The Hawthorne effect on surgical studies.

C Demetriou[1], L Hu[1], TO Smith[2], CB Hing[3]

Running Head: The Hawthorne effect on surgical studies

1: South West London Elective Orthopaedic Centre (SWLEOC), Dorking Road, Epsom KT18 7EG

2: Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Windmill Road, Oxford, OX3 7LD

3: St George's University Hospitals NHS Foundation Trust, Blackshaw Road, Tooting, London SW17 0QT

Charis Demetriou, MBBChir, Research Fellow, SWLEOC, Email: charis.demetriou@nhs.net

Lisi Hu, MBBS, Research Fellow, SWLEOC, Email: lisi.hu@nhs.net

Toby Smith, MSc, PhD, Senior Researcher in Rehabilitation, NDORMS, Email: toby.smith@ndorms.ox.ac.uk

Caroline Hing, MBBS, MD, Trauma and Orthopaedics Consultant, St George's University Hospitals NHS Foundation Trust, Blackshaw Road, Tooting, London SW17 0QT,

Telephone Number: 020 8672 1255 Email: caroh2712@me.com, Corresponding Author

Abstract Word Count: 250 Words, Main Text Word Count: 2069 Words,

Figures/Tables: 8 and 1 Supplementary Table

ABSTRACT

Background: The Hawthorne effect or 'observer effect' describes a change in normal behaviour when individuals are aware they are being observed. This may have an impact on effects estimates in clinical trials. The purpose of this study was to determine if the Hawthorne effect had been recorded as a risk of bias in surgical studies.

Methods: A PRISMA compliant literature search was conducted to March 2019. Eligible studies included those reporting or not reporting the Hawthorne effect in surgical studies from the following databases: MedLine, EMBASE, CINAHL, AMED, BNI, HMIC, PsycINFO, Web of Science, Cochrane Library, Google Scholar and OpenGrey. Two reviewers independently reviewed the papers, extracted data and appraised study methods using the Newcastle Ottawa Scale or the Cochrane risk of bias tool. Data were analysed descriptively.

Results: 842 papers were identified, of which 16 were eligible. Six (37%) observational studies identified with the aim of measuring the Hawthorne effect on their outcome with five reporting that the Hawthorne effect was responsible for the improvements in outcomes and one reporting no change in outcome due to the Hawthorne effect.

Ten (63%) studies were identified, of which eight used the Hawthorne effect as an explanation to improvements seen in the control group or their secondary outcomes and two to compare their results with other studies.

Conclusions: There is considerable between-study heterogeneity in how the Hawthorne effect relates to surgical outcomes. Further consideration on reporting and considering the importance of the Hawthorne effect in the design of surgical trials is warranted.

Keywords: Hawthorne effect, Observer Effect, Surgery, Surgical Studies, Bias

INTRODUCTION

The Hawthorne effect or 'observer effect' describes the modification of activity when individuals are aware that they are being observed. The Hawthorne effect was used as a term to explain the change in behaviour seen as a result of being observed. [1,2]

The Hawthorne effect has since been interpreted in different ways in industrial, social psychology and healthcare studies [1,3,5-7]. It can act in different ways in research, either influencing the behaviour of the participants by direct observation by making them aware of being studied or by answering questionnaires[4,5]. It is suggested that the awareness of the participants of being observed leads to a generation of beliefs around outcomes expected by the researchers or observers, leading to a change in their natural behaviour[4]. The Hawthorne effect can also affect the behaviour of the researchers providing or assessing the intervention in a study.

The Hawthorne effect can undermine the generalisability and the external validity of medical studies[6]. Wartolowska et al. reported non-specific effects with participating in surgical trials such as an interaction with healthcare staff and bias from participation in the study[7]. These effects can lead to biases within studies and thus lead to incorrect conclusions regarding the effectiveness of surgical interventions[7].

The use of the Hawthorne effect in surgical studies has received little attention. There remains uncertainty as to its impact within surgical studies and how this has been considered in such trials. The aim of this systematic review was to identify whether the Hawthorne effect was recorded as a risk of bias and whether the effect was measured in relation to surgical outcomes.

METHODS

Search strategy and study identification

A systematic literature search was performed to 13th March 2019 using the databases: MedLine, EMBASE, CINAHL, AMED, BNI, HMIC, PsycINFO, Web of Science, Cochrane Library, Google Scholar. OpenGrey was searched for grey literature relating to the Hawthorne effect in Surgery. The PRISMA guidelines were followed for this systematic review. Two reviewers (CD, LH) independently reviewed citations and assessed study eligibility based on the following criteria:

Inclusion criteria

1. The intervention affected by the Hawthorne effect must be surgical or related to the technical steps in a surgical procedure.
2. Measuring the Hawthorne effect was stated in the aims or objectives of the study or offered as an explanation for a reported change in an outcome either clinically relevant or affecting the technical steps of a surgical procedure.
3. Randomised and non-randomised clinical trials (RCTs) or observational studies.
4. The studies must have clinically relevant post-operative outcomes for a surgical procedure.
5. Full-text papers written in English.

Exclusion criteria

1. Hawthorne effect is not used to explain any study outcomes.
2. Exclude studies related to anaesthesia, i.e. administration of IV drugs intra-operatively, observation monitoring intra-operatively.
3. Exclude studies for dental surgery.
4. Exclude studies of invasive procedures performed in medicine.
5. Non primary research articles, such as reviews or study protocols.

