



OPEN ACCESS

EDITED BY

Maria Isabel Fernandez-San-Martin,
University Institute for Primary Care
Research (IDIAP Jordi Gol), Spain

REVIEWED BY

Francesc Orfila,
Catalan Health Institute (ICS), Spain
Rocio Casañas,
Blanquerna - Ramon Llull University,
Spain

*CORRESPONDENCE

Christophe Berkhout
christophe.berkhout@univ-lille.fr

SPECIALTY SECTION

This article was submitted to
Family Medicine and Primary Care,
a section of the journal
Frontiers in Medicine

RECEIVED 31 August 2022

ACCEPTED 18 October 2022

PUBLISHED 08 November 2022

CITATION

Berkhout C, Berbra O, Favre J,
Collins C, Calafiore M, Peremans L
and Van Royen P (2022) Defining
and evaluating the Hawthorne effect
in primary care, a systematic review
and meta-analysis.
Front. Med. 9:1033486.
doi: 10.3389/fmed.2022.1033486

COPYRIGHT

© 2022 Berkhout, Berbra, Favre,
Collins, Calafiore, Peremans and Van
Royen. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Defining and evaluating the Hawthorne effect in primary care, a systematic review and meta-analysis

Christophe Berkhout^{1,2*}, Ornella Berbra¹, Jonathan Favre¹,
Claire Collins³, Matthieu Calafiore^{1,4}, Lieve Peremans^{2,5} and
Paul Van Royen²

¹Department of General Practice/Family Medicine, Université de Lille, Lille, France, ²Department of Family Medicine and Population Health, Universiteit Antwerpen, Antwerp, Belgium, ³Irish College of General Practitioners, Dublin, Ireland, ⁴ULR 2694 METRICS, Université de Lille, Lille, France, ⁵Department of Nursing and Midwifery, Universiteit Antwerpen, Antwerp, Belgium

In 2015, we conducted a randomized controlled trial (RCT) in primary care to evaluate if posters and pamphlets dispensed in general practice waiting rooms enhanced vaccination uptake for seasonal influenza. Unexpectedly, vaccination uptake rose in both arms of the RCT whereas public health data indicated a decrease. We wondered if the design of the trial had led to a Hawthorne effect (HE). Searching the literature, we noticed that the definition of the HE was unclear if stated. Our objectives were to refine a definition of the HE for primary care, to evaluate its size, and to draw consequences for primary care research. We designed a Preferred Reporting Items for Systematic reviews and Meta-Analyses review and meta-analysis between January 2012 and March 2022. We included original reports defining the HE and reports measuring it without setting limitations. Definitions of the HE were collected and summarized. Main published outcomes were extracted and measures were analyzed to evaluate odds ratios (ORs) in primary care. The search led to 180 records, reduced on review to 74 for definition and 15 for quantification. Our definition of HE is “an aware or unconscious complex behavior change in a study environment, related to the complex interaction of four biases affecting the study subjects and investigators: selection bias, commitment and congruence bias, conformity and social desirability bias and observation and measurement bias.” Its size varies in time and depends on the education and professional position of the investigators and subjects, the study environment, and the outcome. There are overlap areas between the HE, placebo effect, and regression to the mean. In binary outcomes, the overall OR of the HE computed in primary care was 1.41 (95% CI: [1.13; 1.75]; $I^2 = 97%$), but the significance of the HE disappears in well-designed

studies. We conclude that the HE results from a complex system of interacting phenomena and appears to some degree in all experimental research, but its size can considerably be reduced by refining study designs.

KEYWORDS

effect modifier/epidemiologic, scientific experimental error, systematic review, primary healthcare, Hawthorne effect

Introduction

By autumn of every year, the main French mandatory health insurance scheme conducts a promotional campaign for seasonal influenza vaccination in mass media and in health facilities. General practice surgeries can participate in this campaign by hanging posters and making pamphlets available in their waiting rooms. Advertising using posters and pamphlets in waiting rooms shows no evidence of effectiveness in terms of increasing knowledge or changing the health behavior of patients (1). We conducted a cluster-randomized controlled trial (RCT) with 10,597 patients assessing the 2014–2015 campaign in France confirming these findings (2). No difference was demonstrated in vaccination uptake between waiting rooms advertising for influenza vaccination (intervention) or not (controls) ($P = 0.561$). However, the immunization rate increased by about 3% in both arms of the trial compared to the baseline (previous year). At the same time, a decrease in coverage of 2.4% was observed district wide by public health authorities. As our trial targeted a change in behavior in primary healthcare, we considered the possibility of a Hawthorne effect (HE) to explain this difference and felt the need to have greater insight regarding this effect (3).

The Hawthorne effect (HE) was first observed in relation to six, partly overlapping, experiments carried out from 1924 to 1933 at the Hawthorne plant, a large factory complex of the Western Electric Company in Cicero (Illinois, USA), also reputed to have generated Al Capone's original fortune (4). The most thorough publication was issued by Roethlisberger and Dickson which presented data from the six experiments (5). Elton Mayo, a Harvard business professor, was not the director of the studies, but as he became the main interpreter of the Hawthorne experiments, his name remains associated with the research (6). The study group examined the effects of various incentives on the productivity of two groups of volunteer workers, and the good story was that whatever experiment was applied, the trend of productivity was upward in both groups (7). However, this does not fit with the two last experiments (6). The term "Hawthorne effect" or "observer effects" to describe the performance or behavior improvement of

people involved in research, arising exclusively when under observation, was first used in 1953 (8). In 1974, Parsons described the HE as a failure of the experimenters to realize how the consequences of subjects' performance affect what subjects do (9). Indeed, the internal validity of the Hawthorn experiments was biased by the selection of a small number of volunteer participants, attrition due to the removal of operators because of gross insubordination, and potential antagonism between management and employees (Dickson was an officer of the Western Electric Company) (6). In 2011, Levitt and List recovered the original results of the Hawthorne illumination experiments and reanalyzed the outcomes, finding "some weak evidence that workers respond more to experimental manipulations than to naturally occurring changes in light (10)."

In 2010, French and Sutton published a narrative review calling the changes in the people being measured in an experimental environment "measurement reactivity." They merged this designation with other terms including "assessment reactivity," "mere measurement," "question-behavior effect," or "self-generated validity" (11). Further, in 2017 Paradis and Sutkin recommended the use of the phrase "participant reactivity" when considering the triad participant, observer, and research question (12). One common point of all effects appearing in an experimental environment, whatever their designation, is the considerable heterogeneity of their size across studies (13, 14).

In 2014, McCambridge et al. published an often-cited systematic review to elucidate the existence of the HE, the conditions of its appearance, and its estimated size (15). They noted that it was relevant to clear the term HE in health sciences, as it was evoked in relation to a range of methodological phenomena. To define the HE, they stated that "awareness of being observed or having behavior assessed engenders beliefs about researcher expectations. Conformity and social desirability considerations then lead behavior to change in line with these expectations." They came to the conclusion that "Further research on this subject should be a priority for the health sciences, in which we might expect change induced by research participation to be in the direction of better health and thus likely to be confounded with the outcomes being studied (15)".

In 2020, Pursell et al. conducted a systematic review and meta-analysis regarding the HE in hand hygiene (HH), based on the many publications in the field related to the guidance for HH promoted by the World Health Organization (WHO) (“My Five Moments for Hand Hygiene” initiative) in 2009 (16). It confirmed the considerable heterogeneity in outcomes, with the HE ranging from -6.9 to 65.3%. Probably in line with this heterogeneity, they did not complete the meta-analysis (17). Hand-hygiene behaviors have markedly changed since the COVID-19 outbreak (18). For this reason, the outcomes regarding hand hygiene in hospital wards as in the community are probably outdated.

Noting the considerable inconsistency regarding the phenomenon, the primary objectives of this review were (1) to refine the definition of the HE and outline the progress of research since 2012 (last inclusions in McCambridge’s review) on the HE in terms of its existence and characteristics and (2) to estimate its size in primary care studies, expecting the already described heterogeneity.

Materials and methods

Eligibility criteria, information sources, and search strategy

Considering the definition, publications related to research in the medical field, in particular those regarding health professionals and patients, were included. Reports needed to contain a clear definition or outcome measuring the HE. Included methodologies were clinical trials and their reanalysis, quasi-experimental or observational studies, or historical comparisons. Reports published in French or English, with an available abstract, were included. Only reports published after the review by McCambridge were considered (publication range: January 2012 to March 2022). We ensured that no reports were overlapping with McCambridge’s review (15).

Reports outside the field of medicine or human behavior related to health and those citing the HE without definition or outcome measurement were excluded. Narrative or systematic reviews with meta-analysis were considered for discussion and to retrieve unnoticed reports from the reference lists, but excluded from this review. Didactic records and letters to the author or editor were also excluded.

Considering the appraisal of the size of the HE, included reports had to be conducted in primary care, in outpatient clinics, or in healthy persons. Only published outcomes were considered and only primary outcomes were computed, without limitation. Included designs were RCTs, *post-hoc* analysis of RCTs, historical comparisons (pre–post comparisons), or observational studies. Studies conducted in hospital wards, in particular HH studies, were excluded.

The use of the term “Hawthorne effect” in health sciences is gradually increasing though its definition remains unclear. It is still more often used without any connection to the original studies in the Hawthorne plant, with a meaning of alteration of behavior related to an experimental background. In other disciplines, its meaning has mutated over time to become still more controversial (15). As our purpose was to investigate the HE in primary care research, we limited our investigations to medical research and our information sources to Medline and to the reference lists of the reviews. We hypothesized that the research in the reference lists of the reviews would provide any material that we would have missed by not exploring other sources. Besides this, PsycINFO and the Web of Science were searched to discuss the results.

