

# **Assessing Systematic Literature Review Bias: Kaizen Events in Hospitals Case Study**

**Fernando Gonzalez Aleu, Francisco Flores, Jimena Perez, and Roy Gonzalez**

School of Engineering and Technology  
Universidad de Monterrey  
San Pedro Garza Garcia, NL 45000, Mexico  
[fernando.gonzalezaleu@udem.edu](mailto:fernando.gonzalezaleu@udem.edu)

**Jose Arturo Garza-Reyes**

Centre for Supply Chain Improvement  
University of Derby, Derby, DE22 1GB, UK  
[J.Reyes@derby.ac.uk](mailto:J.Reyes@derby.ac.uk)

## **Abstract**

A systematic literature review (SLR) is a protocol used to identify publications, select relevant publications, collect data, conduct scientometric analyses, and report research results (SLR outcomes or findings). Despite the increasing use of SLR to assess the maturity or evolution of a research field, as Engineering Management, there are a limit number of publications focused to test SLR biases. Therefore, the purposes of this investigation are to test search field bias (precise SLR vs. sensitive SLR) and to identify statistically significant differences between SLR outcomes. In order to achieve these goals, a three steps methodology was used in three platforms/databases. First, a precise SLR in ProQuest (search terms only in abstract) was conducted to identify publications describing a single Kaizen event in a hospital. From these publications, five metrics were assessed: new authors per year, number of authors per paper, number of publications per year, Kaizen event duration (days), and number of tools used during the Kaizen event per paper. Second, a sensitive SLR in ProQuest (search term in full text) was conducted using the same search terms, exclusion criteria, and metrics from the first step. Third, t-test hypotheses were conducted in SPSS version 20 to identify statistically significant difference for each metric between precise SLR vs. sensitive SLR. The same three steps were used in two more platforms/databases: EBSCOhost and Scopus. Initial results from this ongoing investigation show statistically significant differences between precise SLR and sensitive SLR for some of the five metrics assessed, such as the number of publications per year. Final results will be available in November 2018.

## **Keywords**

Systematic literature review, kaizen event, rapid improvement event, hospital, bias

## **1. Introduction**

Year to year, the number of published scientific papers is increasing due to the accessibility of information obtained by technology. This vast range of information has become one of the main challenges for researchers; needing to read or scan several articles found for later selecting the pertinent ones for the study (filter information). Because of this, researchers spend a lot of their time using search engines to locate the material on the vast and unorganized web [1]. Currently, more and more researchers have adopted a systematic way of search and analyze information obtained from the literature available conducting systematic literature reviews. A systematic literature review (SLR) is a protocol used to identify publications, select relevant publications, collect data, conduct scientometric analyses, and report research results (SLR outcomes or findings) [2]. Despite the increasing use of SLR to assess the maturity or evolution of a research field (see Figure 1), as Engineering Management; there are a limit number of publications focused to test SLR biases (over-estimation or an under-estimation of SLR outcomes).