Studies that did not have the Hawthorne effect in their aims or objectives but fulfilled the remaining inclusion or exclusion criteria were retrieved and the whole article was reviewed to identify if the Hawthorne effect was mentioned in the study.

An example of the search strategy is shown in **Table 1**. The search terms used for each database are shown in **Supplementary Table 1**.

Data Extraction

Data were extracted by two reviewers (CD, LH). The demographics of the patients included in each study, the main intervention and the comparator used in each group were recorded. For all studies that aimed to measure the Hawthorne effect their primary outcome was extracted as reported in the original study and whether the Hawthorne effect was measured or not. For all studies that offered the Hawthorne effect as a possible explanation for a secondary outcome, the secondary outcome affected and any explanation about how that outcome had occurred were recorded.

Outcome

The primary outcome was the frequency to which studies reported and/or quantified the Hawthorne effect as a potential bias for their results. The presence of the Hawthorne effect was recorded if the authors of the study had provided quantitative information showing a possible effect on the outcome affected. No *a priori* secondary outcomes or subgroup analyses were planned.

Assessment of Methodological Quality

Methodological quality was assessed independently by two reviewers (CD, LH). Disagreements were resolved through discussion and consensus. The Newcastle Ottawa Scale[8,9] was used to assess the quality of cohort studies and the Cochrane risk of bias tool version 1.0[10,11] was used for the assessment of RCTs.

Data Analysis

The frequency to which the Hawthorne effect was reported and/or quantified was determined and presented as a frequency (percentages). The characteristics of trials reporting the Hawthorne effect were described using descriptive statistics (frequency, mean, standard deviation, median) to answer the research question.

RESULTS

Search Results

842 papers were identified after excluding duplicates. After screening, 16 studies were eligible for inclusion in the final review. The PRISMA flowchart summarising how the studies included in this review were identified is shown in **Figure 1**. The studies included in the review are summarised in **Tables 2 - 6**. We identified a study by Ikpeze et al.[12], that was very similar to Buckley et al.[13], which had also looked at the QuickDASH score before and after consenting for a carpal tunnel release. This study did not identify a Hawthorne effect. We have not included this study in the summary table as it was considered to be a duplicate of Buckley et al[12,13].

Quality Assessment of the studies

Four of the six cohort studies measuring the Hawthorne effect were of poor quality[14–17]; two cohort studies were of good quality[13,18] (**Table 7**). A recurrent limitation with the cohort studies was poor matching of demographic characteristics when comparable analyses were undertaken. One of the interventions in Agarwal et al. comprised decolonisation of patients prior surgery, with the reduction in infection rates being statistically significant only after the use of decolonisation[15]. Three studies had shorter follow-up periods post-intervention, thus making it possible that the improved outcomes were due to inadequate length of follow-up[15,16,17].

The observational studies that mentioned the Hawthorne effect as an explanation for some of their outcomes were all good quality studies[19–22]. Five RCTs were of poor

quality, mainly due to personnel not being blinded to the intervention[23–27] and one was of fair quality[28] as shown in **Table 8**.

Demographics of the studies included

The 16 included studies reported data from 10,432 adults. Four trials were based in orthopaedics (three observational studies; one RCT), three in obstetrics (two observational studies; one RCT), three in neurosurgery (two observational studies; one RCT), three in general surgery (one observational study; two RCTs), one in cardiothoracics (observational), ENT (observational) and maxillofacial surgery (RCT). Thirteen studies reported participant gender, being 3178 females (69%) and 1409 males (31%)[13,14,16,17,19–26,28]. The mean age of participants was 61.1 years (**Table 2)**.

The use of the Hawthorne effect

Six observational studies (37%) were identified with the aim of measuring the Hawthorne effect on their outcome. Five studies suggested the Hawthorne effect as the main reason for improvement in the study outcomes[14–18], one reported no Hawthorne effect[13]. These studies are summarised in **Table 3**.

Ten studies (63%) were identified, including six RCTs and four observational studies that use the Hawthorne effect as an explanation to secondary study outcomes or to compare with results outside their study. These studies are summarised in **Tables 4, 5 and 6**. Two studies (13%) used the Hawthorne effect to explain difference in results reported by their study when compared with other studies in the literature[19,24] (**Table 5**). The other eight studies used the Hawthorne effect to explain some of their outcomes, with

four (25%) using the Hawthorne effect as a justification for unexpected improvements seen in their control groups[20,23,25,27] (**Table 4**) and four (25%) using the Hawthorne effect to justify improvement in subjective outcomes reported by the patients[21,22,26,28] (**Table 6**).

DISCUSSION

The results from this systematic review suggest that there are three general trends for the acknowledgement of the Hawthorne effect in surgical studies: 1) studies that acknowledged the possible bias of the Hawthorne effect on their finding and thus tried to quantify it[13–18], 2) studies that mentioned the Hawthorne effect as a way to justify unexpected results in their studies[20–23,25–28] and 3) studies that used the Hawthorne effect to explain differences seen between their results and the results of similar studies[19,24].

A very heterogeneous group of surgical studies was included in this review with different outcome measures affected by the Hawthorne effect in each study. From the studies that aimed to quantify Hawthorne effect (**Table 3**) there is some evidence that the Hawthorne effect can affect the behaviour of healthcare staff and the way they deliver interventions, thus leading to improved outcomes. However, since most of these studies were of poor quality **(Table 8)** and different outcomes were recorded in each study, no estimations can be made regarding the size of the Hawthorne effect on the

outcome measured. Not much can be said about any effect on patient participation as there was only one study that measured this with no Hawthorne effect seen[13].