The search used PubMed as the main search engine. As McCambridge (15) and Pursell (17) did, we used the “Hawthorne effect” as the only keyword, though it is not a MeSH term (which is “effect modifier”). Filters were set for the availability of an abstract, for language (English, French), and for date range (2012-01-01 to 2022-03-31), as McCambridge’s last included report was published in January 2012. We deliberately chose not to use the keywords “observer effect*,” “participant reactivity,” or merely “reactivity” with another complementary term, in order to be consistent with McCambridge’s approach. The main difference with our search is that beside reports quantifying the HE, we also searched for reports giving a definition of the term. The terms “reactivity,” “placebo effect,” and “regression to the mean” were explored to discuss their interaction with the HE.

Selection process

Initial selections of records were independently undertaken by two reviewers based on the availability of the record, the type of report, the title, and the abstract. All full-text reports meeting the inclusion criteria at this point were read. Reports retrieved from the reference lists of the papers and meeting the inclusion criteria were treated similarly. A consensus meeting of the two reviewers led to the final list of reports included in this review. All reports included were independently fully analyzed by the same two researchers.

Synthesis methods and bias assessment

The same two researchers independently appraised the risk of bias and the level of evidence during the review of the selected full-text reports using the Cochrane tool (19).

Publication bias was assessed by a funnel plot using Review Manager 5.3®.

The narrative results regarding the definition of the HE have been summarized in [Supplementary Table 1](#) with the

description of the study, definition the authors used and a quality appraisal.

All published binary outcome measures of the mean outcome in studies conducted in primary healthcare, outpatient clinics, or healthy persons (e.g., students) have been included in a Microsoft Excel® table. Studies included in the meta-analysis are summarized in **Supplementary Table 2**. Unpublished measures were not sought. Retrieved studies and measures were imported into Cochrane Review Manager 5.3® to compute effect sizes and standard error. The generic inverse variance was used, adjusting for the direction of the HE (i.e., increase or decrease). The odds ratio (OR) and 95% confidence interval (95% CI) were computed using random effects in the context of an important difference in weight of the studies. Heterogeneity was computed using the I^2 statistic. The result is presented as a forest plot. A supplementary sensitivity analysis was computed to differentiate odds ratios and heterogeneity by study design (**Table 1**) and by the level of evidence of the studies (**Table 2**) as the size of the HE appears to be associated with the quality of the research.

Ethics statement and reporting

No ethical statement is required in France for systematic reviews reusing already published data (research method classification MR-004).

The redaction of this review followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement update 2020 (20).

Results

Study selection

Of the 180 records found on Medline, two were excluded because of unavailable abstracts. Forty-four reviews provided

two supplementary records from citation searching. Twenty-nine records were excluded based on title and abstract. Twenty reports were excluded after full reading because they cited the HE without definition or outcome measures. Twice two records reporting on the same study were included as they were complementary reports regarding the outcomes: Buckley (21), Ikpeze (22), Dal-Ré (23), and Pate (24). After the final selection, 74 new English-language reports were included and analyzed for definition and 15 for evaluation of the size of the HE in primary healthcare or outpatient clinics or healthy persons. No report in the French language was found (**Figure 1**).

Study characteristics

Of the 74 selected reports in the definition branch, 15 were randomized controlled trials (RCTs) (25–39), two were not randomized controlled trials (40, 41), three were studies nested in RCTs (42–44), seven were retrospective reanalysis or discussions of RCTs (23, 24, 45–49), three were pilot studies prior to an RCT (50–52), and one was an RCT protocol (53). Further, there were 18 observational studies (54–71), 18 pre–post intervention studies or audits (21, 22, 72–87), one diagnostic accuracy study (88), four qualitative or mixed-method studies (89–92), one mixed-method study protocol (93), and finally one methodology protocol to build up research quality guidelines (13) (**Supplementary Table 1**).

Of the 15 purposely selected reports in primary care, outpatient clinics, or healthy subjects in the meta-analysis branch, the appraisal of the HE was based on a retrospective cohort pre–post intervention analysis in one study (72), in three studies on a *post-hoc* comparison of the RCT population to a non-RCT population (24, 30, 94), in three studies on the comparison of study parameters between enrollment and randomization in an RCT (28, 43, 51), in two studies on the comparison of persons consenting vs. not consenting to participate in a study (45, 67), in two studies on the follow-up of study populations exposed to repeated measurements

TABLE 1 Odds ratio and heterogeneity by study design.

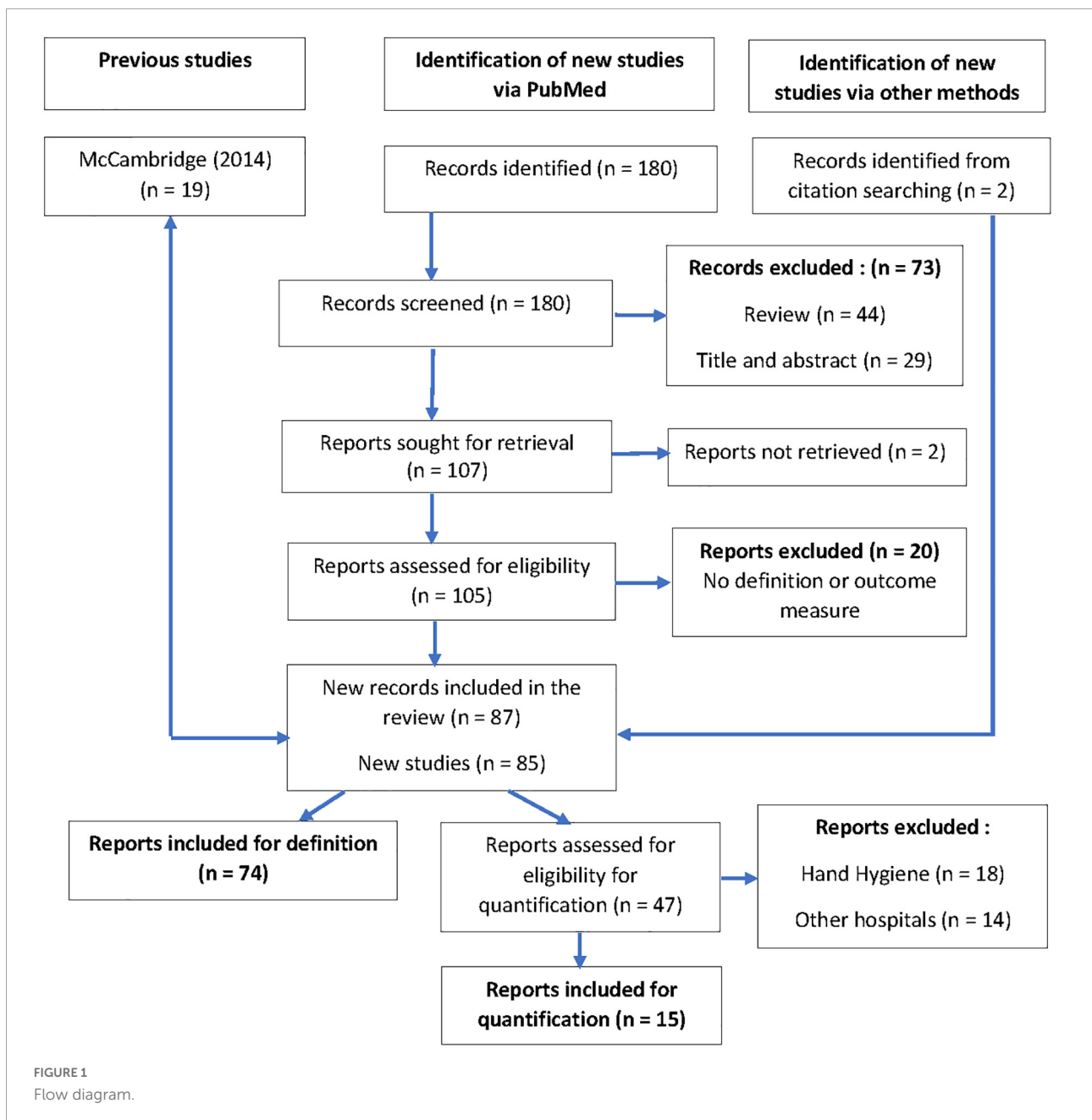
Design	N	OR	95% CI	Chi ²	df	P	I ² (%)
RCTs and pilot RCTs	6	1.08	[0.98; 1.19]	11.58	5	0.04	57
Quasi-experimental and <i>post-hoc</i>	4	1.19	[0.99; 1.44]	0.04	3	1.00	0
Observational	5	1.80	[1.22; 2.66]	126.32	4	<0.0001	97

N, number; OR, odds-ratio; 95% CI, 95% confidence interval; df, degrees of freedom.

TABLE 2 Odds ratio and heterogeneity by level of evidence.

Level of evidence	N	OR	95% CI	Chi ²	df	P	I ² (%)
High/moderate	8	1.04	[0.99; 1.09]	8.02	7	0.33	13
Low	7	1.79	[1.27; 2.50]	128.67	6	<0.0001	95

N, number; OR, odds-ratio; 95% CI, 95% confidence interval; df, degrees of freedom.



