BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

A systematic review of clinician-directed nudges in healthcare contexts

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-048801
Article Type:	Original research
Date Submitted by the Author:	08-Jan-2021
Complete List of Authors:	Last, Briana; University of Pennsylvania, Psychology Buttenheim, Alison; University of Pennsylvania School of Nursing, Department of Family and Community Health; University of Pennsylvania Perelman School of Medicine, Center for Health Incentives and Behavioral Economics (CHIBE) Timon, Carter; University of Pennsylvania, College of Liberal and Professional Studies Mitra, Nandita; University of Pennsylvania Perelman School of Medicine, Department of Biostatistics, Epidemiology & Informatics Beidas, Rinad; University of Pennsylvania Perelman School of Medicine, Department of Psychiatry; University of Pennsylvania Perelman School of Medicine, Department of Medical Ethics and Health Policy
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

A systematic review of clinician-directed nudges in healthcare contexts

Briana S. Last[1], Alison M. Buttenheim,[2.3,4] Carter E. Timon,[5] Nandita Mitra[6], & Rinad S. Beidas[3,4,7-9]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA

[2]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA

[3]Center for Health Incentives and Behavioral Economics (CHIBE), University of Pennsylvania, Philadelphia, PA

[4]Penn Implementation Science Center at the Leonard Davis Institute of Health Economics (PISCE@LDI), University of Pennsylvania, Philadelphia, PA

[5]College of Liberal and Professional Studies, University of Pennsylvania, Philadelphia, PA

[6] Department of Biostatistics, Epidemiology & Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[7]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

For peer review only

[8]Department of Medical Ethics and Health Policy, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[9]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

For peer review only

Author Note

The authors would like to thank Mitesh Patel, Anne Larrivee, Melanie Cedrone, Pamela Navrot, and Amarachi Nasa-Okolie for their assistance in the project.

Correspondence concerning this article should be addressed to Briana S. Last, Department of Psychology, Stephen A. Levin Building, 425 S University Ave, Philadelphia, PA 19104.

Contact: brishiri@sas.upenn.edu

For peer review only

Abstract

Objective: Nudges are interventions that alter the way options are presented, making the optimal option more likely to be chosen. Health systems and researchers have tested nudges to shape clinician decision-making with the aim of improving health service delivery. We aimed to systematically study the use and effectiveness of nudges designed to improve clinicians' decisions in healthcare settings.

Design: A systematic review was conducted to collect and consolidate results from studies testing nudge strategies and to determine whether nudges directed at improving clinical decisions in healthcare settings across clinician types were effective. We systematically searched seven databases (EBSCO Megafile, EconLit, Embase, PsycInfo, PubMed, Scopus and Web of Science) and used a snowball sampling technique to identify peer-reviewed published studies available between 1 January 1989 and 22 April 2020. Eligible studies were critically appraised and narratively synthesized. We categorized nudges according to a nudge taxonomy derived from the Nuffield Council on Bioethics. Included studies were appraised using the Cochrane Risk of Bias Assessment Tool.

Results: We screened 3,586 studies and included 39 studies that met our criteria. The majority of studies (90%) were conducted in the United States and 38% were randomized controlled trials. The most commonly studied nudge strategy (46%) framed options for clinicians, often through social comparisons. Nudges that changed the default options or enabled choice for clinicians were also frequently studied (29%). Default nudges were effective, whereas evidence for the effectiveness of other nudge types was mixed. Given the inclusion of non-experimental designs, only a small portion of studies were at minimal risk of bias (33%) across all Cochrane criteria.

For peer review only

Conclusions: Nudges that change the default options or enable choice in the electronic health record are frequently studied and show promise in improving clinical decision-making. Future work should examine how nudges compare to policy interventions in improving healthcare.

For peer review only

Strengths and limitations of this study

- This systematic review synthesizes the growing research applying nudges in healthcare contexts to improve clinical decision-making.
- Our review uses both systematic search strategies and a snowball sampling approach, the latter of which is useful for identifying relatively novel literature.
- Meta-analysis was not possible due to heterogeneity in methods and outcomes.
- The systematic review was not designed to synthesize research wherein study authors did not identify the intervention as a nudge.

For peer review only

A systematic review of clinician-directed nudges in healthcare contexts

Rationale

Research from economics, cognitive science, and social psychology have converged on the finding that human rationality is "bounded" (1). The intractability of certain decision problems, constraints on human cognition, and scarcity of time and resources lead individuals to employ mental shortcuts to make decisions. These mental shortcuts, often called heuristics, are strategies that overlook certain information in a problem with the goal of making decisions more quickly than more deliberative methods (2). While heuristics can often be more accurate than more complex mental strategies, they can also lead to errors and suboptimal decisions (2,3). Researchers have discovered strategies to harness the predictable ways in which human judgment is biased to improve decisions. These strategies, known as nudges, reshape the "choice architecture," or the way options are presented to decision-makers to optimize choices (4). Nudges have been applied to retirement savings, organ donation, consumer health and wellness, and climate catastrophe mitigation demonstrating robust effects (5–8).

As with retirement savings and dietary choices, clinical decision-making—the process healthcare providers undergo when determining who needs what and when—is complex and error-prone. Clinicians often use heuristics when making diagnostic and treatment decisions (9–11). For example, clinicians are influenced by whether treatment outcomes are framed as losses or gains (e.g., doctors prefer to choose a riskier treatment when the outcome is framed in terms of lives lost rather than lives saved) (12). Heuristics can lead to medical errors (13). In the face of complex medical decisions, clinicians tend to choose the default treatment option (despite clinical guidelines) or conduct clinical examinations that confirm their priors (14,15).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Choice architecture guides clinicians' behavior regardless of whether clinicians are conscious of it, creating opportunities for nudges. Clinical decisions are increasingly made within digital environments such as electronic health record (EHR) systems (16). More than 90% of US hospitals now use an EHR (17,18). Researchers have explored the potential to use these ubiquitous electronic support systems to shape clinical decisions through nudges. They have subtly modified the EHR choice architecture by changing the default options for opioid prescription quantities or by requiring physicians to provide free-text justifications for antibiotic prescriptions (19). Even when nudges are not implemented in the EHR (e.g., peer comparison nudges) researchers extract aggregate data from the EHR, suggesting its increasing role in the study of clinical decision-making (20).

As health systems and researchers have embraced nudges in recent years, there is growing interest in understanding which nudges are most effective to improve clinical decision-making. Taxonomizing nudges is advantageous because many nudges explicitly target heuristics, revealing the mechanism of behavior change, which several strategies to improve clinical decisions cannot do (21). If nudges that leverage people's tendency to adhere to social norms are consistently more effective than nudges that exploit clinicians' default bias, then future nudges can be designed with this insight. Two systematic reviews were recently conducted to evaluate the effectiveness of healthcare nudges. Though both reviews demonstrate promise for the effectiveness of nudges, they offer conflicting evidence on the most studied and most effective nudge types, suggesting that an additional review may be useful (22,23). Our review offers complementary and non-overlapping insights on the study of nudges in healthcare settings for the following reasons: (1) we do not exclusively study physicians as our target population,

For peer review only

instead we include all healthcare workers (23) and (2) we do not restrict our research to randomized controlled trials reported in the Cochrane Library of systematic reviews (22).

Our review also makes use of a nudge taxonomy derived from the widely cited Nuffield Council on Bioethics intervention ladder wherein interventions increase in potency and constrain choice with each new rung (24,25). Interventions on the bottom of the ladder tend to be more passive, offering decision makers information and reminders. Interventions in the middle of the ladder leverage psychological insights to motivate decision-makers either through social influence or by prompting planning action. At the top of the ladder, interventions are more assertive and reduce decisions to a limited set of choices or by creating default options. The Nudge Ladder categorizes nudges by the psychological mechanisms by which they operate, the degree to which they maintain autonomy, and have the additional advantage of aligning with existing public health and quality improvement literature that make use of the Nuffield Council ladder (4,26). The Nudge Ladder offers insights on the heuristics most relevant to the clinical decision-making process and can support health systems in selecting and applying nudges to improve clinical decision-making.

Objective

We systematically evaluated nudge interventions directed at clinicians in healthcare settings to determine the types of nudges that are most studied and most effective in improving clinical decision-making compared with other nudges, non-nudge interventions, or usual care. All quantitative study designs were included in our review.

Methods

Protocol and Registration

For peer review only

Before initiating this review, we searched the international database PROSPERO to avoid duplication. After establishing that no such review was underway, we prospectively registered our review (https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=123349).

Eligibility Criteria

Types of Participants

We included only empirical studies published in peer-reviewed journals studying nudges directed at clinicians working in healthcare settings. Clinicians were defined as workers who provide healthcare to patients in a hospital, skilled nursing facility, or clinic. Examples of clinicians include physicians, nurses, medical assistants, physician assistants, clinical psychologists, clinical social workers, and lay health workers. Studies that exclusively nudged patients were not included.

Types of Intervention

Nudges were defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (4). Alterations to choice architecture included changes to the information provided to the clinician (e.g., translating information, making information visible, providing a social reference point), altering the decision structure of the provider (e.g., changing choice defaults, changing option-related effort, changing the range or composition of options, or changing option consequences) and providing decision assistance (e.g., providing reminders or commitment devices) (27). The study authors did not need to identify the intervention as a nudge to be considered for study inclusion, however given the systematic search string, which includes several behavioral economics terms (see Appendix A), studies that did not self-identify as behavioral economic interventions were unlikely to be included.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Interventions that required sustained education or training were not considered nudges. No options could be forbidden and there could be no financial incentives (28). Though some financial incentives for clinicians may be considered nudges, most studies on financial incentives for clinicians involve significant compensation or "pay for performance"—of which there is already an existing literature (29).

Nudges guided clinicians to make improved clinical decisions, including (but not limited to) increasing the uptake of evidence-based practices (EBPs), adherence to health system or policy guidelines, and reducing healthcare service costs. EBPs refer to clinical techniques and interventions that integrate the best available research evidence, clinical expertise, and patient preferences and characteristics (30). Study authors had to provide the evidentiary rationale for the nudge.

We did not include studies that analyzed the sustainability of nudges in the same study setting and/or sample of providers. In order to analyze studies with independent samples, we included the primary paper and not follow-up papers.

Types of Studies

All study designs were included that had a control or baseline comparator—the control or baseline could be usual care or another intervention (nudge or non-nudge). For studies with parallel intervention groups, we did not require that allocation of interventions be randomized (i.e., quasi-experimental studies were included). Exclusively qualitative studies were not included. See Table 1 for Eligibility Criteria.

Search

Snowball Sampling

For peer review only

The initial search strategy was based on a snowball sampling method (31) using the references from a published commentary referred to BSL on the uses of nudges in healthcare contexts (19). Reviews identified during the preliminary stage of the systematic search process were also used to snowball articles, though these largely resulted in duplicates. Articles were reviewed at the title level to immediately identify those to be excluded. Those tentatively included were reviewed at the abstract level, followed by the full text for those meeting criteria. Following completion of screening of records retrieved via snowball, a systematic search of several databases was completed.