Kwaan et al., Argudo et al., Teernstra et al. and Nakayama et al. had noticed unexpected improvements in their outcomes in the control group when compared to similar outcomes reported in their centres in earlier studies or expected results for some of their cohorts and thus, attributed these results to the Hawthorne effect due to better care and observation by the healthcare staff involved in the study[20,23,25,27] (**Table 4**). This highlights that the Hawthorne effect can occur when healthcare staff are aware that patients are part of a study thus affecting the outcomes and the validity of a study.

Two studies used the Hawthorne effect as a possible explanation for a difference in results seen in similar studies, however technical differences between the studies could also explain the different results[19,24](**Table 5**). This highlights what has been reported in the reviews by McCambridge and Nguyen et al. about the incorrect use of the Hawthorne as justification for unexpected results[2,4]. The remaining four studies have used the Hawthorne effect as explanation for improved outcomes reported by the patients[21,22,26,28] (**Table 6**). In the case of Bradley et al., a double blinded RCT, the improvement in WOMAC scores in the sham intervention group is more likely to be related to a placebo effect but Hawthorne effect could have partly contributed[28]. Roland et al. and Thornes et al. were cohort studies and Gong et al. was an RCT with unblinded patients, thus the Hawthorne effect is more likely to have affected the patient reported outcomes rather than placebo effect[21,22]. As seen in the two good quality studies in **Table 3**, the Hawthorne effect was reported to improve outcomes when an

objective measure was used (Tip-Apex Distance)[18] but when using Patient Reported Outcome Measures (PROMs) no Hawthorne effect was seen[13]. This highlights that the correct use of validated questionnaires without preloading them with expectations by the researchers leads to accurate results.

To quantify the Hawthorne effect, results need to be recorded from a retrospective time period that the personnel delivering the interventions and the patients involved were unaware of the study taking place and a prospective period in which the personnel involved in the study are aware that data collection is taking place. As McCambridge et al. have noted there is potential for research participation bias to occur in a study due to the interaction of the participation effect on the intervention and this form of bias will not be eliminated completely by randomisation[29]. A proposed study design to overcome research participation bias is the Solomon four-group design, with assessed and unassessed, hence unaware of the study, control and intervention groups[2,30,31].

To conclude, the Hawthorne effect is generally under-recognised as a source of potential bias in surgical studies. The Hawthorne effect has been used loosely or inappropriately in some of the studies. Most of the studies in this review were of poor quality and with very heterogeneous outcomes thus we cannot conclude much about the size of the Hawthorne effect and its influence on outcomes. However, there is some evidence that Hawthorne effect can potentially bias the results of a study either through behaviour modification of healthcare staff or the patients involved. Further well-designed studies

are required to try and measure the size of this effect and identify whether it is a significant bias in surgical studies.

Disclosure Statement:

The authors of this systematic review on the Hawthorne effect on surgical studies have no conflict of interest to declare. No funding was received for the preparation of this review.

References:

1. Kompier, M. A. J. The Hawthorne effect is a myth, but what keeps the story going? *Scand. J. Work. Environ. Health* **32**, 402–12 (2006).

2. Nguyen, V. N., Miller, C., Sunderland, J. & McGuiness, W. Understanding the Hawthorne effect in wound research-A scoping review. *Int. Wound J.* **15**, 1010–1024 (2018).

3. Chiesa, M. & Hobbs, S. Making sense of social research: how useful is the Hawthorne Effect? *Eur. J. Soc. Psychol.* **38**, 67–74 (2008).

4. McCambridge, J., Witton, J. & Elbourne, D. R. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J. Clin. Epidemiol.* **67**, 267–77 (2014).

5. McCambridge, J. & Kypri, K. Can Simply Answering Research Questions Change Behaviour? Systematic Review and Meta Analyses of Brief Alcohol Intervention Trials. *PLoS One* **6**, e23748 (2011).

6. Sedgwick, P. & Greenwood, N. Understanding the Hawthorne effect. *BMJ* **351**, h4672 (2015).

7. Wartolowska, K. *et al.* Use of placebo controls in the evaluation of surgery: systematic review. *BMJ* **348**, g3253 (2014).

8. GA Wells, B Shea, D O'Connell, J Peterson, V Welch, M Losos, P. T. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. (Accessed: 24th April 2019)

9. Deeks, J. *et al.* Evaluating non-randomised intervention studies. *Health Technol. Assess. (Rockv).* **7**, (2003).

10. Higgins JPT, G. S. (editors). *l Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.*

11. Higgins, J. P. T. *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* **343**, d5928 (2011).

12. Ikpeze, T. C., Childs, S., Buckley, T. & Elfar, J. C. Validity of QuickDASH at day of surgery versus day of initial consultation: Does informed consent make a difference? *J. Orthop. Surg.* **26**, 230949901877789 (2018).

13. Buckley, T., Mitten, D. & Elfar, J. The Effect of Informed Consent on Results of a Standard Upper Extremity Intake Questionnaire. *J. Hand Surg. Am.* **38**, 366–371 (2013).

14. Zhang-Rutledge, K. *et al.* An Initiative to Reduce the Episiotomy Rate:

Association of Feedback and the Hawthorne Effect With Leapfrog Goals. *Obstet. Gynecol.* **130**, 146–150 (2017).

15. Agarwal, N. *et al.* Reducing surgical infections and implant costs via a novel paradigm of enhanced physician awareness. *Neurosurgery* **82**, 661–669 (2018).