(77, 95), and in four studies comparing a population being aware of exposure to observation or assessment to a population who were not aware (33, 64, 83, 96). The main binary outcomes that were inputted in the tables of the review manager to compute an effect size and standard error were sleeping time (28), anti-malarial drug prescriptions (33), time up and go measure (51), self-reported alcohol consumption (96), pain intensity (43), and subjective shared decision-making (95) in the RCTs or RCT feasibility studies. It was an antibiotic selection in a quasi-experimental RCT (30). In *post-hoc* analysis of RCTs, it was the influenza vaccination rate of students

(94), acceptance of a video recording (45), and the rate of COPD acute exacerbations (24). In observational studies, we computed fall rates (72), protocol adherence (77), quality of care (64), school enrollment (67), and spontaneous eye blinks (83) (Supplementary Table 2).

Risk of bias within studies

According to the Cochrane tool (19), in the definition branch, six studies had a low risk of bias (27–29, 31, 32, 39), 18 studies had a moderate risk of bias (24, 26, 30, 33–36,

38, 43–46, 48, 50, 52, 58, 69, 79), 38 had an important risk of bias (21, 22, 37, 40–42, 49, 51, 54, 56, 57, 59–68, 70–72, 74–78, 80–88), and two studies had a very important risk of bias (23, 73). Nine studies were not assessable with the tool (protocols or qualitative/mixed methods studies) (13, 25, 47, 53, 89–93).

In the meta-analysis branch, one study had a low risk of bias (28), seven a moderate risk (24, 30, 33, 43, 45, 94, 96), and seven a high risk (51, 64, 67, 72, 77, 83, 95).

Results of individual studies

The included studies covered all five continents. The populations consisted of patients and various health professionals (students, nurses, physicians...) in different hospital wards or primary care and the community. The most commonly studied outcome was the World Health Organization (WHO) guidance for hand hygiene (HH) [“My Five Moments for Hand Hygiene” initiative (16)] in 13 studies (54, 56, 58, 60, 61, 65, 66, 71, 78, 79, 82, 89). It is noticeable that no study targeting this topic was conducted since the COVID-19 outbreak, except two qualitative ones (89, 92). Other outcomes were very heterogeneous and linked to behavioral factors in health professionals and patients (e.g., completion of medical records, management protocol adherence, quality audits, antibiotic prescription, sleep duration, alcohol consumption) or other aspects (e.g., falls, skin infection, glomerular filtration rate, and glycemia).

Results of syntheses

Definition of the Hawthorne effect in medical studies

Based on this review, our definition of the HE in medical studies is “an aware or unconscious complex behavioral change in a study environment, related to the interaction of four biases affecting the study subjects and investigators: selection bias, commitment and congruence bias, conformity and social desirability bias, and observation and measurement bias.”

A selection bias

The subject agreeing to participate in a study is interested in its outcome, expects a benefit, and trusts the investigator (67, 92). Characteristics of people who consent to participate in clinical trials often differ from patients who decline participation (24, 44). The investigator has a special interest in the field of the study, has more knowledge, and is more skilled in this field than the average health professional (45). As participants' health literacy is essential to the ability to adhere to the study intervention as well as the ability to remember the details

of the recommendations made to participants during visits, investigators will tend to include patients with a higher level of literacy (47).

A commitment and congruence bias

Signing the informed consent, the subject agrees to comply with the artificial experimental life rules and is willing to respect these rules as much as possible, far more than in real life (26). This is especially true for ambulatory active patients (like primary care patients) compared to passive inpatients (66). Signing his (or her) contract with the sponsor, the investigator agrees to follow good clinical practices, feels like part of a project, and has often agreed to undergo complementary training (77). In order to minimize the number of patients lost to follow-up, s/he will be particularly careful to strengthen the follow-up rules with the subject (47, 49, 59, 77).

A conformity and social desirability bias

As described by McCambridge, the “awareness of (...) having behavior assessed engenders [in the subject] beliefs about researcher expectations. Conformity and social desirability considerations then lead behavior to change in line with these expectations (15).” This is also true for the investigator: in case of uncertainty in the answers to an assessment scale, the investigator will tend to quote systematically in order to be in line with the expectations of the study that s/he shares (24, 50, 64).

An observation and measurement bias

The HE is often mitigated to the observation bias, without going more in depth into the concerns of this effect. The awareness of being possibly observed, assessed, and singled out engenders in the subject and in the investigator a special emphasis regarding the three previous biases (47, 58, 87). A direct observation (e.g., HH studies) engenders the largest HE (56) but depends on the authority status of the observer (65). If the observation remains distant, but the subject or the investigator has to complete repeated measurements or questionnaires, his/her interest in the field of the questionnaire will tend to change his/her behavior or beliefs (13, 24, 35, 95). This measurement bias is also described as “measurement reactivity” or “reactivity” (11, 13, 35, 97).

Heterogeneity of the Hawthorne effect

We found important differences across studies or within individual studies regarding the HE. Four main groups of factors seem to determine this heterogeneity: education and literacy or professional position, mental health conditions, environmental factors of the study setting, and the type of outcome measures.

The education or professional position of health professionals

There were important differences between nurses (more prone to HE) and physicians, and in physicians between medics

(more prone to HE) and surgeons (14, 79). In subjects, the level of literacy and deprivation had an important influence with less marked HE in subjects with a lower level of education (66), though the embarrassment caused by the attendance of an observer might be higher in this population (57). Further, as already described, investigators tend to enroll in trial patients with a better health literacy as a means to ensure they understand and remember the recommendations made to participants during visits (47).

Mental health conditions modify the Hawthorne effect

The presence of symptoms such as anxiety and depression contribute to enhanced behavioral changes when people are aware of observation (45, 48, 70).

Environmental factors of the study setting

Regarding HH, the effect was clearly more marked in medicine wards than in surgery or anesthesia wards in hospitals (14, 79, 89). Primary care patients, playing an active role in the patient–doctor relationship, were more prone to the HE than more passive patients in a hospital setting. The HE was less pronounced in deprived dwellings, possibly increasing health inequalities (66).

The main outcome measure

The more the main outcome is linked to psychological or behavioral factors [e.g., sleep agendas (28) and alcohol consumption (38)], even when measured with blinded assessors, the more the effect is notable. The baseline level of the variable interferes also: the larger the deviation from the targeted value is at baseline, the more a HE has to be expected (71). However, as we will discuss below, this point has to be mitigated by a regression toward the mean (26, 43, 46). The direction of the targeted variation of the HE is also important: when the variable is expected to diminish [e.g., antibiotic prescription (52)], the relative reduction is more important than when it is expected to increase [e.g., carpal tunnel release (21, 22)].

Duration of the Hawthorne effect

The onset of the Hawthorne effect in a study environment is very fast (61). In HH studies, it was estimated to take 14 min after the appearance of the observer before health professionals altered their hand-washing behavior, increasing further after 50 min (71). In sleep agendas for sleeping trouble, there was a significant improvement in sleeping duration between the baseline measure and the measure at randomization; insulin resistance and fasting glucose improved simultaneously (28). In chronic kidney disease, there was an improvement in the glomerular filtration rate during the 3-month run-in phase of an RCT, in a disease where this usually worsens over time (50). In neck pain, the intensity of the pain diminished between screening and randomization (43).

The HE disappears totally or partially after the end of the observation or when the subject is released (36, 70, 85). In the

case of long-lasting studies, the HE decreases gradually as the study environment becomes commonplace for the participants (33, 72, 87).

Size of the Hawthorne effect

As explained above, we only considered the appraisals of the effect on binary outcomes made in primary care research, outpatient clinics, and persons in good health (students) for the calculation of the size of the HE. Hand-hygiene studies were ruled out of our research since Pursell et al. published their meta-analysis (17). Our findings could only confirm theirs, and we consider these results as outdated as the COVID-19 outbreak considerably changed HH habits (18).

To compute the size of the HE, we purposely selected fifteen studies with different designs where the HE was appraised by different approaches (see study characteristics and [Supplementary Table 2](#)).

We computed in all studies an OR of 1.41, 95% confidence interval [1.13; 1.75] ([Figure 2](#): forest plot). In sensitivity analysis, we analyzed separately the studies by design ([Table 1](#)) and by the level of evidence ([Table 2](#)). It is notable that in RCTs, and in a quasi-experimental or *post-hoc* analysis of RCTs, the HE appeared to be not significant (95% CI respectively [0.98; 1.19] and [0.99; 1.44]) with a weak heterogeneity (I^2 respectively 57 and 0%). The same observation is valid for studies with a high-to-moderate level of evidence (95% CI: [0.99; 1.09], I^2 : 13%). A significant HE with a high level of heterogeneity appears in observational studies and studies with a low level of evidence (95% CI respectively [1.22; 2.66] and [1.27; 2.50], and I^2 respectively 97 and 95%).