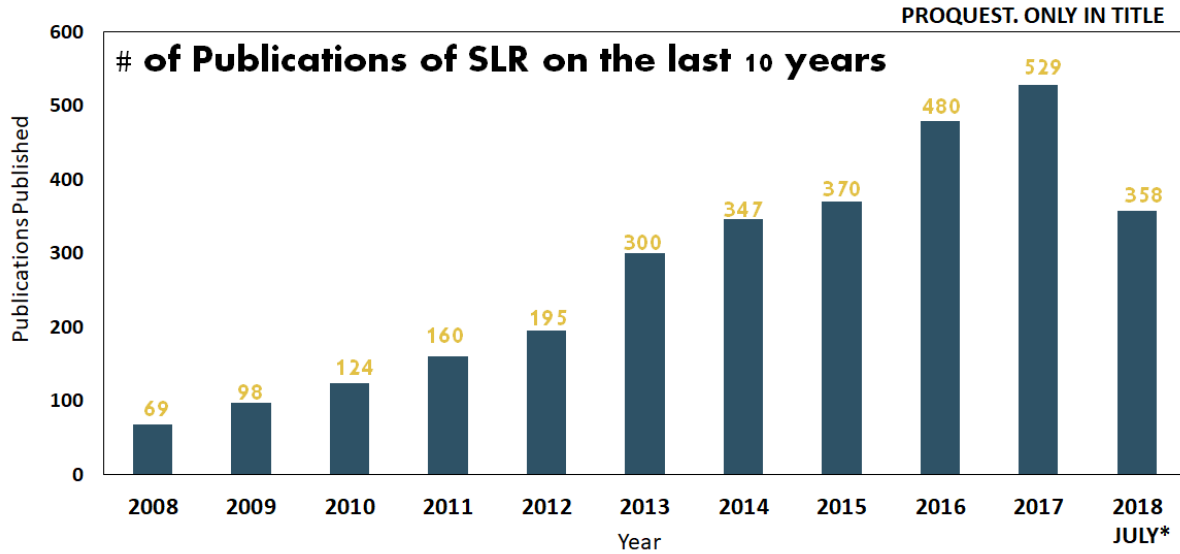


Figure 1: Number of SLR Publications in the Last 10 Years

Several different types of biases could be found during a traditional literature review or SLR [3,4], such as search field bias, publication bias, dissemination bias, and database bias. The purposes of this investigation are to test search field bias (precise SLR vs. sensitive SLR) and to identify statistically significant differences between SLR outcomes or metrics. The goal of a precise SLR or search is to identify those publications highly related to the research field to be assessed. On the other hand, during a sensitive search the researcher is interested to collect any publication that could be related to his/her research field; spending a considerable amount of time during the screening process [5]. Therefore, during a precise search, there is a possibility that a researcher lost valuable publications, affecting SLR outcomes.

Keathley-Herring et al., [6] offer an extensive list of metrics used to assess the maturity of a research field. In order to achieve these goals the research team selected four metrics (average of new authors per year, average of publications per year, average of Kaizen event duration, and average of authors per paper), the research team select Kaizen event or rapid improvement event in a hospital as the research field, and each of the three full-time researchers conducted two SLRs (precise and sensitive) using three platforms/databases: EBSCO, Scopus, and ProQuest. A total of XX hypotheses were identified and tested in this investigation (see Table 1).

Table 1: Hypotheses tested

Database	Hypotheses		
	No.	Description	Equations*
EBSCO (Researcher A)	H01	There is no difference in the <u>average of new author per year</u> between an abstract and a full-text search	$H_0: \mu_{NAE} = \mu_{NFE}$ $H_A: \mu_{NAE} \neq \mu_{NFE}$
	H02	There is no difference in the <u>average of publication per year</u> between an abstract and a full-text search	$H_0: \mu_{PAE} = \mu_{PFE}$ $H_A: \mu_{PAE} \neq \mu_{PFE}$
	H03	There is no difference in the <u>average Kaizen event duration</u> between an abstract and a full-text search	$H_0: \mu_{KAE} = \mu_{KFE}$ $H_A: \mu_{KAE} \neq \mu_{KFE}$
	H04	There is no difference in the <u>average of authors per Paper</u> between an abstract and a full-text	$H_0: \mu_{QAE} = \mu_{QFE}$ $H_A: \mu_{QAE} \neq \mu_{QFE}$
Scopus (Researcher B)	H05	There is no difference in the <u>average of new author per year</u> between an abstract and a full-text search	$H_0: \mu_{NAS} = \mu_{NFS}$ $H_A: \mu_{NAS} \neq \mu_{NFS}$
	H06	There is no difference in the <u>average of publication per year</u> between an abstract and a full-text search	$H_0: \mu_{PAS} = \mu_{PFS}$ $H_A: \mu_{PAS} \neq \mu_{PFS}$
	H07	There is no difference in the <u>average Kaizen event duration</u> between an abstract and a full-text search	$H_0: \mu_{KAS} = \mu_{KFS}$ $H_A: \mu_{KAS} \neq \mu_{KFS}$
	H08	There is no difference in the <u>average of authors per Paper</u>	$H_0: \mu_{QAS} = \mu_{QFS}$

		between an abstract and a full-text	$H_A: \mu_{QAS} \neq \mu_{QFS}$
ProQuest (Researcher C)	H09	There is no difference in the <u>average of new author per year</u> between an abstract and a full-text search	$H_O: \mu_{NAP} = \mu_{NFP}$ $H_A: \mu_{NAP} \neq \mu_{NFP}$
	H10	There is no difference in the <u>average of publication per year</u> between an abstract and a full-text search	$H_O: \mu_{PAP} = \mu_{PFP}$ $H_A: \mu_{PAP} \neq \mu_{PFP}$
	H11	There is no difference in the <u>average Kaizen event duration</u> between an abstract and a full-text search	$H_O: \mu_{KAP} = \mu_{KFP}$ $H_A: \mu_{KAP} \neq \mu_{KFP}$
	H12	There is no difference in the <u>average of authors per Paper</u> between an abstract and a full-text	$H_O: \mu_{QAP} = \mu_{QFP}$ $H_A: \mu_{QAP} \neq \mu_{QFP}$

\*Equation Notations: average of new author per year (N), average of publications per year (P), average of Kaizen even duration (K), average of author per paper (Q), abstract (A), full-text (T), EBSCO (E), Scopus (S), and ProQuest (P).