16. Leung, W. C., Lam, H. S. W., Lam, K. W., To, M. & Lee, C. P. Unexpected reduction in the incidence of birth trauma and birth asphyxia related to instrumental deliveries during the study period: was this the Hawthorne effect? *BJOG An Int. J. Obstet. Gynaecol.* **110**, 319–322 (2003).

17. Borer, A; Gilad, J; Meydan, N; Riesenberg, K; Schlaeffer, F; Alkan, M; Schlaeffer, P. Impact of active monitoring of infection control practices on deep sternal infection after open-heart surgery. *Ann. Thorac. Surg.* **72**, 515–520 (2001).

18. Kennedy, M. T. *et al.* The use of weekly departmental review of all orthopaedic intra-operative radiographs in order to improve quality, due to standardized peer expectations and the "Hawthorne effect". *Surg.* **11**, 10–13 (2013).

19. Sjåvik, K. *et al.* Assessment of drainage techniques for evacuation of chronic subdural hematoma: a consecutive population-based comparative cohort study. *J. Neurosurg.* 1–7 (2017).

20. Argudo, N. *et al.* The use of an algorithm for prophylactic mesh use in high risk patients reduces the incidence of incisional hernia following laparotomy for colorectal cancer resection. *Cir. Esp.* **95**, 222–228 (2017).

21. Roland, P. S., Shoup, A. G., Shea, M. C., Richey, H. S. & Jones, D. B. Verification of Improved Patient Outcomes With a Partially Implantable Hearing Aid, The SOUNDTEC Direct Hearing System. *Laryngoscope* **111**, 1682–1686 (2001).

22. Thornes, B., Shannon, F., Guiney, A. M., Hession, P. & Masterson, E. Suture-button syndesmosis fixation: Accelerated rehabilitation and improved outcomes. *Clin. Orthop. Relat. Res.* 207–212 (2005).

23. Kwaan, M. R. *et al.* Abdominal closure protocol in colorectal, gynecologic oncology, and urology procedures: A randomized quality improvement trial. *Am. J. Surg.* **211**, 1077–1083 (2016).

24. Wilson, L. K., Sullivan, S., Goodnight, W., Chang, E. Y. & Soper, D. The use of blunt needles does not reduce glove perforations during obstetrical laceration repair. *Am. J. Obstet. Gynecol.* **199**, 639.e1-639.e4 (2008).

25. Teernstra, O. P. M. *et al.* Stereotactic Treatment of Intracerebral Hematoma by Means of a Plasminogen Activator: A Multicenter Randomized Controlled Trial (SICHPA). *Stroke* **34**, 968–974 (2003).

26. Gong, X. *et al.* Application of a Computer-Assisted Navigation System (CANS) in the Delayed Treatment of Zygomatic Fractures: A Randomized Controlled Trial.

J. Oral Maxillofac. Surg. **75**, 1450–1463 (2017).

27. Nakayama, H. *et al.* Subcutaneous drainage to prevent wound infection in liver resection: a randomized controlled trial. *J. Hepatobiliary. Pancreat. Sci.* **21**, 509–517 (2014).

28. Bradley, J. D. *et al.* Tidal irrigation as treatment for knee osteoarthritis: A sham-controlled, randomized, double-blinded evaluation. *Arthritis Rheum.* **46**, 100–108 (2002).

29. McCambridge, J., Kypri, K. & Elbourne, D. In randomization we trust? There are overlooked problems in experimenting with people in behavioral intervention trials. *J. Clin. Epidemiol.* **67**, 247 (2014).

30. Solomon, R. L. An extension of control group design. *Psychol. Bull.* **46**, 137–150 (1949).

31. McCambridge, J., Butor-Bhavsar, K., Witton, J. & Elbourne, D. Can Research Assessments Themselves Cause Bias in Behaviour Change Trials? A Systematic Review of Evidence from Solomon 4-Group Studies. *PLoS One* **6**, e25223 (2011).

List of figures and tables:

List of Supporting Information:

Figure 1:

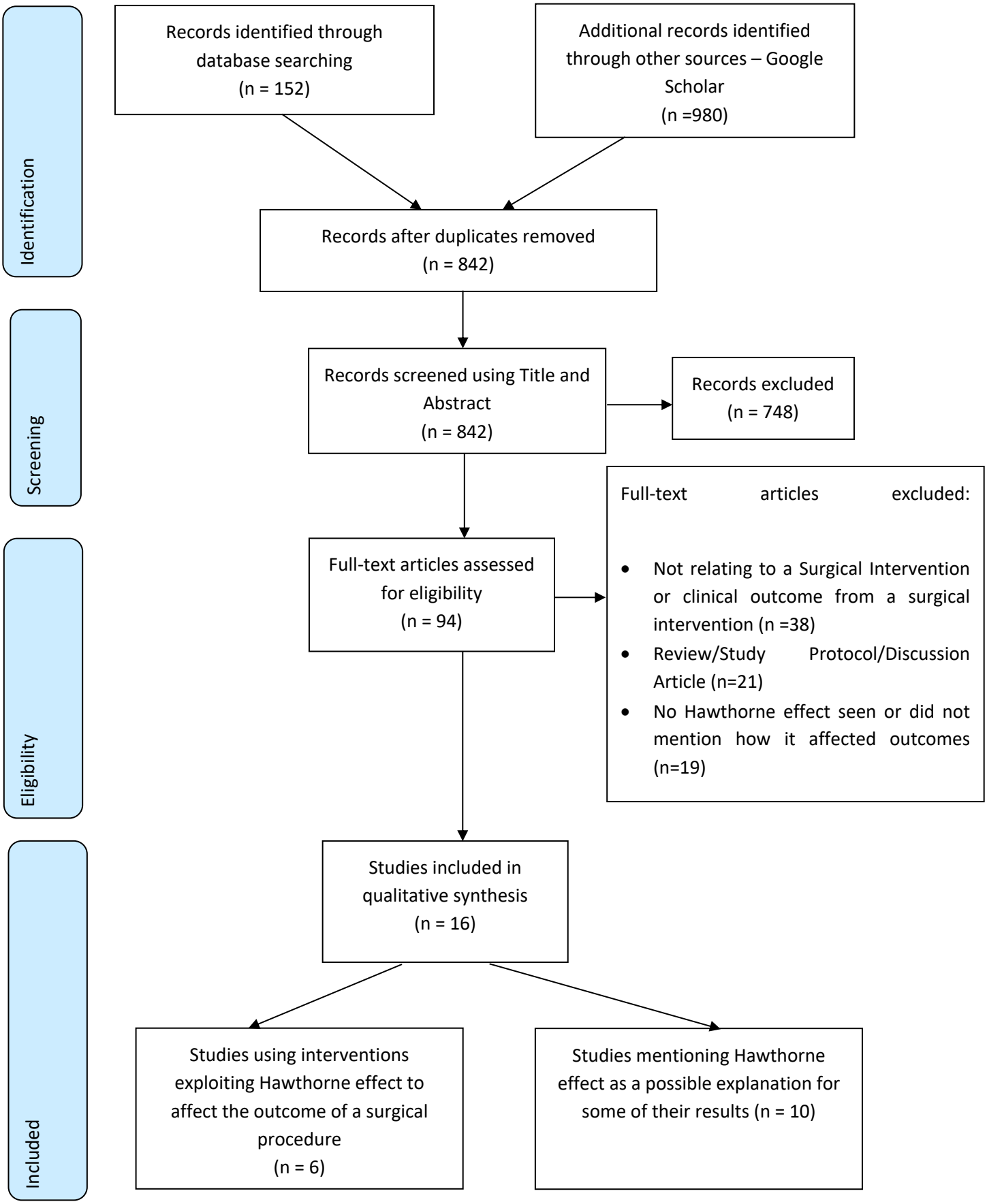
Identification
Records identified through database searching (n = 152)
Additional records identified through other sources – Google Scholar (n =980)
Records after duplicates removed (n = 842)
Screening
Records screened using Title and Abstract (n = 842)
Records excluded (n = 748)
Eligibility
Full-text articles assessed for eligibility (n = 94)
Full-text articles excluded:
Not relating to a Surgical Intervention or clinical outcome from a surgical intervention (n =38)
Review/Study Protocol/Discussion Article (n=21)
No Hawthorne effect seen or did not mention how it affected outcomes (n=19)
Included
Studies included in qualitative synthesis (n = 16)
Studies using interventions exploiting Hawthorne effect to affect the outcome of a surgical procedure (n = 6)
Studies mentioning Hawthorne effect as a possible explanation for some of their results (n = 10)

Table 1:

DATABASE	SEARCH TERMS
MedLine	1: “EFFECT MODIFIER, EPIDEMIOLOGIC”/
	2: (“hawthorne effect”).ti,ab
	3: (1 OR 2)
	4: exp “SPECIALTIES, SURGICAL”/ OR exp “GENERAL SURGERY”/ OR exp “SURGICAL PROCEDURES, OPERATIVE”/
	5: (3 AND 4)

Table 2:

Authors	Type of Study	Specialty	Patient Demographics
Zhang-Rutledge et al.[14]	Prospective Observational Quality Improvement Project	Obstetrics and Gynaecology	1176 female patients
Kennedy et al.[18]	Retrospective Observational Study	Trauma and Orthopaedics	198 patients Group A pre-intervention: 105 patients, Group B post-intervention: 93 patients Mean age: 71.4 years (Range 16-98)
Agarwal et al.[15]	Prospective Observational Study	Neurosurgery	5387 patients
Leung et al.[16]	Prospective Observational Study	Obstetrics and Gynaecology	670 women
Borer et al.[17]	Prospective Observational Study	Cardiothoracic Surgery	118 patients Males: 80, Females:38 Mean age: 62 years (Range 30-82)
Buckley et al.[13]	Both Retrospective and Prospective Observational Study	Trauma and Orthopaedics	74 patients Retrospective cohort: 39 patients, Prospective cohort: 35 patients Females: 47, Males: 27, Mean Age:56 years
Kwaan et al.[23]	Randomised Controlled Trial	Colorectal surgery, Urology, Gynaecology	233 patients (121 in the control group and 112 in the intervention group) Males:86, Females: 147, Mean age 57.5 years
Argudo et al.[20]	Prospective Cohort Study	Colorectal Surgery	235 patients (166 patients had the algorithm applied, 69 patients in the control group) Males: 145, Females:90, Mean age:69.7 years
Teernstra et al.[25]	Randomised Controlled Trial	Neurosurgery	70 patients (36 in the surgical group and 34 in the control group) Males: 40, Females:30, Mean age 31 years
Nakayama et al.[27]	Randomised Controlled Trial	Hepatobiliary Surgery	260 patients (131 in the drainage group and 129 in the non-drainage group) Mean age: 66.5 years
Sjavik et al.[19]	Retrospective Cohort study	Neurosurgery	1260 patients, Males: 878, Females: 372 Mean age: 73.3 years
Wilson et al.[24]	Randomised Controlled Trial	Obstetrics and Gynaecology	438 female patients Mean age: 24.1 years
Gong et al.[26]	Randomised Controlled Trial	Maxillofacial Surgery	78 patients (39 in the control group and 39 in the intervention group) Males: 62, Females:16, Mean age: 31 years (Range 16-60)
Bradley et al.[28]	Randomised Controlled Trial	Trauma and Orthopaedics	180 patients (91 in the sham irrigation group and 89 in the tidal group) Females: 124, Males: 56, Mean age: 55.7 years
Roland et al.[21]	Prospective Within-subjects repeated measures design	ENT	23 patients Females:13, Males:10, Mean age:67.1 years
Thornes et al.[22]	Prospective Cohort Study	Trauma and Orthopaedics	32 patients (16 patients in the suture button and 16 patients in the syndesmosis screw) Males:25, Females:7, Mean age: 31.5 years (Range 17-74)