Reporting biases

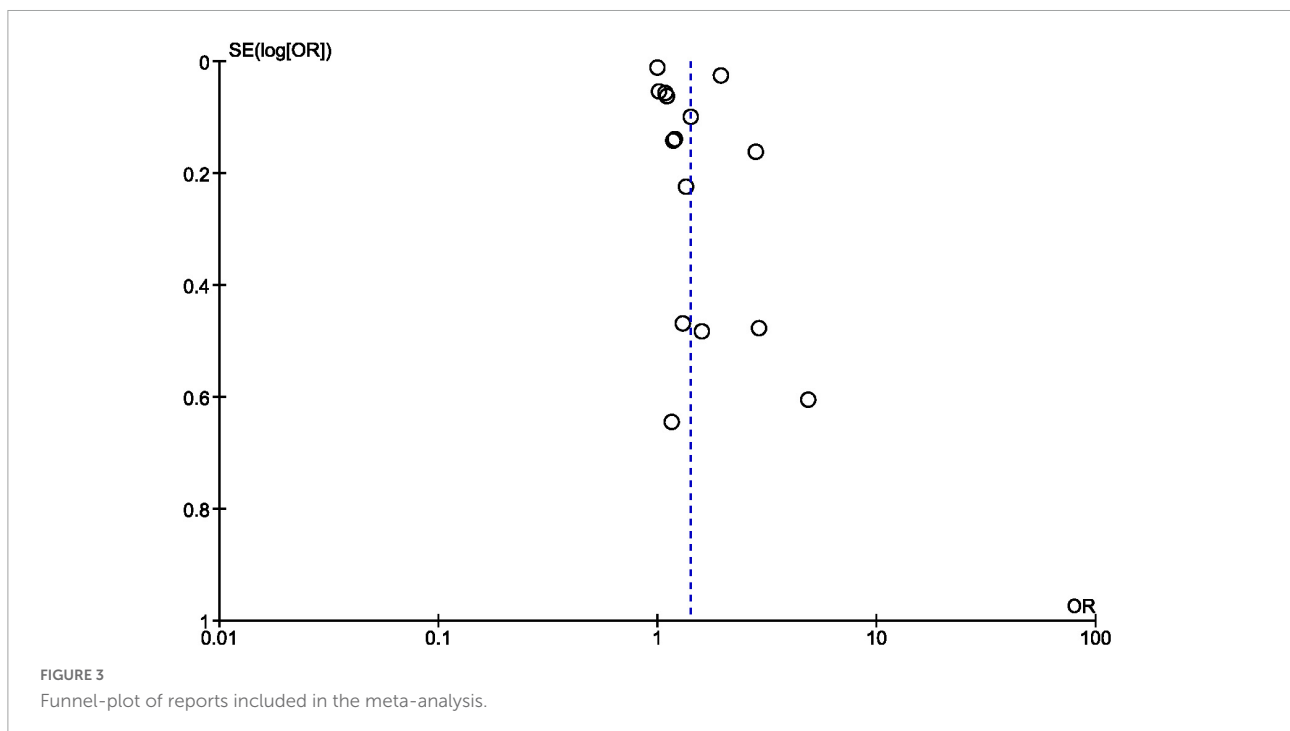
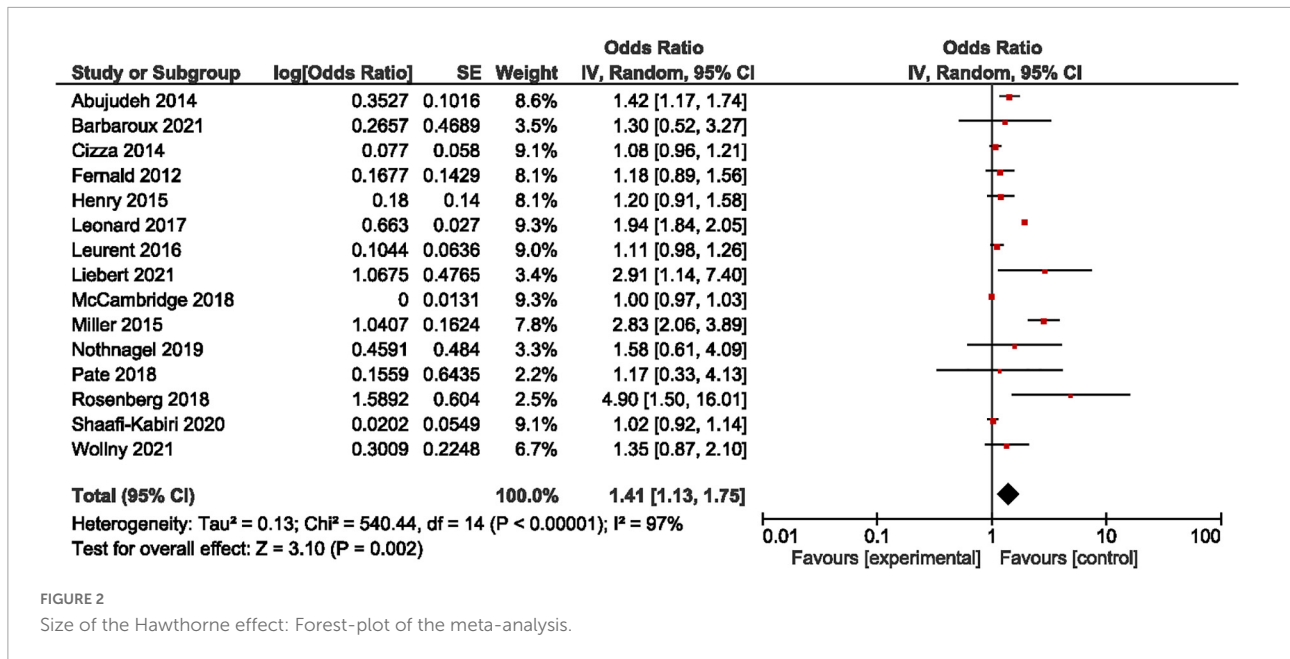
Regarding heterogeneity in the meta-analysis of all the studies, it is notable that the I^2 computing at 97% illustrates that the whole of the variance can be explained by heterogeneity. However, this heterogeneity is to be imputed to observational studies and studies with a poor methodology. Sensitivity analysis found that heterogeneity and the significance of the HE for binary outcomes disappear in well-designed controlled studies.

Regarding the overall publication bias, the chimney plot did not illustrate an exaggerated risk with a well-balanced distribution of the results around the total OR ([Figure 3](#): funnel plot).

Discussion

Summary of evidence

Researchers are still not unanimous regarding the existence of the HE and there is considerable inconsistency concerning



the description and definition of the phenomenon (92). The point is not a denial of an experimental artifact which is unanimously agreed upon. The dissension relates to the description of what happened at the Hawthorne plant (10, 12). Rather than calling this artifact “participant reactivity,” we chose to keep the folkloric name of the Hawthorne effect as it is contemporarily used in health sciences, refining its definition. It is an experimental artifact that reduces the external validity and size effect of studies, with a combined OR for binary outcomes

that can be carefully (due to heterogeneity) estimated at 1.41 (95% CI: [1.13; 1.75]) when considering studies conducted in outpatient clinics and with healthy persons. However, the significance and the heterogeneity of the HE are to be imputed to observational studies and studies with a poor level of evidence, as it disappears in well-designed RCTs or quasi-experimental studies. As a complex system of biases and psychological interferences, all related to a change of behavior in subjects and

investigators, it is more dynamic than the summation of each individual bias.

The size and influence of the HE depend on the population being studied, the educational level and the social position of the investigators and subjects, the mental health status of the investigators and subjects, the studied variable, its initial value and its expected variation, and the duration of the experiment. It is possible to reduce this complex system by analyzing the behavioral beliefs and assessment of the issues of the intervention, the normative beliefs and motivation to comply, and the control beliefs and perceived power as described in the theory of planned behavior or reasoned action (98).

Up until recently, the HE has mainly been linked with observation bias, though the interaction between observation and selection bias has already been described (14, 67). To this point, the use of the term “Hawthorne effect” was of little interest as it was considered to be limited to the fact of observing a subject or an investigator in an experimental environment. The various publications of McCambridge have created a new association with social desirability bias and conformity bias (15, 99, 100). After having completed this review, we acknowledge the reality of what we chose to continue calling the Hawthorne effect, not only as an observation bias or as a summation of biases but also as a complex system that more or less creates an artifact in all research. Describing the HE as selection bias, commitment and congruence bias, conformity and social desirability bias, and observation and measurement bias is enlightening but somewhat simplistic as feedback loops are existing between the research targets, methods, and population explaining the important heterogeneity and temporal instability of the effect (101).

The HE must not be confused with other biases that are not related to bio-psychological, social, or behavioral factors, for example, attrition bias (102) or contamination bias (47). Furthermore, there are important overlap areas between the HE, the regression toward the mean (RTM), and the placebo effect. The RTM is a statistical phenomenon that occurs when repeated measurements are made on the same subject or unit of observation. It happens because values are observed with random error, that is a non-systematic variation in the observed values around a true mean (103). When patients are enrolled into a trial based on a deviating value of the main outcome and randomized a couple of weeks or months later, it can happen at randomization that the deviation of the main outcome is considerably reduced (26, 28, 43, 51). It is then difficult to differentiate the part of the HE and the one of the RTM. Regarding the placebo effect, similar to the HE, its definition is controversial which makes the distinction between the two effects difficult to exemplify. This effect is assumed to be caused by the special type of patient-provider interaction associated with giving and receiving a

treatment, or in other words the treatment ritual (104). This patient-provider interaction can also be described without the prescription of any treatment, for instance, a patient who experiences pain reduction because of an interview with a warm and empathic physician (104). However, in this case the term of placebo effect, related exclusively to the medication, should not be used.

As a consequence, we can assume that all medical research, qualitative or quantitative, is inevitably prone to the HE which limits its external validity, starting with the conscious or unconscious selection of the study population and the investigators, leading to blind spots in medical knowledge.

Strengths and limitations of this study

As an update of McCambridge’s review (15) and a continuation of Pursell et al.’s review (17), we chose to use but one keyword term: “Hawthorne effect.” Hence, we may have missed reports using as keywords the names of biases that are part of the HE (e.g., “observation bias” or “social desirability bias”) or alternative terms of the HE (e.g., “measurement reactivity” or “participant reactivity”). It is probable that our search strategy has been too specific, thus insufficiently sensitive. However, our choice was confirmed during the selection phase by the finding of reports using other terms appointing the same object or pointing to studies using these other terms.