Information Sources & Systematic Search

The methodology for the search was designed based on standards for systematic reviews (32), in consultation with a medical librarian, as well as with two experts from the field of healthcare behavioral economics. The databases used were: EconLit, Embase, EBSCO Megafile, PsycINFO, PubMed, Scopus, and Web of Science.

Search terms included combinations, plurals, and various conjugations of the words relating to identified nudge strategies. The search string and strategy from (6) was used as a basis for search terms, but adjusted to reflect our research question (see Table 1). All peer-reviewed empirical studies published prior to the completion of our search phase (i.e., – 4/2020) were eligible for this review. See Appendix A for the search strings.

Data Collection Process

Following retrieval of all records, duplicates were removed using Zotero (www.zotero.org) and via manual inspection. Article screening involved two stages. First, all records were screened at the title and abstract level by a team of four coders (BSL, CET, and two research assistants) using the web-based application for systematic reviews, Rayyan

For peer review only

(https://rayyan.qcri.org). Criteria in this first-pass screening were more inclusive—that is, all interventions directed at clinicians were included. To establish reliability, the coders screened the same 20 articles and then reviewed their screening decisions together. Any disagreements were resolved by consensus. This process was repeated three additional times until 80 articles were screened by all four coders and sufficient reliability was established. Reliability was excellent (fleiss' $\kappa = 0.96$). For the remainder of the screening process, screening was done independently by all four coders; the team met weekly to discuss edge cases. This screening process was followed by a full text examination to determine eligibility according to more stringent inclusion and exclusion criteria (see Table 1). This screening process was done as a team and determinations of article inclusion were decided by consensus.

Patient and Public Involvement

Patients were not involved in the design or conduct of this research.

Data Items

Study characteristics and outcomes were extracted and tabulated systematically per recommendations for systematic reviews (32). These data included: (1) study characteristics — author names, healthcare setting, study design, country, date of publication, details of the intervention, justification for the nudge, sample size, primary outcomes, main findings, and whether the effect was statistically significant; (2) nudge category; and (3) risk of bias assessment.

BSL and RSB trained the coding team (four Masters level students in a Behavioral and Decision Sciences program) in data extraction. The team coded articles (n=16) together to ensure consensus. RSB reviewed a random sample (n=5) of the final articles to ensure reliability with systematic review reporting standards. BSL subsequently coded the remaining articles (n=18).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Outcomes

We only included studies that included objective measures of clinician behavior in real healthcare contexts. Studies that measured clinicians' choices in vignette or simulation studies were not included. Results could be presented as either continuous (e.g., number of opioid pills prescribed) or binary (e.g., whether physicians ordered influenza vaccinations). Outcomes were measured either directly or indirectly (e.g., using cost to estimate changes in antibiotic prescriptions). Participants could not report on their own behavior because clinicians' self-report can be inaccurate (33). Both absolute measurements and change relative to baseline were accepted.

Risk of Bias in Individual Studies

We evaluated whether the studies included in the systematic review were at risk for bias, using the Cochrane Risk of Bias Tool (32,34). BSL trained CET and they assessed articles (*n*=2) together to ensure consensus. CET independently coded (*n* =12) articles and BSL coded the remaining articles (*n* = 27). The team met weekly to discuss any articles that they were uncertain about and resolved discrepancies by consensus.

Data Synthesis

In order to examine which types of nudges were most studied and most effective, we calculated the number and percentage of studies using each nudge strategy according to the Nudge Ladder (see Figure 1). We reported the effect and statistical significance of the effect when a primary outcome was clearly identified in the study. If no primary outcome was identified by study authors, we determined a primary outcome based on the main research question. For studies that reported multi-component nudges—i.e., interventions that combine several nudges together—we reported the total effect of the intervention. For multicomponent

For peer review only

nudge interventions, we coded them according to the Nudge Ladder with all of the nudge types that apply. For studies with multiple nudge treatment groups, we reported the effect of each treatment arm separately. Only nudge strategies were compared to the control arms.

Due to the differences in the exposure, behavioral outcomes, and study designs interventions could not be directly compared with one another quantitively using effect sizes (35). Hence, meta-analysis of nudge effects was infeasible. To synthesize the results, we used a vote counting method based on the direction of the effect for each study, an acceptable method for synthesis when meta-analysis is not possible (32). If a simple majority of nudges were effective in a nudge category, the category was deemed effective.

Results

Study Selection

The systematic database search identified 3,586 entries, which were combined with another 22 articles of interest identified by the snowball sampling method, totaling 3,608 articles (see Appendix A for yield). After deduplication of records from the respective databases and snowball sampling techniques, 2,486 article records remained. Of the 2,486 articles, 2,486 articles from the systematic search and snowball method were retrievable and screened in the first stage of title and abstract screening, which reduced the total number of full-text screens to 133 unique articles. Of the 133 articles that were full-text screened, 39 articles (20,36–73) met inclusion criteria and the data from these were extracted and evaluated in this review (see PRISMA Diagram in Figure 2).

Study Characteristics

The characteristics of the included studies are summarized in Table 2. The majority (n = 35, 90%) of studies were conducted in the USA; two (5%) were conducted in the United Kingdom, one (3%) in Belgium, and one (3%) in Switzerland. Studies were set in a variety of healthcare

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

contexts (e.g., outpatient clinics, primary care practices, emergency departments, etc.) and targeted a variety of healthcare decisions (e.g., opioid prescriptions, preventative cancer screening, checking hospitalized patients' vitals). Nudges were directed at a variety of medical professionals (including physicians, nurses, medical assistants, and providers with a license to prescribe medication). Many (n = 21, 54%) of the studies did not report the sample size of clinicians interacting with the nudges. Instead, the studies tended to report the sample size in terms of how many patients were affected by the nudge or the number of prescription or lab orders under study. Fifteen (38%) studies were RCTs; 22 studies (56%) were pre-post designs; one study (3%) was a controlled interrupted time series design; and one study (3%) was a non-randomized controlled design. In terms of clustered RCTs, four studies (10%) were parallel cluster RCTs; four studies were stepped wedge cluster RCTs (10%). Most studies (n = 32, 82%) employed a control group/comparator that consisted of usual care or no intervention. One study (3%) used a minimal educational intervention, another study (3%) studying peer comparison letters used a placebo letter and five studies (13%) employed a factorial design in which multiple combined interventions were tested against individual interventions separately.

Of the 39 studies included in the review, 48 nudges were tested. Some studies contained multiple sub-studies, study arms or treatment groups, which were coded and analyzed separately (see Table 3). Given that some interventions (n = 5) were multicomponent (i.e., combinations of multiple nudges) these studies were analyzed separately using the Nudge Ladder (see Table 4).

Analyzing the single component nudges using the Nudge ladder, 6 nudges involved guiding choice through default options (e.g., changing the default opioid prescription quantity in the EHR); 9 nudges involved enabling choice (e.g., electronic prompts to accept or cancel orders for influenza vaccination); 22 nudges involved framing information (e.g., peer comparison letters to the clinicians in the top 50th percentile of antipsychotic prescriptions); two nudges involved prompting implementation intentions (e.g., displaying clinicians' pre-commitment letters in their own examination rooms) and four nudges involved providing information (e.g., an EHR reminder to clinicians when their patients were due for immunizations). Five studies involved multicomponent nudges, with four studies involving a combination of two nudges and one study involving a combination of three nudges (see Table 4).

Risk of Bias of Included Studies

Most studies were judged as high risk for selection bias including random sequence generation (n = 25) and allocation concealment (n = 25). Attrition bias was low risk based on incomplete outcome data (n = 31). A large number of trials were judged as unclear for selective reporting (n = 21). In terms of blinding of participants, most studies were high risk (n=25) and in terms of blinding outcome assessment, 25 studies were judged as having unclear risk of bias. Overall, 13 studies (33%) were considered low risk of bias across all criteria (see Table 5).

Synthesis of Results

With significance defined as ($p<0.05$), 33 of the 48 nudges (73%) significantly affected provider behavior by improving clinical decisions. Thus, overall nudges were effective in improving the targeted clinician behavior. According to the Nudge Ladder, all 6 (100%) of the nudges that involved changing the default option significantly influenced clinician behavior. Seven of the 9 (78%) nudges that enabled choice suggesting their effectiveness. Fifteen of the 22 (68%) nudges that involved framing information significantly shaped behavior, suggesting their effectiveness. One of the two (50%) nudges that prompted implementation intentions, one was significantly effective and the other was not. None of the four (0%) nudges that provided information to clinicians resulted in statistically significant results. The five studies (100%) that combined nudges in multicomponent interventions were all effective.

Changing the default option either by guiding choice through default options or by enabling choice through an "active opt-out" model were the most effective strategies in changing clinician behavior. These nudges also tended to result in the largest effect sizes—default or active choice interventions resulted in clinically significant changes in clinician behavior. One study resulted in a three-fold increase in prescribing behavior (57). Nudges that framed information—the plurality of nudges under study—tended to also significantly influence clinician behavior. However, evidence for framing information was more mixed suggesting that more work needs to be done to establish effectiveness. All of the other types of nudges were inconclusive or had more insignificant findings than significant findings.

Discussion

Summary of Evidence

This systematic review of 39 studies found that a variety of nudge strategies have been tested to improve clinical decisions. Thirty-three of the 48 (73%) clinician-directed nudges significantly improved clinical practice in the hypothesized direction, mostly by altering the default options of the clinicians' choice environment or by framing information by creating social reference points or providing direct feedback to clinicians. Nudges that changed default options or enabled choice were the most effective and nudges framed information for clinicians were also largely effective. Conversely, nudges that provided information to the clinician through reminders and prompting implementation intentions did not conclusively lead to significant changes in clinician behavior.

For peer review only

One strength of nudges and the taxonomy utilized for this review is that we can determine why certain nudges fail over others and the mechanism by which they operate. Drawing on the Nudge Ladder, we find evidence that less aggressive healthcare nudges such as providing information and prompting intentions may be less effective than more aggressive nudges such as changing the default options. This accords with nudge research in other areas outside of healthcare (74). For example, one study comparing various types of nudges that increase the salience of information (e.g., including providing reminders, leveraging social norms, and framing information) with defaults found that only default nudges were effective at changing consumer pro-environmental behavior (8). One large RCT of calorie labeling in restaurants found that posting caloric benchmarks (an informational nudge) paradoxically increased caloric intake for consumers (75).