The remaining sections of this paper include research method, results, and discussion. First, the SLR methodology framework developed by Keathley-Herring et al., [6] was used by the research team in this investigation. Second, results from each of the four metrics were calculated in the three databases and the 12 hypotheses were tested. Third, discussions of the results obtained in this investigation were commented, including research limitations and future works.

## 2. Research Method

Keathley-Herring et al., (2016) developed a six-step SLR framework (see Figure 2): problem definition, scoping study, search strategy, exclusion criteria, data collection, and analysis. First, problem definition was decided in the introduction during the introduction: the number of papers using SLR is increasing, but there is a lack of publications studying SLR bias. Second, considering the different SLR bias identified in other literature [3,4], research leader field of interest (Kaizen event in hospitals), and metrics to assess the maturity of the field (average of new authors per year, average of publications per year, average of Kaizen event duration, and average of authors per publication). Therefore the scope of this investigation is to assess search field bias for publications describing a single Kaizen event in a hospital using four metrics and three platforms/databases.

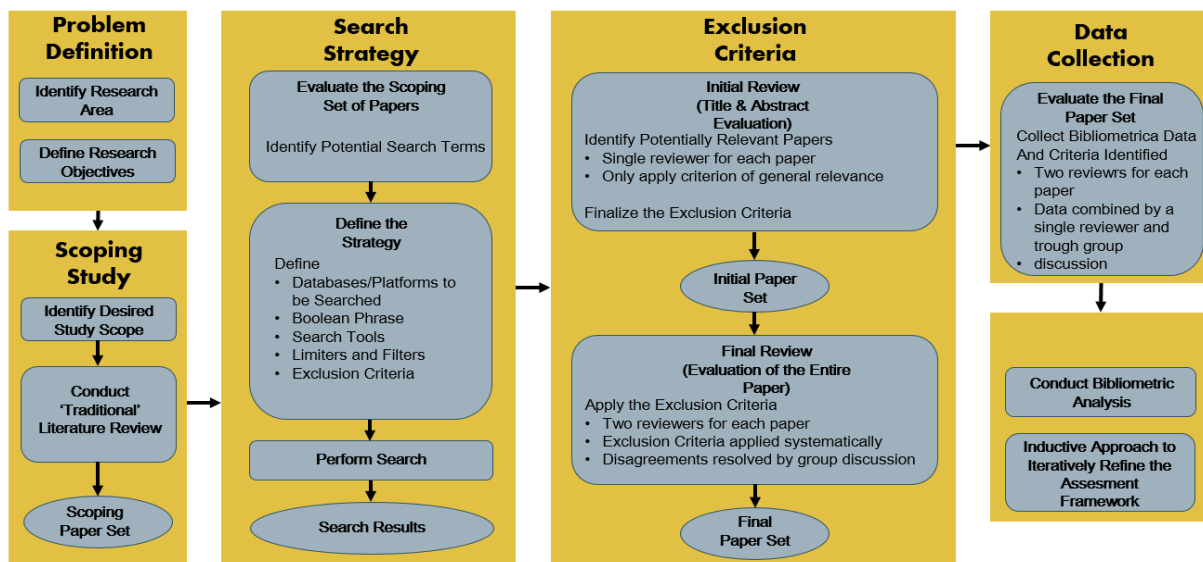


Figure 2: SLR framework

Third, the search strategy consisted of search terms, Boolean operators, and search field (see Table 2). The same search strategy was used in the three databases: EBSCO (researcher A), Scopus (researcher B), and ProQuest (researcher C). Fourth, papers that do not match the following exclusion criteria were removed from this investigation: full publication not available, duplicate publications, publications in a different language than English, publications different than journal papers and conference proceedings, and publications that do not describe the application of a single Kaizen event or rapid improvement event in hospital. Each of the three full-time researchers

in this investigation applied these exclusion criteria. Fifth, also each researcher was responsible to collect the following data each publication: number of new authors per year, number of publications per year, Kaizen event duration (in days), and number of authors per publication. With this data, the four metrics mentioned previously in step two were calculated. Lastly, each researcher used a t-test to assess four hypotheses per platform/database (see Table 1).