The use of the term “Hawthorne effect” is widely used in medical sciences as we could note through the incrementally growing number of records citing it during the last 10 years in our search. It appeared to be relevant to refine the definition of the term as it is used contemporaneously in medical research in general and in primary care in particular. This is evident in 10 years after McCambridge’s review even though they had already noted a dissociation appearing in the meaning of the term in medical sciences in regard to other disciplines (15). For this reason, we only searched reports related to the medical field and we limited our search to Medline and the reference lists of the review articles that we retrieved. This choice might have been too specific and for this reason, we deepened our search using PsycINFO and the Web of Science in order to enlarge the consideration of the results in the discussion. The search in reference lists and other sources found, with two exceptions of reports that were considered in this review, records deriving from other disciplines, mainly from psychology and education sciences. It was notable that psychologists tended to use the term more in line with what happened at the Hawthorne plant and were more critical regarding its use, while medicals were more prone to use the term meaning an experimental artifact connected to behavioral changes in an experimental context, disregarding its origins. Considering the important number of reports that we analyzed and the definitions that were verified,

the risk of having missed a definition due to a too-specific search seems minimal.

The limitation of our search to reports written in English and French might also have been detrimental. We missed two reports in Chinese about acupuncture, one in Japanese regarding HH, one in Dutch about drug effects, one in Spanish about the behavior of diabetic patients, and one in German about clinical coding. None of these reports gave a clear definition of the HE or could have been included in our meta-analysis. Further, the Dutch report might be confused between the HE and the placebo effect.

Some caution in the interpretation of the meta-analysis is necessary related to the fact that binary results (before–after or overt–covert comparisons) cannot exemplify a complex system. We note that adding “apples and oranges” may cause suspicion, but brought up less heterogeneity than HH studies using the same comparator in different hospital wards. This is related to the fact that the computed data for comparison in the meta-analysis are effect sizes and standard errors.

Considering the literature, this heterogeneity in the analysis of all studies was expected and we could have decided not to publish the computation of the meta-analysis as per Pursell et al. (17). In line with some authors, the sensitivity analysis confirmed the association between poor methods and the rise of a HE (11, 12). When analyzing separately RCTs and quasi-experimental studies, or studies with a good level of evidence, we noted that the presence of a HE in binary outcomes was no more significant with an acceptable heterogeneity. Rather, in observational studies or studies with a low level of evidence, the HE appeared to be significant, though with all of the variances possibly explained by heterogeneity.

Implications of the results for future research

Randomized controlled trials

Randomized controlled trials (RCTs) in parallel groups are prone to the HE, but as groups are equally exposed to the effect, its impact on the main outcome might be reduced (99). This might be an explanation of the minor impact of the HE on binary outcomes. This is particularly true when the RCT is blinded, and if possible double blinded. However, blinded studies are often difficult or impossible to implement for ethical, practical, or financial reasons. Blinding would not prevent the selection of subjects to improve the homogeneity of the included population in order to enhance the chance of demonstrating statistically significant differences and reduce attrition bias or the occurrence of serious adverse events in a linear form of reasoning. Concomitantly, it would not prevent the selection of investigators with deeply rooted beliefs (like the role of cholesterol in leading to

cardiovascular diseases) and a conformism that might be strengthened by complementary education, here again to improve homogeneity in completing the clinical record forms (CRF) (105).

Randomized controlled trials are often cluster randomized in primary care for feasibility reasons. The randomization level is mainly the GP investigator, and the cluster is defined as the group of patients of this GP. As a matter of fact, this emphasizes the influence of the selection of investigators on the results. The introduction of the intra-class correlation coefficient in the calculation (ICC) of the sample size is supposed to erase the effect of this bias on the results of the main outcome, but in most cases this ICC is estimated without certainty, based on the literature. Knowing the heterogeneity of the HE, the feasibility of computing exactly this ICC seems inaccessible.

The main risk, when the HE is not correctly mastered in an RCT, occurs when the effect size of the main outcome is small. If the size of the HE turns out to be important, it might overwhelm the results of the main outcome and lead to a negative trial (47). This is an important fact to consider when designing future RCTs in primary care or analyzing the events that led to a negative trial.

As noted, patients change their behavior by the start of the trial, and baseline values are prone to the RTM (24, 28, 43, 51). For these reasons, it can be recommended to separate enrollment in trials and randomization by about 1 month and to repeat outcome measures at the randomization visit. The analyzed baseline measures will be those at randomization, already modified by experimental artifacts, before the implementation of the intervention.

Implementing an RCT in primary care also means a profound disruption in the patient–doctor relationship. The latter changed during the past decades from a paternalistic model to a more balanced model of mutual participation (106). This relation can also be described by the family physician's ongoing commitment to the patient and his/her family as persons (107). The physician will carefully choose among his/her patients, based on this mutual understanding, which patients s/he feels comfortable proposing participation in a trial to. This means that the physician who signed the study contract and the patient who signed the informed consent will both lose their freedom to share decision-making regarding a particular condition of the patient even in trials that try to avoid this barrier (108). In the PaCUDAHG-Gé trial (109), general practitioners had to propose to their insufficiently or unscreened for cervical cancer female patients home vaginal self-sampling or usual physician-sampled cervical smears. Patients included in the study could accept or refuse screening. The interest to include in the study all their eligible patients, whatever their decision, was repeated several times to the investigators by the study team. However, of the 300 included patients, 299 were screened (96 smears and 203 self-sampling) with

only one who refused screening. It is also of note that no never-screened female patient was included. As cervical cancer screening is strongly associated with the level of health literacy, the preference of investigators to include patients with a higher level of literacy contributed to the exclusion of never-screened women (47).

Based on the findings of this review, we assessed whether the RCT we implemented regarding the impact of posters and pamphlets in GPs' waiting rooms had been biased by a HE (2). The design of our study was a cluster-randomized trial, where GP investigators had no CRFs to complete as data were collected from a health insurance claim database. The GP investigators were not affected by the main outcome as it was the delivery of seasonal influenza vaccines in community pharmacies to patients targeted by this vaccination. The intervention was a reshuffle of the wall decoration of their waiting room, pre-existing posters and advertisements being taken away and replaced by one single poster promoting seasonal influenza vaccination, and the available reading material was removed and replaced by pamphlets of the same campaign. GP investigators gave their consent for this transformation without participating in it. GPs from the control group had their waiting room unchanged and had only to give their consent to access their data in the claim database. In this design, the only involvement of the GP investigators that might have biased the study was to give their consent to a study, where the vaccination coverage of their patients was assessed. This means (1) that they believed that seasonal influenza vaccination was important in their patients targeted for this vaccination and (2) that they were confident in doing their best to reach this objective. This means a selection bias of the GP investigators, but no observation bias (the observation of their outcomes being totally remote), no special commitment or congruence bias (their only commitment was signing the consent and accepting the reshuffle by others of their waiting rooms), and no special conformity or social desirability bias unless the one intertwined with the selection bias. It is thus that we believe that the HE in our study was minimal.

Observational studies

The HE probably has more consequences for the outcome of observational studies than RCTs, as it directly influences the results, without the balance of a control group. This statement matches the findings regarding observational studies in our meta-analysis.

The selection of the investigators in primary care will be influenced by the interest of the investigator in the topic and the prevalence of the studied condition among his/her patients. If patients are in general comparable, the way they are managed and educated by their physician might deeply differ due to a different level of commitment (i.e., for patients with addiction mainly managed by a small proportion of highly invested primary care physicians) (110). For similar reasons, the

specialty of the physician can also lead to the selection of more complicated patients (e.g., diabetic patients or hypertensive patients managed by diabetologists or cardiologists are probably more difficult to balance and need heavier interventions than those managed by GPs though there is a lack of literature describing the difference in the burden of disease).

Observational studies will also ignore all the persons who are affected by a condition but are not aware of it or are not willing to address the condition. Similarly, it will ignore people who are not participating in diverse screenings. This highlights the problem of blind spots in primary care research.

Compared to usual care, conformity and social desirability will probably change the managing behavior of the investigator, the level of adherence and compliance of the patient, and the data collected in the CRF. Retrospective data will be altered also by conformity as well as by memory failure, with a trend to embellish vague recollections.

Qualitative research

Qualitative research collecting data rooted in semi-structured individual or group interviews will probably be biased by the HE when the interviewee is a patient or a doctor and the interviewer is a doctor him/herself. The relationship between a patient and a doctor or between two doctors will tend to increase social desirability bias and conformity bias because the interviewee is willing to meet the interviewer's supposed expectations. This deviance might be even more underlined by the signing of a consent form and the recording of the interview that accentuates the need to provide an interest (111). As a criterion of reflexivity, a qualitative researcher is recommended to describe researcher characteristics that may have influenced the research, so including this HE (112).

Along the same lines, people who have a poor level of literacy or education will be more prone to refuse the interview as they are frightened they will not be able to reach the expected level of interest in the interviewer's supposed expectations. Persons who feel guilty about breaking the rules in light of the norms of their social group (e.g., screening secretly for cervical cancer) will refuse the interview due to shame or fear of being discovered, or may not be willing to go further into transgression. In both cases, essential information will be lost to evidence.