The theoretical reasons for why nudges at the bottom of the Nudge Ladder often fail are well established. People have a limited capacity to process information, so providing more data to decision makers can be distracting or cognitively loading (76). The timing of information is also essential—information is beneficial if it is top-of-mind during the decision (77). Some of the social comparison nudges in this review provided information at opportune times, others did not (43). Additionally, information improves decisions only if existing heuristics encourage errors. Often the information individuals receive may not be new to them, falling on deaf ears. Worse still, informational nudges can have negative unintended consequences. For example, alert fatigue describes when clinicians are so inundated by alerts that they become desensitized and either miss or delay their responses to them (78). Finally, often reminders and information frames can be insufficiently descriptive in the course of action they suggest, rendering them

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

futile. Given how much of clinicians' time is spent with the EHR, health system decision supports must be effective and not self-undermining.

Nudges at the top of the Nudge Ladder are successful because they act on several key heuristics (79). Defaults leverage inertia wherein overriding the default requires an active decision (80). When people are busy and their attention scarce, they tend to rely on the status quo (81). Moreover, people often see the default option as signaling an injunctive norm (82). They see the default choice as the recommended choice and don't want to actively override this option unless they are very confident in their private decision. It is unsurprising that our study found that defaults were effective. It is also unsurprising that social comparison nudges tended to also be effective at shaping clinician behavior—clinicians who received messages that their behavior was abnormal compared to their peers, were receiving a signal that helped them update their behavior.

Overall, our results aligned with the results of one (23) of the two recent systematic reviews of nudges tested in healthcare settings (22,23). Differences in findings may be explained by different search strategies. One of these systematic reviews exclusively searched RCTs included in the Cochrane Library of systematic reviews and found that priming nudges—nudges that provide cues to participants—were the most studied and most effective nudges (22). In this review priming encompassed heterogenous interventions that span cues that elude conscious awareness, audit-and-feedback, and clinician reminders—to name a few—which may account for why study authors found those nudges to be the most numerous. The findings from our review conform with the results of the more traditional systematic review, conducted using a systematic search of several databases (23). The latter review, like ours, found that default nudges and social comparison nudges were the most frequently studied and most effective nudges. However, study authors focused their review on physician behavior, and our review is more expansive by studying all healthcare workers.

Limitations

Many of the studies in this review included at least some education (i.e., a non-nudge intervention) such as a reminder of the clinical guidelines. Because many studies (56%) were pre-post designs, they could not use these brief trainings in a control arm to evaluate the independent effect of the nudge. Therefore, we cannot decisively conclude whether nudges alone are responsible for the changes in clinician behavior. Similarly, many of the studies (54%) did not report the number of clinicians involved in the study (often reporting the sample in terms of how many patients or lab orders were affected by the nudge). Though unlikely, many of the effects could presumably be driven by a small portion of clinicians.

There was considerable variability in how researchers operationalized their outcome of interest. The effect of nudges may be contingent upon the behavior under study. One study (72) examining changes in opioid prescriptions only reported the change in the number of 15-pill prescriptions (i.e., the change in "default" orders), whereas other studies examined the change in the total number of opioid pills ordered after an EHR default change (83). The former study found a significant change in the number of 15-pill opioid orders, but the authors did not report whether the total number of opioid pills decreased. Reporting on the latter would enable not only a direct comparison across studies, but would allow us to conclusively determine if the nudge was effective overall at improving clinical decisions.

The considerable number of included papers reporting a statistically insignificant result decreases somewhat the usual concern over publication bias, which would skew the results towards desirable and more statistically significant outcomes. The majority of studies (n =21, 54%) were at unclear risk of selective reporting of outcomes (See Table 5). Moving forward, the field would benefit from reporting of all experimentation, whether its results are successful, unsuccessful, significant, or insignificant. Though not a majority, a large portion of studies (*n* = 12, 31%) were conducted by the same research team in the same health system. To validate that clinician-directed nudges are effective in other settings, other researchers should conduct nudge studies.

Though the nudge taxonomy used in the current review offered a way to classify the nudges described in the studies included, it was not developed empirically. The Nudge Ladder was developed based on a theoretical understanding of public health interventions, not through analysis. It's important to understand whether the conceptual distinctions made between nudge types are scientifically reliable and valid.

Future Research

Behavioral economics recognizes that nudges are "implicit social interactions" between the decision maker and the choice architect (84). When faced with a nudge, people evaluate the beliefs and intentions of the choice architect as well as how their decision will be construed by the choice architect and others. People tend to adhere to the default option when the choice architect is trusted, benevolent, and competent. Several non-healthcare studies found defaults backfiring when consumers distrust the choice architect or feel they are being nudged to spend more money (85). Clinicians may reject nudges when they perceive health systems' preferences to conflict with their patients' interests. Research should attend to how engaged clinicians are in the implementation process and how they make inferences about the intentions and beliefs of the choice architect when interacting with nudges using qualitative methods and surveys.

Nudges are also dependent on how decision makers believe they will be perceived. For example, around 40% of adults seeking care for upper respiratory tract infections want antibiotics (though it is inappropriate) and general practitioners report that patient expectations are a major reason for prescribing antibiotics (86,87). Nudges that attempt to curtail antibiotic prescribing behavior may shape clinicians' behaviors in unexpected ways given clinicians' desire to demonstrate to their patients that they are taking serious action. Subtle features of how nudges are deployed may also influence clinicians' perceptions of the choice architect, heighten awareness of how their own actions may be perceived, and may undermine the intention of the nudge. Investigations of the clinicians' choice environment and clinicians' perspectives using qualitative and survey methods are crucial to the success of nudges.

Future research should also explore how clinician-directed nudges interact with one another in clinicians' choice environments. In our review, all multicomponent nudge studies (n = 5) were effective. However, it is also possible that nudges may crowd each other out when several different clinical decisions are targeted. In addition to alert fatigue, clinicians may experience nudge fatigue and begin to ignore decision support embedded in the EHR. Research should seek to understand how to develop nudges that can work synergistically with one another. Health systems and scientists can work together to understand which guidelines to prioritize and to develop decision support systems within their electronic interfaces that guide providers to make better clinical decisions.

Little work has been done on the sustainability of nudges beyond the study period, with some notable exceptions (88). Particularly for nudges that require continued intervention on the part of the choice architects (e.g., peer comparison interventions), it's necessary to also understand their cost-effectiveness.

Conclusion

This study adds to the growing literature on the study and effectiveness of nudges in healthcare contexts and can guide health systems in their choices of the types of nudges they should implement to improve clinical practice. The review describes how nudges have been employed in healthcare contexts and the evidence for their effectiveness across clinician behaviors, demonstrating potential for nudges, particularly nudges that change default settings or frame information for clinicians. More research is warranted to examine how nudges scale and their global effect on improving clinical decisions in complex healthcare environments.

For peer review only

Authors' Contributions

BSL conceived of and designed the research study; acquired and analyzed the data; interpreted the data; drafted the manuscript and substantially revised it. AMB helped design the research study; analyzed the data; interpreted the data; and substantially revised the manuscript. CET analyzed the data; interpreted the data; and substantially revised the manuscript. NM interpreted the data and substantially revised the manuscript. RSB helped conceive of and design the research study; interpreted the data; and substantially revised the manuscript. All authors approved the submitted version; have agreed to be accountable for the contributions; attest to the accuracy and integrity of the work, even aspects for which the authors were not personally involved.

Competing Interests

The authors declare no financial or non-financial competing interests.

Patient and Public Involvement

Patients and/or the public were not involved in the design, or conduct, or reporting of this research.

Patient Consent for Publication

Not required.

Ethics Approval

Given the nature of systematic reviews, no human participant research was conducted for this original research contribution. Thus, the systematic review was not deemed subject to ethical approval and no human participants were involved in this study.

Funding

For peer review only

Funding for this study was provided by grants from the National Institute of Mental Health (P50 MH 113840, Beidas, Buttenheim, Mandell, MPI). Briana S. Last also receives funding support from the National Science Foundation Graduate Research Fellowship Program (DGE-1321851).

Availability of Data and Materials

Given the nature of systematic reviews, the dataset generated and analyzed for the current study is already available. All studies analyzed for the present review are referenced for readers.

For peer review only

References

1. Simon HA. Models of bounded rationality: Empirically grounded economic reason. Vol. 3. Cambridge, MA: MIT press; 1997.

2. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annual review of psychology. 2011;62:451–82.

3. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974 Sep 27;185(4157):1124.

4. Thaler RH, Sunstein CR. Nudge: Improving decisions about health, wealth, and happiness. Revised and Expanded Edition. New York, NY: Penguin Books; 2009.

5. Szaszi B, Palinkas A, Palfi B, Szollosi A, Aczel B. A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. Journal of Behavioral Decision Making. 2018;31(3):355–66.

6. Arno A, Thomas S. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. BMC Public Health. 2016 Jul 30;16:676.

7. Harbers MC, Beulens JW, Rutters F, de Boer F, Gillebaart M, Sluijs I, et al. The effects of nudges on purchases, food choice, and energy intake or content of purchases in real-life food purchasing environments: a systematic review and evidence synthesis. Nutrition Journal. 2020;19(1):1–27.

8. Momsen K, Stoerk T. From intention to action: Can nudges help consumers to choose renewable energy? Energy Policy. 2014;74:376–82.

9. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. Medical Decision Making. 2015;35(4):539–57.

10. Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009 Aug;84(8):1022–8.

11. Graber M. Diagnostic Errors in Medicine: A Case of Neglect. Joint Commission Journal on Quality and Patient Safety. 2005;31(2):106–13.

12. Almashat S, Ayotte B, Edelstein B, Margrett J. Framing effect debiasing in medical decision making. Patient education and counseling. 2008;71(1):102–7.

13. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC medical informatics and decision making. 2016;16(1):138.

For peer review only

14. Mendel R, Traut-Mattausch E, Jonas E, Leucht S, Kane JM, Maino K, et al. Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. Psychological Medicine. 2011;41(12):2651–9.

15. Redelmeier DA, Shafir E. Medical decision making in situations that offer multiple alternatives. Jama. 1995;273(4):302–5.

16. Sittig DF, Singh H. Electronic health records and national patient-safety goals. Mass Medical Soc; 2012.

17. Chaiyachati KH, Shea JA, Asch DA, Liu M, Bellini LM, Dine CJ, et al. Assessment of Inpatient Time Allocation Among First-Year Internal Medicine Residents Using Time-Motion Observations. JAMA Internal Medicine. 2019;

18. Hsiao C-J, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012. 2012;

19. Patel MS, Volpp KG, Asch DA. Nudge Units to Improve the Delivery of Health Care. The New England journal of medicine. 2018;378(3):214.

20. Sacarny A, Barnett ML, Le J, Tetkoski F, Yokum D, Agrawal S. Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial. JAMA Psychiatry. 2018 Oct 1;75(10):1003–11.

21. Lewis CC, Klasnja P, Powell B, Tuzzio L, Jones S, Walsh-Bailey C, et al. From classification to causality: advancing understanding of mechanisms of change in implementation science. Frontiers in public health. 2018;6:136.