Table 3:

Study	Comparisons	Application of the Hawthorne effect/Intervention	Outcome Measures affected by the Hawthorne effect	Summary of Findings	Hawthorne Effect reported
Zhang-Rutledge et al.[14]	Episiotomy rates before during and after the intervention.	Monthly episiotomy rates recorded without publication or announcement prior to intervention to establish baseline rates. Education of guidelines and feedback of both individual and departmental episiotomy rates were delivered at monthly meetings. In the final six months, individual episiotomy rates were no longer provided.	Departmental episiotomy rates	Baseline episiotomy rate was 9%. After education and monthly departmental performance reports, the rate dropped to 5.9%. After introducing monthly individual episiotomy rates for 6 months, the rated further dropped to 4.4%. The change was sustained for the six months where individual feedback was omitted.	Yes
Kennedy et al.[18]	Tip – Apex Distance (TAD) in the post-op X- rays before and after the introduction of the weekly departmental meetings.	Weekly review of the post-operative DHS X-rays at the departmental meeting. Prior to this the post-operative DHS X-rays were reviewed by the Consultant Surgeon present in theatre on the day of the operation.	TAD in post-operative X- rays for DHS. AP and lateral Distance, Patients with TAD >25mm	AP Distance (mm) Group A: 9.29 +/- 2.85, Group B: 7.33 +/- 2.11, p< 0.0001 Lateral Distance (mm) Group A: 9.52 +/- 3.40, Group B: 7.62 +/- 2.31, p < 0.0001 TAD (mm) Group A 18.81 +/- 5.65, Group B 14.95 +/- 4.01, P< 0.0001 Total number with TAD > 25 mm Group A: 15, Group B: 1, P< 0.0001	Yes
Agarwal et al.[15]	Infection rates before and after intervention (physician education only and education with the decolonisation treatment) for craniotomies. Infection rates before and after physician education alone for ventricular shunt insertion.	From May 2015 to April 2016 all the surgeons were informed of their individual post-operative infection incidence and how this compared with their colleagues at the departmental meetings. From December 2015 to April 2016, physician education and formal infection prevention programme was introduced to identify and decolonise Staphylococcus aureus prior to surgery. Physicians were also made aware the cost of ventricular shunts and alternative devices.	Craniotomy infection incidence rate, ventricular shunt infection incidence rate, combined craniotomy infection and ventricular shunt incidence rate.	Craniotomy Infection Incidence rate prior to intervention: 3.0% Craniotomy Infection Incidence rate after education only: 2.4%, p= 0.471 Craniotomy Infection Incidence rate after education + decolonisation: 2.0%, p= 0.104 Ventricular Shunt Incidence rate prior to intervention: 3.7% Ventricular Shunt Incidence rate after Education: 2.5%, p= 0.327 Combined Craniotomy and Ventricular Infection Incidence rate prior to intervention: 3.2% Craniotomy and Ventricular Infection Incidence rate after education and decolonisation for craniotomies: 2.1%, p=0.041	Yes
Leung et al.[16]	Birth trauma and birth asphyxia rates related to instrumental deliveries before and after the intervention	A codesheet was designed to be used in theatres for characteristics of labour, pelvic examination findings prior to attempting instrumental delivery and neonatal outcomes.	Birth asphyxia and birth trauma rates	Prior to intervention, the birth trauma and birth asphyxia rate was 2.8%. Post-intervention this has dropped to 0.6%. RR=0.27, 95% CI 0.11–0.70	Yes
Borer et al.[17]	Deep Sternal Infection (DSI) rate during the study and 6 months after study completion with DSI rate prior to the study (18 months prior).	Active monitoring of infection control practices in operating theatres and intensive care units by 3 nurses using a specially designed monitoring questionnaire (monitored infection control practices by surgeons, anaesthetists, theatre staff and cardiopulmonary bypass technicians). Surgeons were not made aware about the questions in the questionnaire and were not notified in advance which procedures would be monitored.	DSI and Infection Control Practices between the two study periods	Improved infection control practices between the two study periods in the operating theatres Significant reduction in the rate of DSI in the 6 months after the study when compared to the rate before the study (10% prior, 5.1% during study period - p=0.14, 2.8% 6 months after – p=0.007)	Yes