Conclusion

The Hawthorne effect results from a complex system of interacting psychological and social phenomena and appears in all experimental research thereby diminishing external validity. It combines the mobilization of feedback loops at different levels and time, encompassing social selection, individual motivation, commitment and congruence, social conformity

and desirability, and the awareness of being observed, several times assessed, and singled out. There are overlapping areas with the regression toward the mean and the placebo effect. Observational studies or studies with a poor level of evidence are more prone to a HE.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

CB designed the study, searched, selected, and analyzed the published reports, conducted the meta-analysis, and wrote the manuscript. OB searched, selected, and analyzed the reports and wrote the first draft. JF designed the study and tracked the conformance of data management. MC designed the study and amended the first draft. CC copyedited and revised critically the reports for important intellectual content. LP and PV supervised the design of the study and revised critically the manuscript for important intellectual content. All authors gave their final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

References

- Berkhout C, Zgorska-Meynard-Moussa S, Willefert-Bouche A, Favre J, Peremans L, Van Royen P. Audiovisual aids in primary healthcare settings' waiting rooms. A systematic review. *Eur J Gen Pract.* (2018) 24:202–10. doi: 10.1080/13814788.2018.1491964
- Berkhout C, Willefert-Bouche A, Chazard E, Zgorska-Maynard-Moussa S, Favre J, Peremans L, et al. Randomized controlled trial on promoting influenza vaccination in general practice waiting rooms. *PLoS One.* (2018) 13:e0192155. doi: 10.1371/journal.pone.0192155
- Greenwood N. *Understanding the Hawthorne Effect.* (2022). Available online at: https://core.ac.uk/reader/74393583?utm_source=linkout (accessed April 18, 2022).
- Kelly J. *Cicero [Internet].* (Vol. 46.). American Heritage. (1995). Available online at: <https://www.americanheritage.com/cicero#2> (assessed October 29, 2022).
- Roethlisberger FJ, Dickson WJ. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago. Reprinted.* London: Routledge (2003). 615 p. (The early sociology of management and organizations).
- Kompier MA. The "Hawthorne effect" is a myth, but what keeps the story going? *Scand J Work Environ Health.* (2006) 32:402–12. doi: 10.5271/sjweh.1036
- Gillespie R. *Manufacturing Knowledge: A History of the Hawthorne Experiments.* Cambridge: Cambridge University Press (1991). 282 p. (Studies in economic history and policy).
- Festinger L, Katz D. *Research Methods in the Behavioral Sciences.* Hinsdale, IL: Dryden Press (1953). 684 p.
- Parsons HM. What happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. *Science.* (1974) 183:922–32. doi: 10.1126/science.183.4128.922
- Levitt SD, List JA. Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *Am Econ J Appl Econ.* (2011) 3:224–38.
- French DP, Sutton S. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? *Br J Health Psychol.* (2010) 15:453–68. doi: 10.1348/135910710X492341
- Paradis E, Sutkin G. Beyond a good story: from Hawthorne Effect to reactivity in health professions education research. *Med Educ.* (2017) 51:31–9. doi: 10.1111/medu.13122
- Miles LM, Elbourne D, Farmer A, Gulliford M, Locock L, McCambridge J, et al. Bias due to measurement reactions in trials to improve health (MERIT): protocol for research to develop MRC guidance. *Trials.* (2018) 19:653.
- Wu KS, Lee SSJ, Chen JK, Chen YS, Tsai HC, Chen YJ, et al. Identifying heterogeneity in the Hawthorne effect on hand hygiene observation: a cohort study of overtly and covertly observed results. *BMC Infect Dis.* (2018) 18:369. doi: 10.1186/s12879-018-3292-5
- McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol.* (2014) 67:267–77. doi: 10.1016/j.jclinepi.2013.08.015
- World Health Organization [WHO]. *WHO Guidelines on Hand Hygiene in Health Care: A Summary [Internet].* Geneva: WHO (2009).

Acknowledgments

The authors thank Luc Dauchet for his guidance to complete the meta-analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.1033486/full#supplementary-material>

17. Purssell E, Drey N, Chudleigh J, Creedon S, Gould DJ. The Hawthorne effect on adherence to hand hygiene in patient care. *J Hosp Infect.* (2020) 106:311–7.
18. Moore LD, Robbins G, Quinn J, Arbogast JW. The impact of COVID-19 pandemic on hand hygiene performance in hospitals. *Am J Infect Control.* (2021) 49:30–3.
19. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane handbook for systematic reviews of interventions, version 6.3 [Internet].* (2022). Available online at: <https://training.cochrane.org/handbook> (accessed October 29, 2022).
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* (2021) 372:n71.
21. Buckley T, Mitten D, Elfar J. The effect of informed consent on results of a standard upper extremity intake questionnaire. *J Hand Surg.* (2013) 38:366–71.
22. Ikpeze TC, Childs S, Buckley T, Elfar JC. Validity of QuickDASH at day of surgery versus day of initial consultation: does informed consent make a difference? *J Orthop Surg Hong Kong.* (2018) 26:2309499018777897. doi: 10.1177/2309499018777897
23. Dal-Ré R. Could phase 3 medicine trials be tagged as pragmatic? A case study: the salford COPD trial. *J Eval Clin Pract.* (2018) 24:258–61. doi: 10.1111/jep.12796
24. Pate A, Barrowman M, Webb D, Pimenta JM, Davis KJ, Williams R, et al. Study investigating the generalisability of a COPD trial based in primary care (Salford Lung Study) and the presence of a Hawthorne effect. *BMJ Open Respir Res.* (2018) 5:e000339.
25. Arnold SH, Jensen JN, Kousgaard MB, Siersma V, Bjerrum L, Holm A. Reducing antibiotic prescriptions for urinary tract infection in nursing homes using a complex tailored intervention targeting nursing home staff: protocol for a cluster randomized controlled trial. *JMIR Res Protoc.* (2020) 9:e17710. doi: 10.2196/17710
26. Blondeau P, Hamid M, Ghalie Z. Prospective randomized clinical trial on the effects of latanoprost, travoprost and bimatoprost on latanoprost non-responders. *J Fr Ophthalmol.* (2019) 42:894–9. doi: 10.1016/j.jfo.2019.02.009
27. Briët OJT, Yuchik JO, Pfeiffer C, Miller W, Jaeger MS, Khanna N, et al. The effect of small solar powered ‘Bokoo’ net fans on mosquito net use: results from a randomized controlled cross-over trial in southern Ghana. *Malar J.* (2017) 16:12. doi: 10.1186/s12936-016-1654-2
28. Cizza G, Piaggi P, Rother KI, Csako G, Sleep Extension Study Group. Hawthorne effect with transient behavioral and biochemical changes in a randomized controlled sleep extension trial of chronically short-sleeping obese adults: implications for the design and interpretation of clinical studies. *PLoS One.* (2014) 9:e104176. doi: 10.1371/journal.pone.0104176
29. Edwards KE, Hagen SM, Hannam J, Kruger C, Yu R, Merry AF. A randomized comparison between records made with an anesthesia information management system and by hand, and evaluation of the Hawthorne effect. *Can J Anaesth.* (2013) 60:990–7. doi: 10.1007/s12630-013-0003-y
30. Fernald DH, Coombs L, DeAlaume L, West D, Parnes B. An assessment of the Hawthorne Effect in practice-based research. *J Am Board Fam Med.* (2012) 25:83–6.
31. Garrouste-Orgeas M, Soufir L, Tabah A, Schwebel C, Vesin A, Adrie C, et al. A multifaceted program for improving quality of care in intensive care units: IATROREF Study. *Crit Care Med.* (2012) 40:468–76. doi: 10.1097/CCM.0b013e318232d94d
32. Humalda JK, Klaassen G, de Vries H, Meuleman Y, Verschuur LC, Straathof EJM, et al. A Self-management approach for dietary sodium restriction in patients with CKD: a randomized controlled trial. *Am J Kidney Dis.* (2020) 75:847–56. doi: 10.1053/j.ajkd.2019.10.012
33. Leurent B, Reyburn H, Muro F, Mbakilwa H, Schellenberg D. Monitoring patient care through health facility exit interviews: an assessment of the Hawthorne effect in a trial of adherence to malaria treatment guidelines in Tanzania. *BMC Infect Dis.* (2016) 16:59. doi: 10.1186/s12879-016-1362-0
34. McDermott L, Wright AJ, Cornelius V, Burgess C, Forster AS, Ashworth M, et al. Enhanced invitation methods and uptake of health checks in primary care: randomised controlled trial and cohort study using electronic health records. *Health Technol Assess.* (2016) 20:1–92. doi: 10.3310/hta20840
35. Nair H, Williams LJ, Marsh A, Lele P, Bhattacharjee T, Chavan U, et al. Assessing the reactivity to mobile phones and repeated surveys on reported care-seeking for common childhood illnesses in rural India. *J Glob Health.* (2018) 8:020807. doi: 10.7189/jogh.08.020807
36. Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson’s disease and healthy subjects: is there a Hawthorne effect? *J Appl Biomech.* (2015) 31:189–94. doi: 10.1123/jab.2013-0319
37. Smith JE, Rockett M, S SC, Squire R, Hayward C, Ewings P, et al. PAIN SoluTions In the Emergency Setting (PASTIES)–patient controlled analgesia versus routine care in emergency department patients with pain from traumatic injuries: randomised trial. *BMJ.* (2015) 350:h2988.
38. Smith JL, Dash NJ, Johnstone SJ, Houben K, Field M. Current forms of inhibitory training produce no greater reduction in drinking than simple assessment: a Preliminary Study. *Drug Alcohol Depend.* (2017) 173:47–58. doi: 10.1016/j.drugalcdep.2016.12.018
39. Wolff CM, Nowacki AS, Yeh JY, Hickner JM. A randomized controlled trial of two interventions to improve medication reconciliation. *J Am Board Fam Med.* (2014) 27:347–55.
40. Bhimani R. Prevention of work-related musculoskeletal injuries in rehabilitation nursing. *Rehabil Nurs.* (2016) 41:326–35.
41. Guerrero DM, Carling PC, Jury LA, Ponnada S, Nerandzic MM, Donskey CJ. Beyond the Hawthorne effect: reduction of Clostridium difficile environmental contamination through active intervention to improve cleaning practices. *Infect Control Hosp Epidemiol.* (2013) 34:524–6. doi: 10.1086/670213
42. Ardestani MM, Hornby TG. Effect of investigator observation on gait parameters in individuals with stroke. *J Biomech.* (2020) 100:109602.
43. Nothnagel H, Brown Menard M, Kvarstein G, Norheim A, Weiss T, Puta C, et al. Recruitment and inclusion procedures as “pain killers” in clinical trials? *J Pain Res.* (2019) 12:2027–37.
44. van Wyk L, Boers KE, Gordijn SJ, Ganzevoort W, Bremer HA, Kwee A, et al. Perinatal death in a term fetal growth restriction randomized controlled trial: the paradox of prior risk and consent. *Am J Obstet Gynecol MFM.* (2020) 2:100239. doi: 10.1016/j.ajogmf.2020.100239
45. Henry SG, Jerant A, Iosif AM, Feldman MD, Cipri C, Kravitz RL. Analysis of threats to research validity introduced by audio recording clinic visits: selection bias, Hawthorne effect, both, or neither? *Patient Educ Couns.* (2015) 98:849–56. doi: 10.1016/j.pcc.2015.03.006
46. Morberg BM, Malling AS, Jensen BR, Gredal O, Wermuth L, Bech P. The Hawthorne effect as a pre-placebo expectation in Parkinson disease patients participating in a randomized Placebo-Controlled Clinical Study. *Nord J Psychiatry.* (2018) 72:442–6. doi: 10.1080/08039488.2018.1468480
47. Petersen B, Vesper I, Pachwald B, Dagenbach N, Buck S, Waldenmaier D, et al. Diabetes management intervention studies: lessons learned from two studies. *Trials.* (2021) 22:61.
48. Wainberg ML, Mann CG, Norcini-Pala A, McKinnon K, Pinto D, Pinho V, et al. Challenges and opportunities in the science of research to practice: lessons learned from a randomized controlled trial of a sexual risk-reduction intervention for psychiatric patients in a public mental health system. *Braz J Psychiatry.* (2020) 42:349–59. doi: 10.1590/1516-4446-2019-0737
49. Wong G, Lam E, Chow E, Zhang L, Li CN, Mawdsley G, et al. Do patients enrolled in observational studies have better outcomes than non-participants? A retrospective analysis. *Support Care Cancer.* (2020) 28:5751–61. doi: 10.1007/s00520-020-05417-w
50. Fassett RG, Geraghty DP, Coombes JS. The impact of pre-intervention rate of kidney function change on the assessment of CKD progression. *J Nephrol.* (2014) 27:515–9. doi: 10.1007/s40620-014-0058-z
51. Liebert A, Bicknell B, Laakso EL, Heller G, Jalilabaei P, Tilley S, et al. Improvements in clinical signs of Parkinson’s disease using photobiomodulation: a prospective proof-of-concept study. *BMC Neurol.* (2021) 21:256. doi: 10.1186/s12883-021-02248-y.
52. Persell SD, Doctor JN, Friedberg MW, Meeker D, Friesema E, Cooper A, et al. Behavioral interventions to reduce inappropriate antibiotic prescribing: a randomized pilot trial. *BMC Infect Dis.* (2016) 16:373. doi: 10.1186/s12879-016-1715-8.
53. Janssen M, Heerkens Y, Van der Heijden B, Korzilius H, Peters P, Engels J. A study protocol for a cluster randomised controlled trial on mindfulness-based stress reduction: studying effects of mindfulness-based stress reduction and an additional organisational health intervention on mental health and work-related perceptions of teachers in Dutch secondary vocational schools. *Trials.* (2020) 21:376. doi: 10.1186/s13063-020-4189-3
54. Wu KS, Chen YS, Lin HS, Hsieh EL, Chen JK, Tsai HC, et al. A nationwide covert observation study using a novel method for hand hygiene compliance in health care. *Am J Infect Control.* (2017) 45:240–4. doi: 10.1016/j.ajic.2016.10.010
55. Di Bona D, Minenna E, Albanesi M, Nettis E, Caiaffa MF, Macchia L. Benralizumab improves patient reported outcomes and functional parameters in difficult-to-treat patients with severe asthma: data from a real-life cohort. *Pulm Pharmacol Ther.* (2020) 64:101974. doi: 10.1016/j.pupt.2020.101974
56. El-Saed A, Noushad S, Tannous E, Abdirizak F, Arabi Y, Al Azzam S, et al. Quantifying the Hawthorne effect using overt and covert observation of hand