22. Yoong SL, Hall A, Stacey F, Grady A, Sutherland R, Wyse R, et al. Nudge strategies to improve healthcare providers' implementation of evidence-based guidelines, policies and practices: a systematic review of trials included within Cochrane systematic reviews. Implementation Science. 2020;15(1):1–30.

23. Wang SY, Groene O. The effectiveness of behavioral economics-informed interventions on physician behavioral change: A systematic literature review. Plos one. 2020;15(6):e0234149.

24. Council NB. Public health: ethical issues. Nuffield Council on Bioethics. 2007;

25. Patel MS. Nudges for influenza vaccination. Nature Human Behaviour. 2018;2(10):720.

26. Hillier-Brown FC, Summerbell CD, Moore HJ, Routen A, Lake AA, Adams J, et al. The impact of interventions to promote healthier ready-to-eat meals (to eat in, to take away or to be delivered) sold by specific food outlets open to the general public: a systematic review. Obesity Reviews. 2017;18(2):227–46.

27. Münscher R, Vetter M, Scheuerle T. A review and taxonomy of choice architecture techniques. Journal of Behavioral Decision Making. 2016;29(5):511–24.

For peer review only

28. Sunstein CR. Nudges vs. shoves. Harv L Rev F. 2013;127:210.

29. Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. Cochrane database of systematic reviews. 2011;(7).

30. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. British Medical Journal Publishing Group; 1996.

31. Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. Sociological methods & research. 1981;10(2):141–63.

32. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. 2008;

33. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. Jama. 2006;296(9):1094–102.

34. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj. 2011;343:d5928.

35. Nugent WR. The (Non) Comparability of the Correlation Effect Size Across Different Measurement Procedures: A Challenge to Meta-Analysis as a Tool for Identifying "Evidence Based Practices." Journal of evidence-based social work. 2011;8(3):253–74.

36. Allen J.M., Dunn R., Bush J. Effect of prescriber peer comparison reports on fluoroquinolone use across a 16-facility community hospital system. JACCP JAm Coll Clin Pharm. 2019;2(5):502–8.

37. Andereck J.W., Reuter Q.R., Allen K.C., Ansari S., Quarles A.R., Cruz D.S., et al. A Quality Improvement Initiative Featuring Peer-Comparison Prescribing Feedback Reduces Emergency Department Opioid Prescribing. Jt Comm J Qual Patient Saf. 2019;45(10):669–79.

38. Arora VM, Mochado N, Anderson SL, Desai N, Marsack W, Blossomgame S, et al. Effectiveness of SIESTA on Objective and Subjective Metrics of Nighttime Hospital Sleep Disruptors. J Hosp Med. 2019 Jan;14(1):38–41.

39. Bourdeaux CP, Davies KJ, Thomas MJ, Bewley JS, Gould TH. Using 'nudge'principles for order set design: a before and after evaluation of an electronic prescribing template in critical care. BMJ Qual Saf. 2014;23(5):382–8.

40. Buntinx F, Knottnerus JA, Essed GG, Crebolder HF. Long-term effect of feedback and peer comparison on the sampling quality of cervical smears--a randomized controlled trial. Eur J Cancer Prev. 1995 Apr;4(2):153–7.

For peer review only

41. Chiu AS, Jean RA, Hoag JR, Freedman-Weiss M, Healy JM, Pei KY. Association of Lowering Default Pill Counts in Electronic Medical Record Systems With Postoperative Opioid Prescribing. JAMA Surg. 2018 Nov 1;153(11):1012–9.

42. Delgado MK, Shofer FS, Patel MS, Halpern S, Edwards C, Meisel ZF, et al. Association between Electronic Medical Record Implementation of Default Opioid Prescription Quantities and Prescribing Behavior in Two Emergency Departments. Journal of general internal medicine. 2018;1–3.

43. Hemkens LG, Saccilotto R, Reyes SL, Glinz D, Zumbrunn T, Grolimund O, et al. Personalized prescription feedback using routinely collected data to reduce antibiotic use in primary care: a randomized clinical trial. JAMA internal medicine. 2017;177(2):176–83.

44. Hempel D, Pivetta E, Kimberly HH. Personalized peer-comparison feedback and its effect on emergency medicine resident ultrasound scan numbers. Crit Ultrasound J. 2014 Jan 14;6(1):1.

45. Hsiang EY, Mehta SJ, Small DS, Rareshide CAL, Snider CK, Day SC, et al. Association of an Active Choice Intervention in the Electronic Health Record Directed to Medical Assistants With Clinician Ordering and Patient Completion of Breast and Colorectal Cancer Screening Tests. Jama Network Open. 2019 Nov;2(11):e1915619.

46. Kim RH, Day SC, Small DS, Snider CK, Rareshide CAL, Patel MS. Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination. JAMA Netw Open. 2018 Sep 7;1(5):e181770.

47. Kullgren JT, Krupka E, Schachter A, Linden A, Miller J, Acharya Y, et al. Precommitting to choose wisely about low-value services: a stepped wedge cluster randomised trial. BMJ Qual Saf. 2018;27(5):355–64.

48. Lewis S, Young B, Thurley P, Shaw D, Cranwell J, Skelly R, et al. Evaluation of a nudge intervention providing simple feedback to clinicians of the consequence of radiation exposure on demand for computed tomography: a controlled study. Clinical Medicine. 2019 Jul 1;19(4):290–3.

49. Meeker D, Knight TK, Friedberg MW, Linder JA, Goldstein NJ, Fox CR, et al. Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. JAMA Intern Med. 2014 Mar;174(3):425–31.

50. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. Jama. 2016;315(6):562–70.

51. Nguyen C.T., Davis K.A. Evaluating the impact of peer comparison on vancomycin dose order verification among pharmacists. JACCP JAm Coll Clin Pharm. 2019;2(2):137–42.

For peer review only

52. O'Reilly-Shah VN, Easton GS, Jabaley CS, Lynde GC. Variable effectiveness of stepwise implementation of nudge-type interventions to improve provider compliance with intraoperative low tidal volume ventilation. BMJ Qual Saf. 2018 Dec;27(12):1008–18.

53. Olson J, Hollenbeak C, Donaldson K, Abendroth T, Castellani W. Default settings of computerized physician order entry system order sets drive ordering habits. J Pathol Inform. 2015;6:16.

54. Orloski CJ, Tabakin ER, Shofer FS, Myers JS, Mills AM. Grab a Seat! Nudging Providers to Sit Improves the Patient Experience in the Emergency Department. Journal of Patient Experience. 2019 Jun;6(2):110–6.

55. Parrino TA. The nonvalue of retrospective peer comparison feedback in containing hospital antibiotic costs. Am J Med. 1989 Apr;86(4):442–8.

56. Patel MS, Volpp KG, Small DS, Wynne C, Zhu J, Yang L, et al. Using active choice within the electronic health record to increase influenza vaccination rates. Journal of general internal medicine. 2017;32(7):790–5.

57. Patel M., Day S.C., Halpern S., Hanson B., Martinez J., Honeywell S., et al. Change in generic medication prescribing rates after health system-wide redesign of default options within the electronic health record. J Gen Intern Med. 2016;31(2):S148.

58. Patel MS, Volpp KG, Small DS, Wynn C, Zhu J, Yang L, et al. Using active choice within the electronic health record to increase physician ordering and patient completion of high-value cancer screening tests. Healthc (Amst). 2016 Dec;4(4):340–5.

59. Patel MS, Day S, Small DS, Howell III JT, Lautenbach GL, Nierman EH, et al. Using Default Options Within the Electronic Health Record to Increase the Prescribing of Generic-Equivalent Medications. Annals of Internal Medicine. 2014 Nov 19;161:S44–52.

60. Patel MS, Kurtzman GW, Kannan S, Small DS, Morris A, Honeywell S, et al. Effect of an Automated Patient Dashboard Using Active Choice and Peer Comparison Performance Feedback to Physicians on Statin Prescribing The PRESCRIBE Cluster Randomized Clinical Trial. Jama Network Open. 2018 Jul;1(3):e180818.

61. 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 Aug 5;16:373.

62. Ryskina K, Dine CJ, Gitelman Y, Leri D, Patel M, Kurtzman G, et al. Effect of social comparison feedback on laboratory test ordering for hospitalized patients: a randomized controlled trial. Journal of general internal medicine. 2018;33(10):1639–45.

63. Sedrak MS, Myers JS, Small DS, Nachamkin I, Ziemba JB, Murray D, et al. Effect of a price transparency intervention in the electronic health record on clinician ordering of inpatient laboratory tests: the PRICE randomized clinical trial. JAMA internal medicine. 2017;177(7):939–45.

For peer review only

64. Sharma S, Guttmann D, Small DS, Rareshide CAL, Jones J, Patel MS, et al. Effect of Introducing a Default Order in the Electronic Medical Record on Unnecessary Daily Imaging During Palliative Radiotherapy for Adults With Cancer: A Stepped-Wedge Cluster Randomized Clinical Trial. Jama Oncology. 2019 Aug;5(8):1220–2.

65. Shively N.R., Buehrle D.J., Wagener M.M., Clancy C.J., Decker B.K. Improved antibiotic prescribing within a veterans affairs primary care system through a multifaceted intervention centered on peer comparison of overall antibiotic prescribing rates. Antimicrob Agents Chemother [Internet]. 2020;64(1). Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L2004514290

66. Srinivasan M., Huntman J., Nelson M., Mathew S. Use of peer comparison, provider education, and electronic medical record triggers to increase influenza vaccination rates in hospitalized children. Hosp Pediatr. 2020;10(1):76–83.

67. Suffoletto B., Landau A. Nudging Emergency Care Providers to Reduce Opioid Prescribing Using Peer Norm Comparison Feedback: A Pilot Randomized Trial. Pain Med [Internet]. 2019;((Suffoletto B.) Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States). Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L630281105

68. Szilagyi PG, Serwint JR, Humiston SG, Rand CM, Schaffer S, Vincelli P, et al. Effect of provider prompts on adolescent immunization rates: a randomized trial. Academic pediatrics. 2015;15(2):149–57.

69. Trent SA, Havranek EP, Ginde AA, Haukoos JS. Effect of Audit and Feedback on Physician Adherence to Clinical Practice Guidelines for Pneumonia and Sepsis. Am J Med Qual. 2018 Sep 12;1062860618796947.

70. Wigder HN, Ballis SFC, Lazar L, Urgo R, Dunn BH. Successful implementation of a guideline by peer comparisons, education, and positive physician feedback. Journal of Emergency Medicine. 1999 Oct;17(5):807–10.

71. Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Barnett GO. Improving physician performance through peer comparison feedback. Med Care. 1984 Jun;22(6):527–34.

72. Zivin K, White JO, Chao S, Christensen AL, Horner L, Petersen DM, et al. Implementing Electronic Health Record Default Settings to Reduce Opioid Overprescribing: A Pilot Study. Pain Med. 2019 Jan 1;20(1):103–12.