Buckley et al.[13]	QuickDASH score between retrospective and prospective cohorts both before and after the procedure. Compared pre-operative score pre- and post-consent in the prospective cohort.	Retrospective cohort identified patients who completed both pre- and post-operative questionnaires (QuickDASH). The prospective cohort were enrolled on the day of surgery and made aware of the study aiming to ascertain the Hawthorne effect. Pre- and post-operative questionnaires were completed in the same way as the retrospective cohort. After consenting to enrolment, a second pre-operative questionnaire was completed.	QuickDASH score after patients have consented to enrol in the study	Preoperative QuickDASH: Retrospective: 40, Prospective: 40 – after consent NS, p=0.86 Postoperative QuickDASH: Retrospective: 27, Prospective: 19 NS, p=0.41 Prospective Cohort, Pre-operative QuickDASH score: Pre-consent 39.0, After Consent 39.7 p=0.98	No

Table 4:

Study	Intervention	Comparisons/Primary Outcome	Outcome that the authors attributed to Hawthorne effect	Explanation of the noticed outcome
Kwaan et al.[23]	Development of an abdominal closing tray protocol, which involved the following steps: 1) Instruments, sponges, suction tips, and devices, including electrocautery were removed from the surgical field. 2) All surgical personnel at the operative field changed their gloves. A surgical gown change was optional. 3) The operative field was re-draped with freshly opened sterile towels or half-sheets. 4) A sterile closing tray was opened onto an unused sterile surface and only those instruments and sutures were used for the remainder of the procedure.	Control Group: Usual standard of care for closing the laparotomy wound Intervention Group: adoption of the abdominal closing tray protocol Primary End point: Surgical Site Infections (SSI) at one month post-operatively	The SSI rate for both groups in this study was 50% lower (12%) compared to the SSI rate in earlier years (24%) at the same centre. There was no statistically significant difference in SSI rates between the 2 groups.	Possibly related to unmeasured changes to daily practice among the surgeons and providers during the study (Hawthorne effect)
Argudo et al.[20]	Application of a decision algorithm to decide which patients require prophylactic mesh augmentation of the laparotomy incision to prevent incisional hernia. Patients who were considered low risk for developing incision hernia underwent closure with simple suture.	Patients with decision algorithm vs patients where the algorithm was not used. Primary outcome: Incidence of incisional hernia during follow-up	The authors have reported that in the low-risk group, the incisional hernia rate (14.3%) was lower than the rate (31.1%) seen in the same centre in a previous retrospective study.	They have attributed this to the Hawthorne effect and the fact the being part of this study has led to improved quality of the abdominal wall closure by the surgeons.
Teernstra et al.[25]	Surgical intervention involved stereotactically placed catheter in the centre of the haematoma, injecting 5000IU of urokinase and gentle suction of the liquefied haematoma after 6 hours. This was repeated for 8 times over 48 hours.	Non-surgical group had standard medical care. Surgical group had urokinase injections as explained. Primary Outcome: Mortality rate at 6 months No statistically significant difference between the 2 groups in mortality rates	The predicted mortality used for this study was 88%. The observed mortality in the non-surgical group was 59% and 56% for the intervention group.	The authors have attributed the reduction in mortality in the non-surgical group to the Hawthorne effect and the increased monitoring of these patients by trial coordinators and the monitoring committee at regular intervals, which might have caused an overall increase in supportive care for these patients
Nakayama et al.[27]	The drainage group underwent hepatectomy with closed irrigation drain inserted intra-operatively. 10 Fr drain was placed subcutaneously and connected to a low pressure (under 20- 80 cm H2O) aspiration reservoir to allow drainage of the full length of the wound The non-drainage group did not have the subcutaneous drain inserted.	Primary Outcome: Superficial or deep surgical site infection within 30 days post-surgery between drainage and non-drainage groups. No statistically significant difference in wound infection between the 2 groups.	The authors reported that the wound infection incidence rate in this study has fallen by 3% compared to the retrospective data they have for wound infection rates at their centre.	They attributed this decrease in wound infection rates to the Hawthorne effect.

Table 5:

Study	Intervention	Comparisons/Primary Outcome	Outcome that the authors attributed to Hawthorne effect	Explanation of the noticed outcome
Sjavik et al.[19]	Comparison of three drainage techniques: continuous irrigation and drainage (n=166), passive subdural drainage (n=330) and active subgaleal drainage (n=764)	Comparisons between the 3 treatment groups Primary End point: Recurrence of haematoma requiring reoperation within 6 months of index surgery	Recurrence of haematomas in the passive drainage group was 20% (66 patients). The authors compared the recurrence rate for passive drainage of chronic subdural haematomas with an RCT that reported a recurrence rate of 9.3%.	The authors have mentioned that this difference between the 2 studies might be partly explained by the Hawthorne effect as applied to patients in the RCT. However, there were technical differences between the 2 studies: the surgeons in the RCT used 2 burr holes with drain removal at 48h compared to 1 burr hole and drain removal at 24h in this study.
Wilson et al.[24]	Patients were randomised to either blunt or sharp needles to repair obstetric lacerations. Surgeon gloves were collected immediately after the procedure to assess for perforation by needles.	Control Group: Using Sharp needles to repair obstetric lacerations Intervention (n=221) Group: Using Blunt Needles to repair obstetric lacerations (n=217) Primary End point: Glove Perforation assessed at the end of the procedure using a validated water test	5 glove perforations in the sharp needles group (2.26%) and 4 in the blunt needles group (1.84%) Relative Risk, 0.79 (95% CI, 0.2-2.95), not statistically significant. The authors compared with other studies that had perforation rates between 10-20%	They have mentioned that the difference in perforation rates between the studies might be due to the Hawthorne effect. However, they reported that in 2006 the FDA has lowered the acceptable rate of surgical glove defects from 2.5% to 1.5%, thus there are fewer pre-existing defects in surgical gloves with modern manufacturing technique.