- hygiene at a tertiary care hospital in Saudi Arabia. *Am J Infect Control.* (2018) 46:930–5.
57. Goodwin MA, Stange KC, Zyzanski SJ, Crabtree BF, Borawski EA, Flocke SA. The Hawthorne effect in direct observation research with physicians and patients. *J Eval Clin Pract.* (2017) 23:1322–8.
58. Hagel S, Reischke J, Kesselmeier M, Winning J, Gastmeier P, Brunkhorst FM, et al. Quantifying the Hawthorne effect in hand hygiene compliance through comparing direct observation with automated hand hygiene monitoring. *Infect Control Hosp Epidemiol.* (2015) 36:957–62. doi: 10.1017/ice.2015.93
59. Hameed W, Ishaque M, Gul X, Siddiqui JUR, Hussain S, Hussain W, et al. Does courtesy bias affect how clients report on objective and subjective measures of family planning service quality? A comparison between facility- and home-based interviews. *Open Access J Contracept.* (2017) 9:33–43. doi: 10.2147/OAJC.S153443
60. Kovacs-Litman A, Wong K, Shojania KG, Callery S, Vearncombe M, Leis JA. Do physicians clean their hands? Insights from a covert observational study. *J Hosp Med.* (2016) 11:862–4. doi: 10.1002/jhm.2632
61. Kurtz SL. Measuring and accounting for the Hawthorne effect during a direct overt observational study of intensive care unit nurses. *Am J Infect Control.* (2017) 45:995–1000. doi: 10.1016/j.ajic.2017.03.022
62. Lakomek F, Lukas RP, Brinkrolf P, Mennewisch A, Steinsiek N, Gutendorf P, et al. Real-time feedback improves chest compression quality in out-of-hospital cardiac arrest: a Prospective Cohort Study. *PLoS One.* (2020) 15:e0229431. doi: 10.1371/journal.pone.0229431
63. Malchow C, Fiedler G. Effect of observation on lower limb prosthesis gait biomechanics: preliminary results. *Prosthet Orthot Int.* (2016) 40:739–43.
64. Miller NP, Amouzou A, Hazel E, Degefe T, Legesse H, Tafesse M, et al. Assessing the quality of sick child care provided by community health workers. *PLoS One.* (2015) 10:e0142010. doi: 10.1371/journal.pone.0142010
65. Pan SC, Tien KL, Hung IC, Lin YJ, Sheng WH, Wang MJ, et al. Compliance of health care workers with hand hygiene practices: independent advantages of overt and covert observers. *PLoS One.* (2013) 8:e53746. doi: 10.1371/journal.pone.0053746
66. Quick A, Böhnke JR, Wright J, Pickett KE. Does involvement in a Cohort Study improve health and affect health inequalities? A natural experiment. *BMC Health Serv Res.* (2017) 17:79. doi: 10.1186/s12913-017-2016-7
67. Rosenberg M, Pettifor A, Twine R, Hughes JP, Gomez-Olive FX, Wagner RG, et al. Evidence for sample selection effect and Hawthorne effect in behavioural HIV prevention trial among young women in a rural South African community. *BMJ Open.* (2018) 8:e019167. doi: 10.1136/bmjopen-2017-019167
68. Srigley JA, Furness CD, Baker GR, Gardam M. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ Qual Saf.* (2014) 23:974–80. doi: 10.1136/bmjqs-2014-003808
69. Steward WT, Koester KA, Guzé MA, Kirby VB, Fuller SM, Moran ME, et al. Practice transformations to optimize the delivery of HIV primary care in community healthcare settings in the United States: a program implementation study. *PLoS Med.* (2020) 17:e1003079. doi: 10.1371/journal.pmed.1003079
70. Vickers J, Reed A, Decker R, Conrad BP, Olegario-Nebel M, Vincent HK. Effect of investigator observation on gait parameters in individuals with and without chronic low back pain. *Gait Posture.* (2017) 53:35–40. doi: 10.1016/j.gaitpost.2017.01.002
71. Yin J, Reisinger HS, Vander Weg M, Schweizer ML, Jesson A, Morgan DJ, et al. Establishing evidence-based criteria for directly observed hand hygiene compliance monitoring programs: a prospective, multicenter cohort study. *Infect Control Hosp Epidemiol.* (2014) 35:1163–8. doi: 10.1086/677629
72. Abujudeh HH, Aran S, Besheli LD, Miguel K, Halpern E, Thrall JH. Outpatient falls prevention program outcome: an increase, a plateau, and a decrease in incident reports. *Am J Roentgenol.* (2014) 203:620–6. doi: 10.2214/AJR.13.11982
73. Afsarlar CE, Ryan SL, Donel E, Baccam TH, Jones B, Chandwani B, et al. Standardized process to improve patient flow from the emergency room to the operating room for pediatric patients with testicular torsion. *J Pediatr Urol.* (2016) 12:233.e1–4.
74. Chandok N, Speechley M, Ainsworth PJ, Chakrabarti S, Adams PC. The impact of population-based screening studies on hemochromatosis screening practices. *Dig Dis Sci.* (2012) 57:1420–2.
75. Kennedy MT, Ong JCY, Mitra A, Harty JA, Reidy D, Dolan M. 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”. *The Surgeon.* (2013) 11:10–3. doi: 10.1016/j.surge.2011.10.002
76. Laborie S, Abadie G, Denis A, Touzet S, Fischer Fumeaux CJA. Positive impact of an observational study on breastfeeding rates in two neonatal intensive care units. *Nutrients.* (2022) 14:1145. doi: 10.3390/nu14061145
77. Leonard KL, Masatu MC. Changing health care provider performance through measurement. *Soc Sci Med.* (2017) 181:54–65.
78. McDonald EG, Smyth E, Smyth L, Lee TC. Hand hygiene ‘hall monitors’: leveraging the Hawthorne effect. *Am J Infect Control.* (2018) 46:706–7. doi: 10.1016/j.ajic.2017.11.030
79. McLaws ML, Kwok YLA. Hand hygiene compliance rates: fact or fiction? *Am J Infect Control.* (2018) 46:876–80.
80. Rampersad SE, Martin LD, Geiduschek JM, Weiss GK, Bates SW, Martin LD. Video observation of anesthesia practice: a useful and reliable tool for quality improvement initiatives. *Pediatr Anesth.* (2013) 23:627–33. doi: 10.1111/pan.12198
81. Rezk F, Åstrand H, Acosta S. Antibiotic prophylaxis with trimethoprim/sulfamethoxazole instead of cloxacillin/cefotaxime increases inguinal surgical site infection rate after lower extremity revascularization. *Int J Low Extrem Wounds.* (2019) 18:135–42. doi: 10.1177/1534734619838749
82. Sánchez-Carrillo LA, Rodríguez-López JM, Galarza-Delgado DÁ, Baena-Trejo L, Padilla-Orozco M, Mendoza-Flores L, et al. Enhancement of hand hygiene compliance among health care workers from a hemodialysis unit using video-monitoring feedback. *Am J Infect Control.* (2016) 44:868–72. doi: 10.1016/j.ajic.2016.01.040
83. Shaafi Kabiri N, Brooks C, Comery T, Kelley ME, Fried P, Bhangu J, et al. The Hawthorne effect in eye-blinking: awareness that one’s blinks are being counted alters blink behavior. *Curr Eye Res.* (2020) 45:1380–4. doi: 10.1080/02713683.2020.1752736
84. Spector JM, Agrawal P, Kodkany B, Lipsitz S, Lashoer A, Dziekan G, et al. Improving quality of care for maternal and newborn health: prospective Pilot Study of the WHO safe childbirth checklist program. *PLoS One.* (2012) 7:e35151. doi: 10.1371/journal.pone.0035151
85. Wander P, Fahrbruch C, Rea T. The dispatcher assisted resuscitation trial: indirect benefits of emergency research. *Resuscitation.* (2014) 85:1594–8.
86. White E, Proudlove N, Kallon D. Improving turnaround times for HLA-B*27 and HLA-B*57:01 gene testing: a Barts Health NHS Trust quality improvement project. *BMJ Open Qual.* (2021) 10:e001538. doi: 10.1136/bmjopen-2021-001538
87. Zhang-Rutledge K, Clark SL, Denning S, Timmins A, Dildy GA, Gandhi M. An initiative to reduce the episiotomy rate: association of feedback and the Hawthorne effect with leapfrog goals. *Obstet Gynecol.* (2017) 130:146–50. doi: 10.1097/AOG.0000000000002060
88. Barron CVM, Heenan HF, Thompson H, Chan H, Ngu J, Lunt H. Detecting dysglycaemia in compensated liver cirrhosis: comparison of oral glucose tolerance test and glycated haemoglobin, with continuous glucose monitoring. *Diabet Med.* (2022) 39:e14778. doi: 10.1111/dme.14778
89. McKay KJ, Li C, Sotomayor-Castillo DC, Ferguson PE, Wyer DM, Shaban PRZ. Healthcare workers’ experiences of video-based monitoring of hand hygiene behaviours: a Qualitative Study. *Am J Infect Control.* (2022):S0196655322001511. [Epub ahead of print]. doi: 10.1016/j.ajic.2022.03.010
90. Petri LA, Thottathil P, Shih G, Henderson A, Pasquariello C, Black SA. Ask the question, be the solution: fostering well-being through contextualized assessment and strategy development. *Pediatr Anesth.* (2021) 31:68–73. doi: 10.1111/pan.14087
91. Rea J, Stephenson C, Leasure E, Vaa B, Halvorsen A, Huber J, et al. Perceptions of scheduled vs. unscheduled directly observed visits in an internal medicine residency outpatient clinic. *BMC Med Educ.* (2020) 20:64. doi: 10.1186/s12909-020-1968-1
92. Rezk F, Stenmarker M, Acosta S, Johansson K, Bengnér M, Åstrand H, et al. Healthcare professionals’ experiences of being observed regarding hygiene routines: the Hawthorne effect in vascular surgery. *BMC Infect Dis.* (2021) 21:420. doi: 10.1186/s12879-021-06097-5
93. Płaszewski M, Krzepakowska W, Grantham W, Wroński Z, Makaruk H, Trębska J. Knowledge, behaviours and attitudes towards Evidence-Based Practice amongst physiotherapists in Poland. A nationwide cross-sectional survey and focus group study protocol. *PLoS One.* (2022) 17:e0264531. doi: 10.1371/journal.pone.0264531
94. Barbaroux A, Benoit L, Raymondie RA, Milhabet I. Nudging health care workers towards a flu shot: reminders are accepted but not necessarily effective. A randomized controlled study among residents in general practice in France. *Fam Pract.* (2021) 38:410–5. doi: 10.1093/fampra/cmab001
95. Wollny A, Löffler C, Drewelow E, Altiner A, Helbig C, Daubmann A, et al. Shared decision making and patient-centeredness for patients with poorly