73. Zwank M.D., Kennedy S.M., Stuck L.H., Gordon B.D. Removing default dispense quantity from opioid prescriptions in the electronic medical record. Am J Emerg Med. 2017;35(10):1567–9.

74. Sunstein CR. Nudges that fail. Behavioural Public Policy. 2017;1(1):4–25.

For peer review only

75. Downs JS, Wisdom J, Wansink B, Loewenstein G. Supplementing menu labeling with calorie recommendations to test for facilitation effects. American Journal of Public Health. 2013;103(9):1604–9.

76. Marois R, Ivanoff J. Capacity limits of information processing in the brain. Trends in cognitive sciences. 2005;9(6):296–305.

77. Karlan D, McConnell M, Mullainathan S, Zinman J. Getting to the top of mind: How reminders increase saving. Management Science. 2016;62(12):3393–411.

78. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN advanced critical care. 2013;24(4):378–86.

79. Jachimowicz JM, Duncan S, Weber EU, Johnson EJ. When and why defaults influence decisions: a meta-analysis of default effects. Behavioural Public Policy. 2019;3(2):159–86.

80. Madrian BC, Shea DF. The power of suggestion: Inertia in 401 (k) participation and savings behavior. The Quarterly journal of economics. 2001;116(4):1149–87.

81. Kahneman D, Knetsch JL, Thaler RH. Anomalies: The endowment effect, loss aversion, and status quo bias. Journal of Economic perspectives. 1991;5(1):193–206.

82. Everett JA, Caviola L, Kahane G, Savulescu J, Faber NS. Doing good by doing nothing? The role of social norms in explaining default effects in altruistic contexts. European Journal of Social Psychology. 2015;45(2):230–41.

83. Zwank MD, Kennedy SM, Stuck LH, Gordon BD. Default versus open text narcotic prescription writing in the emergency department electronic medical record. Acad Emerg Med. 2014;21(5):S87–8.

84. Krijnen JM, Tannenbaum D, Fox CR. Choice architecture 2.0: Behavioral policy as an implicit social interaction. Behavioral Science & Policy. 2017;3(2):i–18.

85. Brown CL, Krishna A. The skeptical shopper: A metacognitive account for the effects of default options on choice. Journal of consumer research. 2004;31(3):529–39.

86. Fletcher-Lartey S, Yee M, Gaarslev C, Khan R. Why do general practitioners prescribe antibiotics for upper respiratory tract infections to meet patient expectations: a mixed methods study. BMJ open. 2016 Oct 24;6(10):e012244–e012244.

87. Linder JA, Singer DE. Desire for antibiotics and antibiotic prescribing for adults with upper respiratory tract infections. Journal of general internal medicine. 2003 Oct;18(10):795–801.

88. Olshan D, Rareshide CAL, Patel MS. Longer-Term Durability of Using Default Options in the Electronic Health Record to Increase Generic Prescribing Rates. J Gen Intern Med [Internet]. 2018;((Olshan D.; Rareshide C.A.L.; Patel M.S., mpatel@upenn.edu) Penn Medicine Nudge Unit, Philadelphia, PA, United States). Available from:

For peer review only

http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L624764295

For peer review only

Table 1. Eligibility Criteria

Inclusion Criteria	Full-text empirical journal articles
	English language
	Published in a Peer-Reviewed Journal
	The studies in the paper empirically investigated one or more behavioral intervention techniques that were considered nudges or were connected to the choice architecture literature by the original authors. These interventions are all clinician-directed (e.g., nurses, doctors, residents, medical assistants), not patient-directed.
	The studies in the paper had behavioral outcome variables, not preferences or attitudes (e.g., prescribing behavior).
Exclusion Criteria	Abstracts unavailable in the first-pass screen
	Review articles, conference abstracts, textbooks, chapters, and conference papers.
	Studies without a control group or baseline comparator
	The studies in the paper applied interventions that restrict the freedom of choice of the target population, included significant economic incentives, ongoing education, complex decision support systems, or consultation.

For peer review only

Figure 1. Ladder of nudge interventions.

Note. Adapted from (24,25).

For peer review only

Figure 2. PRISMA flow diagram

For peer review only

Table 2. Study Characteristics

Authors (Year, Country)	Setting	Design	Intervention	Justification	Sample size	Outcomes Measured	Main findings	Significance
Allen, Dunn, & Bush (2019) USA (36)	Health system (16 community hospitals across 8 counties)	Prospective, pre-post design	Peer comparison reports were distributed to eligible prescribers (by email, fax, or in-person, etc.) provided to eligible prescribers on a quarterly basis. Eligible prescribers, defined as prescribers that accounted for 75%-80% of total prescribed antibiotic days for their peer group. Peer comparison report recipients were not aware that they were high-volume antibiotic prescribers.	Reduce the prescription of antibiotics. Among potential targets for antibiotic use reduction, fluoroquinolones (FQs) are an attractive drug class due to their broad spectrum of activity, known adverse event profile, and availability of other less toxic therapeutic options	Internal medicine; hospitalists; family medicine (n = 189). Critical care; pulmonology (n = 67). Infectious diseases (n = 60)	The primary study outcome was FQ days of therapy/1000 patient days (PDs). A FQ day of therapy was defined as the receipt of at least one dose of a FQ in a 24-hour period, as noted on each facility's bar-coded medication administration (BCMA) records.	FQ DOT/1000 PD declined 29% compared with baseline. FQ DOT/1000 PD declined for all facilities included in the study.	p<0.001
Andereck et al. (2019) USA (37)	Large urban academic Emergency Department	Prospective pre-post design (QI initiative)	Quarterly feedback by e-mail. Individual prescribing rates were grouped by prescriber level and then sorted from lowest to	The opioid epidemic; and unnecessary prescribing patterns.	In the preintervention period, a total of 35,636 ED visits were discharged (mean per block: 3,960).	The primary outcome of this evaluation was the overall ED discharge opioid prescribing rate. Prescribing rate was defined as	Departmental opioid prescribing rates during the evaluation period declined.	p<0.001

For peer review only

			highest within each prescriber class. Prescribers could identify their individual prescribing rate on a de-identified, peer-contextualized chart of their peers. Included a brief "pharmacy fact " with each provision of e-mailed prescribing as well as a formal lecture provided by the pharmacist team member in the setting of a weekly educational conference that is required for all resident physicians and incentivized for attending physicians.		On average, 44 attending physicians, 30 senior resident physicians, and 33 junior resident physicians and advanced practice providers per block met the threshold for inclusion by discharging more than 20 patients per block. In the postinterventio n period, a total of 18,830 ED visits were discharged (mean per block: 3,672). On average, 40 attending physicians, 30 senior residents, and 35 junior residents and advanced practice providers per block met the threshold for inclusion.	the proportion of discharged patient encounters resulting in an opioid prescription for the entire department in a given scheduling block.		

Arora et al. (2019) USA (38)	Two general medicine inpatient units	Prospective, cross-sectional pre-post design	Changing the EHR, creating a default to monitor patient's vital signs; Customized office signage directed at nurses educating them about best "sleep-friendly" vitals monitoring practices during sleep; pocket-cards with information; 20-minute education session.	Sleep is critical to patient recovery in the hospital; hospitalization is not restful, and inpatient sleep deprivation has been linked to poor health outcomes.	*n* = ? providers (1,083 general-medicine patients, 1,669 EHR general medicine orders)	Changes in the mean percentage of "sleep-friendly" (i.e., non-nocturnal) orders for checking vital signs and venous thromboembolism prophylaxis; changes in nighttime room entries/disruptions	Increases in the mean percentage of sleep-friendly orders rose for both vital signs; Decreases in nighttime entries.	p<0.001
Bourdeaux et al. (2014) UK (39)	Inpatient Intensive Care Unit	Retrospective Pre-post design	A prescribing template with some commonly used drugs and fluids preprescribed. Admitting doctors can choose to use the template when compiling the electronic drug chart at admission	Chlorhexidine mouthwash has been shown to reduce the rate of ventilator associated pneumonia in ventilated critically ill patients. It is a low cost intervention with widespread acceptance among clinicians. Hydroxyethyl starch (HES) is an intravenous fluid used to support circulation.	2231 ventilated patients were identified as appropriate for treatment with chlorhexidine, 591 before the intervention and 1640 after. 55.3% were prescribed chlorhexidine before the change and 90.4% after (p<0.001). 6199 patients were considered in the HES intervention, 2177 before the	Evaluate the impact of changes to the design of an order set on the delivery of chlorhexidine mouthwash and hydroxyethyl starch (HES) to patients in the intensive care unit.	The mean volume of HES infused per patient fell and the percentage of patients receiving HES fell.	p<0.001 for both

For peer review only

					intervention and 4022 after			
Buntinx et al. (1993) Belgium (40)	Department of Pathology	Randomized Controlled Trial	Interventions, four groups. Some arms had feedback and then advice. One arm had peer comparison	Cervical screening is important and can help prevent cancer.	183 doctors	Percentage of smears lacking endocervical cells.	A significantly larger decrease in the percentage of smears lacking endocervical cells was found in the groups receiving monthly overviews of their results with peer comparison, when compared with the groups not receiving this type of feedback.	P<0.05
Chiu et al. (2018) USA (41)	Health System (5 hospitals)	Prospective pre-post design	Changing the EHR, lowered the default number of pills on all electronic opioid prescriptions from 30 to 12.	Reliance on prescription opioids for postprocedural analgesia has contributed to the opioid epidemic	*n* = ? providers (1447 procedures before default change and 1463 procedures after the default change)	Changes in prescription rates, the median number of opioid pills prescribed per operation.	Decreases in the median number of opioid pills prescribed	p<0.01
Delgado et al. (2018) USA (42)	Two emergency departments	Prospective pre-post design	Changing the EHR, lowered the default number of	Reliance on prescription opioids for	*n* = ? providers	Changes in the mean number of Oxy/Apap tablets	No change in the mean number of	p<0.001