Table 2: Search Strategy

No.	Search Strategy
1	Abstract (“Kaizen event” OR “rapid improvement event” OR “lean event” OR “Kaizen blitz”) AND Abstract (hospital OR “healthcare” OR “health care”)
2	Full-text (“Kaizen event” OR “rapid improvement event” OR “lean event” OR “Kaizen blitz”) AND Full-text (hospital OR “healthcare” OR “health care”)

### 3. Results

Table 3 shows the initial number of publications identified for each search strategy and platform/database, as well as the final set of publications included in this investigation. The minimum rate of publications admitted in this investigation was 1.7% in full-text ProQuest; similar to other SLRs [7].

Table 3: Initial set of papers identify per search strategy in each platform/database

Papers	EBSCO (Research A)		Scopus (Research B)		ProQuest (Research C)	
	Abstract	Full-text	Abstract	Full-text	Abstract	Full-text
Initial	69	1253	62	555	55	1190
Exclusion criteria	58	1231	52	526	48	1170
Final Set	11	22	10	29	7	20

After the three researchers collected the data mentioned in the previous section, each metric was calculated and the 12 hypotheses were tested (see Table 4). There are three main findings to highlight from these results: four out of twelve hypotheses were rejected (33%), average of new authors per year in ProQuest shows a high difference between abstract search strategy and full-text strategy is considerable high, and SD levels are high in most of the metrics (especially in average of new authors per year, average of publications per year, and Kaizen event duration). Therefore, it is important to discuss these findings in the following section.

Table 4: Results per metric, search strategy and platform/database

Metrics	Statistics/Hypothesis	EBSCO (Research A)		Scopus (Research B)		ProQuest (Research C)	
		Abstract	Full-text	Abstract	Full-text	Abstract	Full-text
Average of new authors per year	Mean	5.36	8.45	3.18	9.18	2.00	5.00
	SD	4.95	5.85	3.76	4.42	2.93	4.49
	p-value	0.196		0.003		0.079	
	Hypothesis result	H01: Accepted		<b>H05: Rejected</b>		H09: Accepted	
Average of publications per year	Mean	1.00	2.00	0.82	2.64	0.64	1.73
	SD	0.89	1.18	0.98	1.36	0.81	1.35
	p-value	0.037		0.002		0.032	
	Hypothesis results	<b>H02: Rejected</b>		<b>H06: Rejected</b>		<b>H10: Rejected</b>	
Kaizen event duration (days)	Mean	5.27	4.98	4.38	4.85	6.14	6.44
	SD	4.98	3.87	3.28	3.26	3.61	6.37
	p-value	0.853		0.727		0.908	
	Hypothesis results	H03: Accepted		H07: Accepted		H11: Accepted	
Average of authors per publication	Mean	5.36	4.32	4.00	3.69	4.00	3.50
	SD	2.54	2.23	1.32	1.93	1.15	1.93
	p-value	0.235		0.656		0.528	
	Hypothesis results	H04: Accepted		H08: Accepted		H12: Accepted	

#### **4. Discussion**

The purposes of this investigation were to test search field bias (precise SLR vs. sensitive SLR) and to identify statistically significant differences between SLR outcomes or metrics. After applied the SLR framework (see Figure 2) and test 12 hypotheses (see Table 4), the research team found that four hypotheses were rejected. This evidence suggests, in some cases, that SLR outcomes or metrics could be statistically different when you conduct abstract search terms vs. full-text search terms. It is interesting to observe that the hypotheses related to the average of publications per year (a metric used frequently by a researcher in SLR papers) were rejected in the three platforms/databases. Additional hypotheses could be rejected also, but the high level of SD values influence these results; for example, the average of new authors per year in ProQuest and Kaizen event duration in the three platforms/databases. Although the research leader selected these four metrics considering that they are frequently used by other researchers with published SLRs in other fields, these metrics involve three out of eight dimensions used to assess the maturity or evolution of a research field [6]: authors' characteristics (average of new authors per year), average of publications per year (publication characteristics), Kaizen event duration (content characteristics), and average of authors per paper (publication characteristics). Therefore, another way to interpret finding from this investigation is that two out of eight dimensions (67%) show a statistically significant difference between abstract search field and full-text search field in at least one platform/database. These findings should be considered by researchers interesting to conduct SLRs and publish their results in order to define the maturity or evolution of a research field using quantitative metrics.