- controlled type 2 diabetes mellitus in primary care—results of the cluster-randomised controlled DEBATE trial. *BMC Fam Pract.* (2021) 22:93. doi: 10.1186/s12875-021-01436-6
96. McCambridge J, Wilson A, Attia J, Weaver N, Kypri K. Randomized trial seeking to induce the Hawthorne effect found no evidence for any effect on self-reported alcohol consumption online. *J Clin Epidemiol.* (2019) 108:102–9. doi: 10.1016/j.jclinepi.2018.11.016
97. Cledes SA, Deans NK. Presence and duration of reactivity to pedometers in adults. *Med Sci Sports Exerc.* (2012) 44:1097–101.
98. Albarracín D, Johnson BT, Fishbein M, Muellierelle PA. Theories of reasoned action and planned behavior as models of condom use: a meta-analysis. *Psychol Bull.* (2001) 127:142–61. doi: 10.1037/0033-2909.127.1.142
99. 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.* (2014) 67:247–53. doi: 10.1016/j.jclinepi.2013.09.004
100. McCambridge J. From question-behaviour effects in trials to the social psychology of research participation. *Psychol Health.* (2015) 30:72–84. doi: 10.1080/08870446.2014.953527
101. García JM. *Theory and Practical Exercises of System Dynamics: Modeling and Simulation with Vensim PLE. Preface John Sterman. Independently Published.* Chicago, IL: Juan Martin Garcia (2020).
102. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ Evid-Based Med.* (2018) 23:21–2.
103. Barnett AG. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol.* (2004) 34:215–20.
104. Hróbjartsson A. What are the main methodological problems in the estimation of placebo effects? *J Clin Epidemiol.* (2002) 55:430–5.
105. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. *Am J Commun Psychol.* (2009) 43:267–76.
106. Kaba R, Sooriakumaran P. The evolution of the doctor-patient relationship. *Int J Surg Lond Engl.* (2007) 5:57–65.
107. McWhinney IR. Continuity of care in family practice. Part 2: implications of continuity. *J Fam Pract.* (1975) 2:373–4.
108. Légaré F, O'Connor AM, Graham ID, Wells GA, Tremblay S. Impact of the Ottawa decision support framework on the agreement and the difference between patients' and physicians' decisional conflict. *Med Decis Making.* (2006) 26:373–90. doi: 10.1177/0272989X06290492
109. Berkhout C. *Participation in Screening for Cervical Cancer: Interest of a Human Papillomavirus (HPV) Self-sampling Device Provided by the General Practitioner; a Cluster Randomized Clinical Trial [Internet]. Report No.: NCT02749110.* (2020). Available online at: <https://clinicaltrials.gov/ct2/show/NCT02749110> (accessed April 26, 2021).
110. Levesque D, Umanzor C, de Aguiar E. Stage-based mobile intervention for substance use disorders in primary care: development and test of acceptability. *JMIR Med Inform.* (2018) 6:e1. doi: 10.2196/medinform.7355
111. Galdas P. Revisiting bias in qualitative research: reflections on its relationship with funding and impact. *Int J Qual Methods.* (2017) 16:1609406917748992.
112. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med J Assoc Am Med Coll.* (2014) 89:1245–51.