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

			pills on electronic opioid prescriptions to 10 pills.	postprocedural analgesia has contributed to the opioid epidemic	(3264 prescriptions were written across the two EDs)	prescribed per week	Oxy/APAP tablets prescribed per week. However, increase in proportion of prescriptions for 10 tablets.	
Hemkens et al. (2017) Switzerland (43)	Nationwide	Pragmatic RCT	Personalized antibiotic prescription feedback by mail and an online dashboard and a letter on antibiotic prescribing guidelines	Clinicians often inappropriately prescribe antibiotics for acute respiratory tract infections	2,900 primary care physicians	Changes in the prescribed defined daily doses (DDD) of any antibiotic to any patient per 100 consultations in the first year analyzed by intention-to-treat.	No change in prescribing behavior	N.S.
Hempel et al. (2014) USA (44)	Emergency department	Prospective pre-post design (Differences-in-differences)	Peer comparison feedback on emergency medicine resident ultrasound scan numbers.	Clinician-performed ultrasound has been incorporated into EM residency curricula; need for effective teaching.	44 emergency medicine residents	Changes in number of scans performed in the three months after intervention	Increase in number of scans performed	$p<0.05$
Hsiang et al. (2019) USA (45)	Health System (25 primary care practices)	Retrospective Difference-in-Differences Approach (Intervention vs control practices during post-intervention year compared to	3 Health System primary care practices implemented an active choice intervention in the EHR using a best-practice alert in EPIC directed to medical assistants. Prior to meeting	US Preventive Services Task Force guidelines for breast and colorectal cancer screening	*n* = ? providers The sample eligible for breast cancer screening comprised 26,269 women. The sample	The primary outcome was clinician ordering of the screening test during the primary care visit. The secondary outcome was patient completion of a	Breast cancer screening tests and Colorectal cancer screening test increased.	$p<0.001$

For peer review only

		the 2 preintervention years)	with the clinician, patients met with a medical assistant to check their vital signs and prepare for the visit. At that time, the EHR checked for patient eligibility for breast and colorectal cancer screening and prompted medical assistants to accept or cancel an order for it. If accepted, the order would be templated (a pending order is made for the clinician to review and sign during the patient visit).		eligible for colorectal cancer screening comprised 43,647 men.	screening test (not necessarily linked to the order from the visit) within 1 year of the primary care visit.		
Kim et al. (2018) USA (46)	11 Primary Care Practices	Prospective, cross-sectional pre-post design (Differences-in-differences)	Changing the EHR, "Active choice" intervention in the EHR using a best practice alert directed to medical assistants—prompt to accept or cancel an order for the flu vaccine. If accepted, the order would be	Center for Disease Control recommends influenza vaccination for everyone	*n* = ? providers (96, 291 patients)	Changes in flu vaccination rates	Increase in flu vaccination rates	p<0.001

			templated for the physician to review and sign during the patient visit.					
Kullgren et al. (2018) USA (47)	6 adult primary care practices	12-month stepped wedge cluster RCT, randomization by clinic	Clinicians precommited to Choosing Wisely recommendations against low-value/inappropriate services. They received 1–6 months of point-of-care precommitment reminders as well as patient education handouts and weekly emails with resources to support communication about low-value services.	Clinicians often excessively order costly and inappropriate tests as well as inappropriately prescribe antibiotics for acute respiratory tract infections	45 primary are physicians and advanced practice providers	Difference between control and intervention period percentages of visits with potentially low-value orders.	No change in in the percentage of visits with potentially low-value orders overall, for headaches or for acute sinusitis	N.S.
Lewis et al. (2019) UK (48)	Acute medical hospital	A controlled interrupted time series design.	The intervention comprised the addition of the message below to the bottom of all inpatient and outpatient paper and electronic CT reports. It was designed to highlight the type of patient who is most at risk after exposure to	CT scans are known to expose individuals to radiation.	n = ? providers	Immediate change in level or a gradual trend change in CT counts in electronic reports.	There was a significant reduction in CT scans.	p = 0.002

For peer review only

			ionising radiation and asks the provider if they informed the patient.					
Meeker et al., (2014) USA (49)	5 primary care clinics	RCT, randomization by clinician	Displaying poster-sized commitment letters in the clinicians' personal examination rooms for 12 weeks. These letters, featuring clinician photographs and signatures, stated their commitment to avoid inappropriate antibiotic prescribing for acute respiratory infections	Clinicians often inappropriately prescribe antibiotics for acute respiratory tract infections	14 clinicians (11 physicians and 3 nurse practitioners) (954 eligible adult patients)	Differences in antibiotic prescribing rates for antibiotic-inappropriate acute respiratory infection diagnoses in baseline and intervention periods.	Decrease in inappropriate antibiotic prescribing rate relative to control	$p<0.05$
Meeker et al. (2016) USA (50)	47 primary care practices in 2 different health systems	2 × 2 × 2 factorial RCT (Practices received 0, 1, 2, or 3 interventions)	1- Changes in EHR, Suggested alternatives presented electronic order sets suggesting nonantibiotic treatments 2- Changes in EHR, Accountable justification prompted clinicians to enter free-text justifications for	Clinicians often inappropriately prescribe antibiotics for acute respiratory tract infections	248 clinicians (14, 753 visits at baseline and 16, 959 during intervention period)	Changes in rates of inappropriate antibiotic prescribing behavior	1- No significant change in inappropriate antibiotic prescriptions 2- Decrease in inappropriate antibiotic prescriptions 3- Decrease in inappropriate antibiotic prescriptions	1 – NS; 2- $p<0.001$; 3- $p<0.001$. No statistically significant interactions between interventions .

For peer review only

			prescribing antibiotics into patients' electronic health records 3- Peer comparison emails sent to clinicians that compared their antibiotic prescribing rates with those of "top performers" (those with the lowest inappropriate prescribing rates).				No statistically significant interactions between interventions.	
Nguyen & Davis (2019) USA (51)	One multi-specialty academic medical center	Single center, prospective, quasi-experimental pre-post design	The interventions consisted of providing peer comparison reports describing the percentage of appropriately verified vancomycin orders based on the individual pharmacist. In intervention phase I, these reports were blinded. In intervention phase II, the reports were unblinded. Both intervention phases were compared with a	Order verification by a pharmacist serves as an important safeguard to prevent medication errors, as order verification errors have the potential to result in patient harm. Vancomycin is one of the most commonly utilized in hospitalized patients.	*n* = ? providers 1,625 vancomycin orders were included for evaluation (537 orders in the control group, 549 orders in intervention phase I, and 539 orders in intervention phase II).	Appropriate vancomycin dose order verification, defined as adherence to the institution's vancomycin dosing guidelines.	Unblinded peer comparison was associated with an increase in the percentage of appropriate vancomycin dose order verification. The percentage of appropriately verified vancomycin orders significantly improved in the unblinded intervention	$p < 0.001$

For peer review only

			pre-intervention control group.				phase II compared with the control group.	
O'Reilly-Shah et al. (2018) USA (52)	Department of Anesthesiology in a large health system (two academic hospitals, two private practice hospitals and two academic surgery centers)	Stepwise RCT, Randomization by facility (n = 5) of two interventions	1- Audit and feedback on provider level and department-level compliance with lung-protective ventilation (LPV) for attending physicians. 2- Audit and feedback for advance practice providers and residents 2- Changes to the EHR, default setting on anaesthesia machines for tidal volume was decreased from 700 mL to 400 mL.	Compliance with anesthesiology surgical quality metrics needs improvement.	*n = ?* providers (5 facilities, Total surgical case count (n) = 14, 793 Unique patients (n) = 12,785. 5 facilities.)	Rates of compliance with low tidal wave ventilation	Introduction of attending physician dashboards resulted in a 41% increase in the odds of compliance (OR 1.41, 95% CI 1.17 to 1.69). the addition of advanced practice provider and resident dashboards lead to an additional 93% increase in the odds of compliance (OR 1.93, 95% CI 1.52 to 2.46). Modifying ventilator defaults led to a 376% increase in the odds of compliance (OR 3.76, 95% CI 3.1 to 4.57).	1- p = 0.002 2- p<0.001 3- p<0.001

For peer review only

Olson et al. (2015) USA (53)	Clinical pathology, hematology, and oncology departments in a health system	Prospective Pre-post design (multiple baseline)	Changes in the EHR, default order sets dealing for posttransfusion hematocrits and platelet counts changed from "optional" to "preselected." The default settings for platelet count was later changed back to "optional"	Need to improve the monitoring of posttransfusion outcomes	> 500 residents and fellows. (7578 orders for red blood cell transfusion, 3285 total orders for platelet transfusion)	Rates of lab test ordering for post-transfusion counts after default change and post default change	Increase in hemocrit and platelet posttransfusio n count orders after default for order was set to "pre-selected" After switch back to "optional", significant decrease in orders	$p < 0.001$
Orloski et al. (2019) USA(54)	2 urban, academic emergency departments	prospective, controlled pre–post trial	Study interventions included (a) placement of branded folding seats and (b) an educational campaign. Only the intervention ED received folding seats.	Patient Satisfaction is important	2827 patients were surveyed	The primary outcome examined the influence of provider sitting on patient satisfaction. A secondary outcome examined the frequency of provider sitting.	Sitting at any point during an ED encounter improved responses to satisfaction questions for all measures). The odds of provider sitting increased 30% when a seat was placed in the room.	$p<0.0001$
Parrino (1989) USA (55)	One tertiary referral hospital	Prospective pre-post design	Monthly peer comparison letters sent to two cohorts (surgical and nonsurgical physicians) who are in the upper 50 percentiles of	Antibiotics are overutilized and expensive	202 clinically active physicians, surgical (n = 83) and nonsurgical (n = 119)	Changes in expenditures (total dollars) on antibiotics	No significant change in total dollars spent on antibiotics	N.S.

For peer review only

			physicians for antibiotic prescription expenditures					
Patel et al. (2014) USA(59)	One general internal medicine and one family medicine practice	Retrospective cross-sectional pre-post design	Changing the EHR default from displaying brand and generic medications to displaying initially only generics, with the ability to opt out.	Generic medications are less costly than brand medications	Internal and Family Medicine Attending physicians (IM, n = 38; FM, n = 17) and residents (IM, n = 166; FM, n = 34)	Monthly prescriptions of brand-name and generic equivalent beta-blockers, statins, and proton-pump inhibitors.	Increase in generic prescribing behavior	p<0.001
Patel et al. (2016) USA (57)	Three internal medicine practices	Prospective cross-sectional pre-post design (Differences in differences)	Changing the EHR, "Active choice" intervention in the EHR using a best practice alert directed to medical assistants and physicians—prompt to accept or cancel an order for a colonoscopy, mammography, or both. Physician needed to review and sign during the patient visit.	Guidelines to increase early cancer detection.	*n = ?* providers One intervention practice, 2 controls. 7560 patients eligible for colonoscopy with 14,546 clinic visits and 8,337 patients eligible for mammography with 14,410 clinic visits.	Percentage of patients eligible for screening who received a cancer screening order	Increase in mammograph y and colonoscopy orders	p<0.001
Patel et al. (2016) USA (58)	All specialties across a Health System	Pre-post design, difference-in-differences approach	Active Choice Nudge in the HER. Instead of changing EHR display defaults, an opt-out checkbox labeled "dispense as written" was	Using generic medications has been associated with higher adherence and improved clinical outcomes.	*N* = ? providers Pre-intervention data: (611 068 of 811 561 prescriptions) during the 10-month	Generic prescribing rates for 10 medical conditions i.e., 10 drugs	The overall generic prescribing rate increased significantly.	p < 0.001

			added to the prescription screen, and if left unchecked the generic-equivalent medication was prescribed.		preintervention period to 98.4% (644 587 of 655 011 prescriptions) during the 7-month postinterventio n period (P < .001)			
Patel et al. (2017) USA (56)	Three Internal Medicine practices	Prospective cross-sectional, pre-post design (Differences-in-differences)	Changing the EHR, "Active choice" intervention in the EHR using a best practice alert directed to medical assistants and physicians—prompt to accept or cancel an order for the flu vaccine. Physician needed to review and sign during the patient visit.	Center for Disease Control recommends influenza vaccination for everyone	*n = ?* providers (One intervention practice, 2 control practices. 45,926 patients)	Changes in flu vaccination rates	Increase in vaccination rates	p<0.001
Patel et al (2018) USA (60)	One health system, 32 primary care practices	3-arm Cluster randomized Clinical trial	1- "Active choice" and "accountable justification. "Physicians received an email indicating the number of his or her patients who met guidelines for statin therapy but had not been prescribed a statin. PCPs were	50% of patients who are eligible do not receive statins	96 PCPs (4774 patients not previously receiving statin therapy)	Percentage of eligible patients receiving statin prescription orders	1- Not a significant increase in statin prescription rates compared to usual care arm 2- Increase in statin prescription compared to usual care	1- NS; 2- p<0.01