However, these findings also should be understood considering mainly three limitations. First, this investigation was focused to assess one of several biases identified during an SLR [3,4]. Second, this investigation was limited to four metrics from the extensive identify by Keathley-Herring et al., [6] but represent three out of eight possible dimensions to be assessed. Third, the sample size (number of publications) used to calculate the four metrics in both search strategies (abstract and full-text) in the three platforms/databases are relatively small. This investigation was designed to be conducted during four months using three full-time undergraduate students (each student worked in this project an average of 25 hours per week). Therefore, the research leader selected a research field that matches with his investigation topics (Kaizen event in hospitals) and with a reasonable scope or number of publications to be identified and screened. Future work should be focused on the following topics: assess other SLR biases, assess search strategy bias using additional metrics, and assess search strategy using a different research field.

#### **References**

1. Lawrence, S., and Giles, C. L., 1999, "Accessibility of information on the web," *Nature*, 400 (6740), 107.
2. Tranfield, D., David D., and Palminder, S., 2003, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review." *British journal of management* 14(3), 207-222.
3. Petticrew, M., and Roberts, H., 2008, "Systematic reviews—do they 'work' in informing decision-making around health inequalities?," *Health Economics, Policy, and Law*, 3(2), 197-211
4. Booth, A., Sutton, A., and Papaioannou, D., and 2012, *Systematic Approaches to a Successful Literature Review*. Sage Publications Inc.
5. Lefebvre, C., Manheimer, E., and Glanville, J., 2011, "Chapter 6: Searching for studies", appears in *Cochrane Handbook for systematic reviews of interventions*.
6. Keathley-Herring, H., Van Aken, E. M., Gonzalez-Aleu, F., Deschamps, F., Letens, G., and Cardenas-Orlandini, P., 2016, "Assessing the maturity of a research area: bibliometric review and proposed framework," *Scientometrics*, 109(2), 927-951.
7. Gonzalez-Aleu, F. and Van Aken, E. M., 2017, "Continuous improvement projects: an authorship bibliometric analysis," *International Journal of Health Care Quality Assurance*, 30(5), 467-476

#### **Biographies**

**Fernando Gonzalez Aleu**, is an Associate Professor at the Universidad de Monterrey in Mexico. He received a BS in Mechanical and Management Engineering at UDEM (Mexico) in 1993, a MSc with specialty on Manufacturing Systems at ITEMS (Mexico) in 1999 and a MSc in Industrial and Systems Engineering at Virginia Tech in 2015, a Ph.D. in Industrial and Systems Engineering at Virginia Tech in

2016. His research focuses on the application of continuous improvement programs and projects, (e.g., Kaizen Events, Six Sigma, and Lean Six Sigma) and performance excellence models. He has more than 15 year of professional experience implementing quality, environmental, and management systems in Mexico, Chile, and Peru. He is member of the Institute of Industrial Engineers (IISE), the American Society for Engineering Management (ASEM), the American Society for Quality (ASQ), and the Industrial Engineering and Operations Management Society (IEOM).

**Francisco Flores** is an Industrial and Systems Engineering undergraduate senior student at Universidad de Monterrey in México.

**Jimena Perez** is an Industrial and Systems Engineering undergraduate senior student at Universidad de Monterrey in México.

**Roy Gonzalez** is an Industrial and Systems Engineering undergraduate senior student at Universidad de Monterrey in México.

**Jose Arturo Garza-Reyes** is a Professor of Operations Management and Head of the Centre for Supply Chain Improvement at the College of Business, Law and Social Sciences, University of Derby, UK. He is actively involved in industrial projects where he combines his knowledge, expertise and industrial experience in operations management to help organisations achieve excellence in their internal functions and supply chains. He has also led and managed international research projects funded by the British Academy, British Council and Mexico's National Council of Science and Technology (CONACYT). As a leading academic, he has published over 150 articles in leading scientific journals, international conferences and five books in the areas of operations management and innovation, manufacturing performance measurement and quality management systems. Areas of expertise and interest for Professor Garza-Reyes include general aspects of operations and manufacturing management, business excellence, quality improvement, and performance measurement.