Table 6:

Study	Intervention	Comparisons/Primary Outcome	Outcome that the authors attributed to Hawthorne effect	Explanation of the noticed outcome
Gong et al.[26]	Marker-assisted surgical navigation intra-operatively in conjunction with computer-assisted design steps using pre-operative computed tomography (CT).	Control Group: No navigation system used intraoperatively Intervention Group: Computer assisted navigation system used intraoperatively Primary Outcome: Absolute bilateral differences of the ZMC eminence and width based on CT measurements 48 to 72 hours after surgery	Visual Analogue Score used to subjectively evaluate the postoperative recovery of facial soft tissue symmetry. Clinician median VAS was higher for the navigation group (8 vs 7; P = 0.043). Patients median VAS was not significantly different between the groups (9 vs 8; P = 0.328).	Authors attributed the difference between clinician and patients VAS to the Hawthorne effect No clear explanation as how this might be attributed to the Hawthorne effect.
Bradley et al.[28]	Tidal Irrigation: 14-gauge needle inserted into the knee capsule via the lateral suprapatellar port and 30-50ml aliquots of saline were injected into the knee and aspirated repeatedly until 1 litre of saline washed the knee joint. Sham Irrigation: 14-gauge needle advance up to the capsule via the lateral suprapatellar port but did not puncture the knee capsule. Aliquots of 40-50ml saline were injected in the subcutaneous tissue and aspirated back until 1 litre of saline has passed through.	Change in pain and function domains of the WOMAC score over the next 3, 6 and 12 months between the tidal and the sham irrigation groups There was no statistically significant difference in the WOMAC scores between the 2 groups.	The authors have noticed a slightly greater improvement (not statistically significant) in the WOMAC scores from baseline that was sustained over the study period	The authors attributed this to both the placebo effect and the Hawthorne effect.
Roland et al.[21]	Implantation of the SOUNDTEC Direct system	Evaluation of the patients pre- and post-implantation Using objective measurements and subjective questionnaires for both study periods and compared the results pre- and post-implantation	The authors have not noticed any objective evidence of improved speech perception in quiet and noise with the SOUNDTEC system. The patients have reported increased satisfaction with the SOUNDTEC System (improved clarity, more natural sound, increased loudness)	The authors attributed the subjective increased satisfaction reported by the patients to placebo and Hawthorne effects.
Thornes et al.[22]	Patients in the control group received syndesmosis screw fixation. Patients in the intervention group had syndesmosis fixation with 2 endobuttons on the tibia and fibula side that were connected with number 5 braided polyester suture and were tightened around the syndesmosis.	Compared outcomes of these patients at 3- and 12-months post-op. The main outcome used was the American Orthopaedic Foot and Ankle Society (AOFAS) score. Patients with endobutton fixation had a statistically significant improvement in the mean AOFAS score at 3 and 12 months.	Patients in the endobutton suture group reported higher satisfaction with the outcome at 12 months compared to the patients in the syndesmosis screw group	The authors attributed the higher satisfaction in the endobutton suture group to the Hawthorne effect, i.e. patients were told they were receiving treatment using a new technique.

Table 7:

	Zhang-Rutledge et al.[14]	**Kennedy et al.**[18]	**Agarwal et al.**[15]	**Leung et al.**[16]	**Borer et al.**[17]	**Buckley et al.**[13]	**Sjavik et al.**[19]	**Argudo et al.**[20]	**Roland et al.**[21]	**Thornes et al.**[22]
Representativeness of the exposed cohort	no description	*	*	*	*	*	*	*	no	No description of the derivation of the cohort
Selection of the non- exposed cohort	*	*	*	*	no exact details of the non-exposed cohort before the study	*	*	*	*	*
Ascertainment of exposure	*	*	*	*	*	*	*	*	*	*
Demonstration that outcome of interest was not present at start of study	*	no	*	no	no	*	*	*	*	*
Comparability of cohorts on the basis of the design or analysis	no	*	no	no	no controls with pre-exposure group	*	**	**	**	**
Assessment of outcome	*	*	*	*	*	*	*	*	*	*
Was follow-up long enough for outcomes to occur	*	*	*	*	*	*	*	*	*	*
Adequacy of follow up of cohorts	*	*	*	*	no details about the patients in the non-exposure cohort prior to the study	*	*	*	*	*
QUALITY	**Poor Quality**	**Good Quality**	**Poor Quality**	**Poor Quality**	**Poor Quality**	**Good quality**	**Good quality**	**Good quality**	**Good quality**	**Good quality**

*: The characteristic assessed was present in the study

Table 8:

	Gong et al.[26]	**Kwaan et al.**[23]	**Wilson et al.**[24]	**Teernstra et al.**[25]	**Nakayama et al.**[27]	**Bradley et al.**[28]
Random Sequence Generation	low	low	low	low	low	low
Allocation Concealment	unclear	low	unclear	low	unclear	high
Selective Reporting	low	low	low	low	low	low
Other Bias	low	low	low	high (finished prematurely due to slow accrual)	low	low
Blinding of Participants and Personnel	High	High	High	unclear	unclear	low
Blinding of Outcome Assessment	low	low	low	unclear	low	low
Incomplete Outcome Data	low	low	low	unclear	low	low
Quality	**Poor**	**Poor**	**Poor**	**Poor**	**Poor**	**Fair**