For peer review only

			asked to make an active choice to prescribe atorvastatin, 20 mg, once daily, atorvastatin at another dose, or another statin or not prescribe a statin and select a reason. 2- Active choice + accountable justification + Peer comparison e-mails.					
Persell et al (2016) USA(61)	General internal medicine clinic	2 × 2 × 2 factorial RCT with 3 interventions	1- "Accountable Justification" Change in EHR settings where physicians receive an alert when inappropriately prescribing antibiotic 2- Suggested Alternatives, change in HER when physicians inappropriately prescribe antibiotics 3- Peer Comparison feedback	Clinicians frequently prescribe antibiotics inappropriately for acute respiratory infections	*n* = ? providers (3,276 visits in the pre-intervention year and 3,099 visits in the intervention year.)	Rate of oral antibiotic prescriptions for nonantibiotic appropriate acute respiratory infection diagnoses	No significant decrease in inappropriate prescribing rates	N.S.
Ryskina et al. (2018) USA(62)	Six general medicine teams in one health system	Single-blinded cluster randomized controlled	Peer comparison e-mails sent to physicians on general medicine	Routine laboratory tests for hospitalized patients are overused	6 attending physicians, 114 interns and residents	Number of routine laboratory orders placed by each	No significant changes in number of laboratory	N.S.

For peer review only

		trial. Randomization by 2-week service block.	teams, received a summary of their routine laboratory test ordering vs. the service average for the prior week			physician per patient day.	orders by each physician	
Sacarny et al (2018) USA (20)	Highest volume primary care prescribers of quetiapine in 2013 and 2014, whose patients have Medicare	RCT (intent to treat) placebo-controlled, parallel-group design with balanced randomization (1:1 ratio) to the control arm (placebo letter) and treatment arm (peer comparison letter).	Mailed peer comparison letters indicating that the prescriber's quetiapine prescribing was under review by CMS and was extremely high relative to the within-state peers. The text of the letter discussed that high quetiapine prescribing could be appropriate but was concerning for medically unjustified use.	Antipsychotic agents, such as quetiapine fumarate, are frequently overprescribed for indications not supported by clinical evidence, potentially causing harm.	5,055 PCPs, 231 (4.6%) were general practitioners, 2428 (48.0%) were in family medicine, and 2396 (47.4%) were in internal medicine.	Total quetiapine days supplied by prescribers from the intervention start to 9 months in intervention vs control.	Decrease in quetiapine days per prescriber in treatment vs control arm	p<0.001
Sedrak et al (2017) USA (63)	Three hospitals in one health system	RCT comparing a 1-year intervention to a 1-year preintervention period, and adjusting for time trends and patient characteristics.	Intervention lab tests that displayed Medicare allowable fees at the time of order and control lab tests did not.	It is estimated that nearly 30% of laboratory testing in the United States may be wasteful. Many health systems are considering increasing price transparency at	*n = ?* providers 60 diagnostic laboratory tests, 30 that are the most frequently ordered and 30 are the most expensive. 142, 921 hospital admissions	The number of tests ordered per patient-day. Secondary outcomes were tests performed per patient-day and Medicare associated fees.	No significant changes in number of tests ordered between intervention and control group	N.S.

		Randomization at test-level		the time of order entry.	representing 98,529 patients			
Sharma et al. (2019) USA (64)	One Health System.	Stepped-wedge cluster randomized clinical trial	A default imaging order in the electronic health record (EHR) to reduce unnecessary daily imaging during palliative radiotherapy. The intervention introduced a default imaging order in the EHR that specified no daily imaging during palliative radiotherapy. Physicians could opt out, selecting another imaging frequency	Daily imaging, using radiography or computed tomography, can augment positioning. Although daily imaging is often used for curative radiotherapy, national guidelines consider it unnecessary for palliative radiotherapy. Unnecessary imaging can increase treatment time and expense for patients in distress.	21 radiation oncologists, 1019 patients who received 1188 palliative radiotherapy courses (n = 747 at the university practice; n = 441 at the community based practices) to bone (52.2%), soft tissue (19.9%), brain (15.7%), or multiple sites (12.3%)	The primary outcome was a binary indicator of radiotherapy courses with daily imaging (defined as imaging during ≥80% of treatments).	The default intervention led to a significant reduction in daily imaging.	p=0.004
Shively et al. (2020) USA (65)	Veterans' Affairs Health System (7 primary care practices)	Prospective pre-post design	An education session, followed by monthly e-mails with their antibiotic prescribing rate, peer prescribing rates, and a system target. (Peer comparison feedback)	Reducing inappropriate outpatient antibiotic use is an important national goal.	Baseline = 65 primary care professionals (PCPs) caring for 40,734 patients in the VA Pittsburgh health care system (VAPHS). During the intervention	Mean rate of monthly antibiotic prescribing rates	The mean rate of monthly antibiotic prescriptions declined. Among reviewed cases, unnecessary antibiotic prescribing declined and	P<0.001

For peer review only

					period = 73 PCPs caring for 41,191 patients. There were 28,402 office visits during the baseline period and 32,982 office visits during the intervention period.		the rate of optimally prescribed antibiotics increased.	
Srinivasan et al. (2020) USA (66)	Inpatient units at a 350-bed, tertiary care freestanding children's hospital	Prospective Pre-post design	Interventions included electronic medical record triggers, provider education, and peer comparison.	American Academy of Pediatrics guidelines recommend annual influenza vaccination for all children 6 months and older	*N = ?* providers A total of 6089 children 6 months and older were admitted to the medical and surgical units during the baseline period, and 6206 were admitted during the QI initiative	The primary outcome measure was the percentage of children discharged from the hospital with at least 1 dose of the influenza vaccine received at either the hospital or before admission	There was a significant increase in the percentage of hospitalized children discharged with at least 1 dose of the vaccine received at either the hospital or before admission during the QI initiative.	p<0.001
Suffoletto & Landeau (2019) USA (67)	Emergency departments in Single hospital system, 16 hospitals	A pilot randomized controlled trial	Audit and feedback intervention vs peer norm comparison intervention emails.	Opioid epidemic; reducing opioid prescriptions	37 emergency medicine providers	Mean prescribing rates of opioids.	The mean reduction in opioid prescriptions was significant for audit and feedback, and for peer norm comparison.	p<0.001

For peer review only

Szilagyi et al. (2014) USA (68)	Practices in two large research networks	RCT, randomization unit by practices in two practice-based research networks	EHR alert at all office visits, indicating the specific recommended immunizations. Staff prompts in the form of a reminder sheet was placed on the provider's desk in the exam room indicating the vaccines due.	Adolescent immunization rates are suboptimal	*n* = ? providers Two networks. One network: 5 intervention, 5 control practices; One network: 6 intervention, 6 control practices	Changes in adolescent immunization rates by study end	Immunization rates at the study end for adolescents who were behind on immunization s at study initiation were not significantly different for intervention versus control practices for any vaccine or combination of vaccines	N.S.
Trent et al. (2018) USA (69)	Health Medical Center, an urban, safety net, Level 1 trauma center	Stepped wedge design and cluster randomization of physicians.	Monthly audit and feedback with blinded peer comparison on guideline adherence for pneumonia and severe sepsis. All physicians in that cluster received an email detailing their adherence to pneumonia and severe sepsis guidelines for every month since the start of the study.	Emergency physician adherence to guidelines for appropriate antibiotic administration for inpatient pneumonia	*n* = ? providers 469 patients were seen during the study period	Guideline-concordant antibiotic selection was determined if the emergency physician administered the appropriate antibiotics, given the patient's risk for multidrug-resistant organisms in accordance with the study institution's specific antibiotic guideline for inpatient pneumonia,	Overall, feedback significantly improved adherence to antibiotic guidelines.	$p<0.05$

						which was readily available electronically both on an antibiotic stewardship smartphone application as well as on the institution's intranet.		
Wigder et al (1999) USA (70)	Emergency department in a suburban tertiary care facility	Prospective, pre-post design	1- Information campaign of Ottawa rule. 2- ED physicians shown their baseline data. 3- Audit and Feedback. Copies of ED charts of knee injury patients were placed in physician mailboxes complimenting them for adherence to the Ottawa knee rule or telling them they didn't adhere to it.	There is an excessive use of X-rays when guidelines for evaluating are less invasive, less costly, and are shown to be effective.	27 ED physicians	Changes in X-rays ordered, Number of Percentage Abnormal Results.	Decrease in number of X-ray studies, increase in number of abnormal X-rays	$p<0.001$
Winickoff et al. (1984) USA (71)	Department of Internal Medicine at one group practice	3 Interventions: Pre-post design for first two interventions. 3rd intervention:	1 - Educational meeting for clinical standard 2 – Peer comparison, meeting at which the rate of group compliance with	Many clinicians do not adhere to guidelines for colorectal screening.	*n = ?* for first two interventions 16 physicians for RCT	Rate of performance of a stool test for blood	1 – No change in stool tests ordered 2- No change in stool tests ordered 3- Increase in number of	1- N.S. 2- N.S. 3- $p <0.001$

For peer review only

		RCT using a crossover design (over a 12 month period, crossover at 6 months)	the standard before and after the educational meeting was presented 3 – Peer Comparison Feedback, monthly feedback about how physicians compare to peers at practice.				stool tests ordered	
Zivin et al (2019) USA (72)	Two health systems	Prospective, pre-post design	Changing the EHR default for all Schedule II opioid prescriptions to a 15-pill count.	Reliance on prescription opioids for postprocedural analgesia has contributed to the opioid epidemic	448 prescribers 6,390 opioid prescriptions	Changes in the proportion of opioid prescriptions for 15 pills	Increase in the proportion of opioid prescriptions for 15 pills increased at both sites	$p<0.001$
Zwank et al. (2017) USA (73)	Emergency department of a Level 1 Trauma Center	Retrospective pre-post design	Changing the EHR, modifying default number of opioid prescriptions from 15 tablets to a number the physician had to enter themselves	Reliance on prescription opioids for postprocedural analgesia has contributed to the opioid epidemic	*n = ?* providers (7,019 eligible prescriptions)	Changes in the total opioid pill quantity per prescription	No significant change in mean number of opioid tablets dispensed by prescription	N.S.

