case of the oil industry. *Journal of Product Innovation Management* 28 (S1): 44–62.

Stevens, E. 2014. Fuzzy front-end learning strategies: Exploration of a high-tech company. *Technovation* 34 (8): 431–40.

Thieme, J. 2007. Perspective: The world's top innovation management scholars and their social capital. *Journal of Product Innovation Management* 24 (3): 214–29.

Van de Kaa, G., and H. J. de Vries. 2015. Factors for winning format battles: A comparative case study. *Technological Forecasting and Social Change* 91: 222–35.

Van Echtelt, F. E., F. Wynstra, A. J. Van Weele, and G. Duysters. 2008. Managing supplier involvement in new product development: A multiple-case study. *Journal of Product Innovation Management* 25 (2): 180–201.

Voss, C., N. Tsikriktsis, and M. Frohlich. 2002. Case research in operations management. *International Journal of Operations and Production Management* 22 (2): 195–219.

Weick, K. 1995. What theory is not, theorizing is. *Administrative Science Quarterly* 40 (3): 385–90.

Welch, C., E. Plakoyiannaki, R. Piekkari, and E. Paavilainen-Mäntymäki. 2012. Legitimizing diverse uses for qualitative research: a rhetorical analysis of two management journals. *International Journal of Management Reviews* 15 (2): 245–64.

Yin, R. K. 2009. *Case study research, design and methods* (5th ed.). London, UK: Sage Publications.

APPENDIX

Table A1 Exemplary Studies—Articles with an Overall Quality Score of "8" or More

1. Antioco, M., R. K. Moenaert, and A. Lindgreen. 2008. Reducing ongoing product design decision-making bias. *Journal of Product Innovation Management* 25 (6): 528-45.
2. Berends, H., M. Jelinek, I., Reymen, and R. Stultiëns. 2014. Product innovation processes in small firms: Combining entrepreneurial effectuation and managerial causation. *Journal of Product Innovation Management* 31 (3): 616-35.
3. Beverland, M. B., P. Micheli, and F.J. Farrelly. 2016. Resourceful sensemaking: Overcoming barriers between marketing and design in NPD. *Journal of Product Innovation Management* 33 (5): 628-48.
4. Brattström, A., and A. Richtnér. 2014. Good cop-bad cop: Trust, control, and the lure of integration. *Journal of Product Innovation Management* 31 (3): 584-98.
5. Cassiman, B., M. Colombo, P. Garrone, and R. Veugelers. 2005. The impact of M&A on the R&D process: An empirical analysis of the role of technological-and market-relatedness. *Research Policy* 34 (2): 195-220.
6. Chen, R. R., Ä. Kannan, and R. P. Narasimhan. 2015. Formal integration archetypes in ambidextrous organizations. *R&D Management* 45 (3), 267-86.
7. Eisingerich, A. B., S. J. Bell, and P. Tracey. 2010. How can clusters sustain performance? The role of network strength, network openness, and environmental uncertainty. *Research Policy* 39 (2): 239-53.
8. Frishammar, J., K. Ericsson, and P. C. Patel. 2015. The dark side of knowledge transfer: Exploring knowledge leakage in joint R&D projects. *Technovation* 41: 75-88.
9. Godener, A., and K. E. Söderquist. 2004. Use and impact of performance measurement results in R&D and NPD: an exploratory study. *R&D Management* 34 (2): 191-219.
10. Haefliger, S., P. Jäger, and G. Von Krogh. 2010. Under the radar: Industry entry by user entrepreneurs. *Research Policy* 39 (9): 1198-213
11. Harmancioglu, N., R. C. McNally, R. J. Calantone, and S. S. Durmusoglu. 2007. Your new product development (NPD) is only as good as your process: an exploratory analysis of new NPD process design and implementation. *R&D Management* 37 (5): 399-424.
12. Hienerth, C. 2006. The commercialization of user innovations: The development of the rodeo kayak industry. *R&D Management* 36 (3): 273-94.
13. Hienerth, C., and C. Lettl. 2011. Exploring how peer communities enable lead user innovations to become standard equipment in the industry: Community pull effects. *Journal of Product Innovation Management* 28 (s1): 175-95.
14. Horváth, A., and E. Enkel. 2014. When general recommendations fail: How to search in single innovation project settings. *R&D Management* 44 (4): 409-26.
15. Kammerlander, N., and M. Ganter. 2015. An attention-based view of family firm adaptation to discontinuous technological change: Exploring the role of family CEOs' noneconomic goals. *Journal of Product Innovation Management* 32 (3): 361-83.
16. Kelley, D. 2009. Adaptation and organizational connectedness in corporate radical innovation programs. *Journal of Product Innovation Management* 26 (5): 487-501.
17. Kester, L., A. Griffin, E. J. Hultink, and K. Lauche, K. 2011. Exploring portfolio decision-making processes. *Journal of Product Innovation Management* 28 (5): 641-61.
18. Kok, R. A., and W. G. Biemans. 2009. Creating a market-oriented product innovation process: A contingency approach. *Technovation* 29 (8): 517-26.
19. Koners, U., and K. Goffin. 2007. Learning from post-project reviews: A cross-case analysis. *Journal of Product Innovation Management* 24 (3): 242-58.
20. Levén, P., J. Holmström, and L. Mathiassen. 2014. Managing research and innovation networks: Evidence from a government sponsored cross-industry program. *Research Policy* 43 (1): 156-68.
21. Micheli, P., J. Jaina, K. Goffin, F. Lemke, and R. Verganti. 2012. Perceptions of industrial design: The "means" and the "ends". *Journal of Product Innovation Management* 29 (5): 687-704.
22. Perks, H., R. Cooper, and C. Jones. 2005. Characterizing the role of design in new product development: An empirically derived taxonomy. *Journal of Product Innovation Management* 22 (2): 111-27.
23. Sala, A., P. Landoni, and R. Verganti. 2016. Small and medium enterprises collaborations with knowledge intensive services: An explorative analysis of the impact of innovation vouchers. *R&D Management* 46 (S1): 291-302.