Note. N.S. denotes a non-significant finding.

For peer review only

Table 3. Studies Organized According to Nudge Ladder

Nudge Ladder	Study	Significant Effect in the Hypothesized Direction?	Majority in Category Significant?
Provide Information	Meeker et al. (2016) USA (50) —Arm 1	N.S.	No
	Persell et al. (2016) USA (61) — Arm 2	N.S.	
	Sedrak et al (2017) USA (63)	N.S.	
	Szilagyi et al. (2014) USA (68)	N.S.	
Frame Information	Allen, Dunn, & Bush (2019) USA (36)	p<0.001	Yes
	Andereck et al. (2019) USA (37)	p<0.001	
	Buntinx et al. (1993) Belgium (40)	p>0.05	
	Hemkens et al. (2017) Switzerland (43)	N.S.	
	Hempel et al. (2014) USA (44)	p<0.05	
	Lewis et al. (2019) UK (48)	p = 0.002	
	Meeker et al. (2016) USA (50) – Arm 2	p<0.001	
	Meeker et al. (2016) USA (50) – Arm 3	p<0.001	
	Nguyen & Davis (2019) USA (51)	p < 0.001	
	O'Reilly-Shah et al. (2018) (52)— Arm 1	P = 0.002	
	O'Reilly-Shah et al. (2018) (52) — Arm 2	P <0.001	
	Parrino (1989) (55) USA	N.S.	
	Persell et al. (2016) USA (61) — Arm 1	N.S.	
	Persell et al. (2016) USA (61) — Arm 3	N.S.	
	Ryskina et al. (2018) USA (62)	N.S.	
	Sacarny et al (2018) USA (20)	p<0.001	
	Shively et al. (2020), USA (65)	P<0.001	
	Suffoletto & Landeau (2019) USA (67)	p<0.001	
	Trent et al. (2018), USA (69)	p<0.05	
	Winickoff et al. (1984) USA (71) — Study 1	N.S.	
	Winickoff et al. (1984) USA (71) — Study 2	N.S.	
	Winickoff et al. (1984) USA (71) — Study 3	p <0.001	
Prompt Implementation Intentions	Kullgren et al. (2018) USA	N.S.	No
	Meeker et al. (2014) USA (49)	p <0.05	
Enable Choice	Bourdeaux et al. (2014) UK (39)	p<0.001 for both	Yes
	Hsiang et al. (2019) USA (45)	<0.001	
	Kim et al. (2018) USA (46)	p<0.001	
	Orloski et al. (2019) USA (54)	p<0.0001	
	Patel et al. (2016) USA (57)	p<0.001	
	Patel et al. (2016) USA (58)	p < 0.001	
	Patel et al. (2017) USA (56)	p<0.001	
	Patel et al. (2018) USA (60) — Arm 1	N.S.	

For peer review only

	Zwank et al. (2017) USA (73)	N.S.	
Guide choice through default options	Chiu et al. (2018) USA (41)	p<0.01	Yes
	Delgado et al. (2018) USA (42)	p<0.001	
	Olson et al. (2015) USA (53)	p < 0.001	
	Patel et al. (2014) USA (59)	p<0.001	
	Sharma et al. (2019) USA (64)	p=0.004	
	Zivin et al. (2019) USA (72)	p<0.001	

Note. Articles that included multiple intervention treatment groups, studies, or study arms are described.

For peer review only

Table 4. Multicomponent Intervention Studies Organized According to Nudge Ladder

Nudge Ladder	**Study**	**Significant Effect in the Hypothesized Direction?**
Provide information + Guide choice through default options	Arora et al. (2019) USA (38)	P < 0.001
Provide Information + Frame Information	Wigder et al. (1999) USA (70)	p<0.001
Enable Choice + Frame Information	Patel et al. (2018) USA (60)— Arm 2	P < 0.001
Frame Information + Guide choice through default options	O'Reilly-Shah et al. (2018) USA (52) — Arm 3	P < 0.001
Provide information + Frame Information + Enable choice	Srinivasan et al. (2020) USA (66)	P < 0.001

For peer review only

Table 5. Cochrane Risk of Bias Assessment Tool.

Authors (Year, Country)	Random Sequence Generation	Allocation Concealment	Blinding (participants and personnel)	Blinding Outcome Assesors	Incomplete Outcome Data	Selective Reporting
Allen, Dunn, & Bush (2019) USA (36)	−	−	−	?	+	?
Andereck et al. (2019) USA (37)	−	−	−	?	+	?
Arora et al. (2019) USA (38)	−	−	−	?	−	?
Bourdeaux et al. (2014) UK (39)	−	−	−	?	+	?
Buntinx et al. (1993) Belgium (40)	+	+	+	+	+	?
Chiu et al. (2018) USA (41)	−	−	−	?	+	+
Delgado et al. (2018) USA (42)	−	−	−	?	+	?
Hemkens et al. (2017) Switzerland (43)	+	+	+	+	+	+
Hempel et al. (2014) USA (44)	−	−	−	?	?	?
Hsiang et al. (2019) USA (45)	−	−	−	?	+	?
Kim et al. (2018) USA (46)	−	−	−	?	+	+
Kullgren et al. (2018) USA (47)	+	+	+	+	+	+
Lewis et al. (2019) UK (48)	−	−	−	?	?	+
Meeker et al., (2014) USA (49)	+	+	+	+	+	+
Meeker et al. (2016) USA (50)	+	+	+	+	+	+
Nguyen & Davis (2019) USA (51)	−	−	−	?	?	?
O'Reilly-Shah et al. (2018) USA (52)	−	−	−	?	+	+
Olson et al. (2015) USA (53)	−	−	−	?	+	?

For peer review only

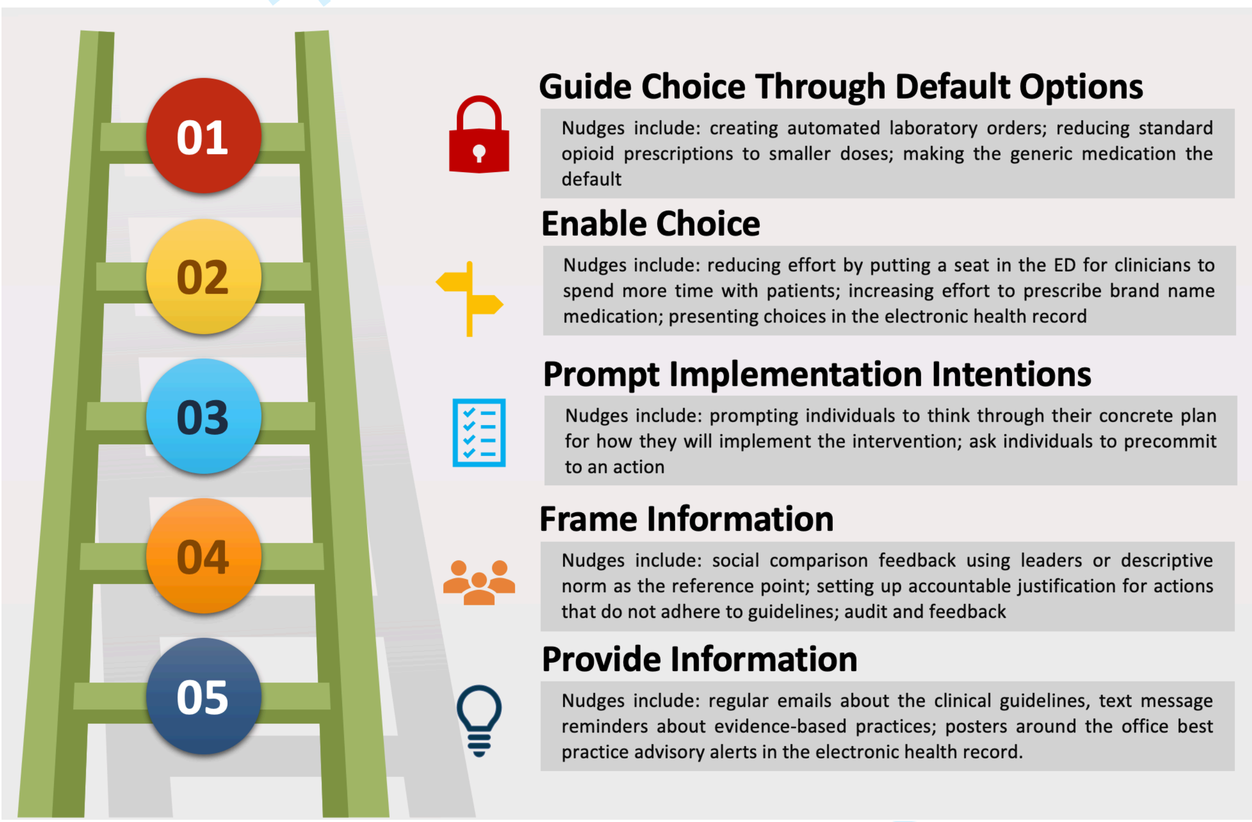
Orloski et al. (2019) USA(54)	−	−	−	?	?	?
Parrino (1989) USA (55)	−	−	−	?	+	?
Patel et al. (2014) USA(59)	−	−	−	?	+	?
Patel et al. (2016) USA (57)	−	−	−	?	+	?
Patel et al. (2016) USA (58)	−	−	−	?	+	?
Patel et al. (2017) USA (56)	−	−	−	?	+	?
Patel et al (2018) USA (60)	+	+	+	+	+	+
Persell et al (2016) USA(61)	+	+	+	+	+	+
Ryskina et al. (2018) USA(62)	+	+	+	+	+	+
Sacarny et al (2018) USA (20)	+	+	+	+	+	+
Sedrak et al (2017) USA (63)	+	+	+	+	+	+
Sharma et al. (2019) USA (64)	+	+	+	+	+	+
Shively et al. (2020) USA (65)	−	−	−	?	?	?
Srinivasan et al. (2020) USA (66)	−	−	−	?	?	?
Suffoletto & Landeau (2019) USA (67)	+	+	+	+	+	+
Szilagyi et al. (2014) USA (68)	+	+	+	+	+	+
Trent et al. (2018) USA (69)	+	+	+	+	+	+
Wigder et al (1999) USA (70)	−	−	−	?	?	?

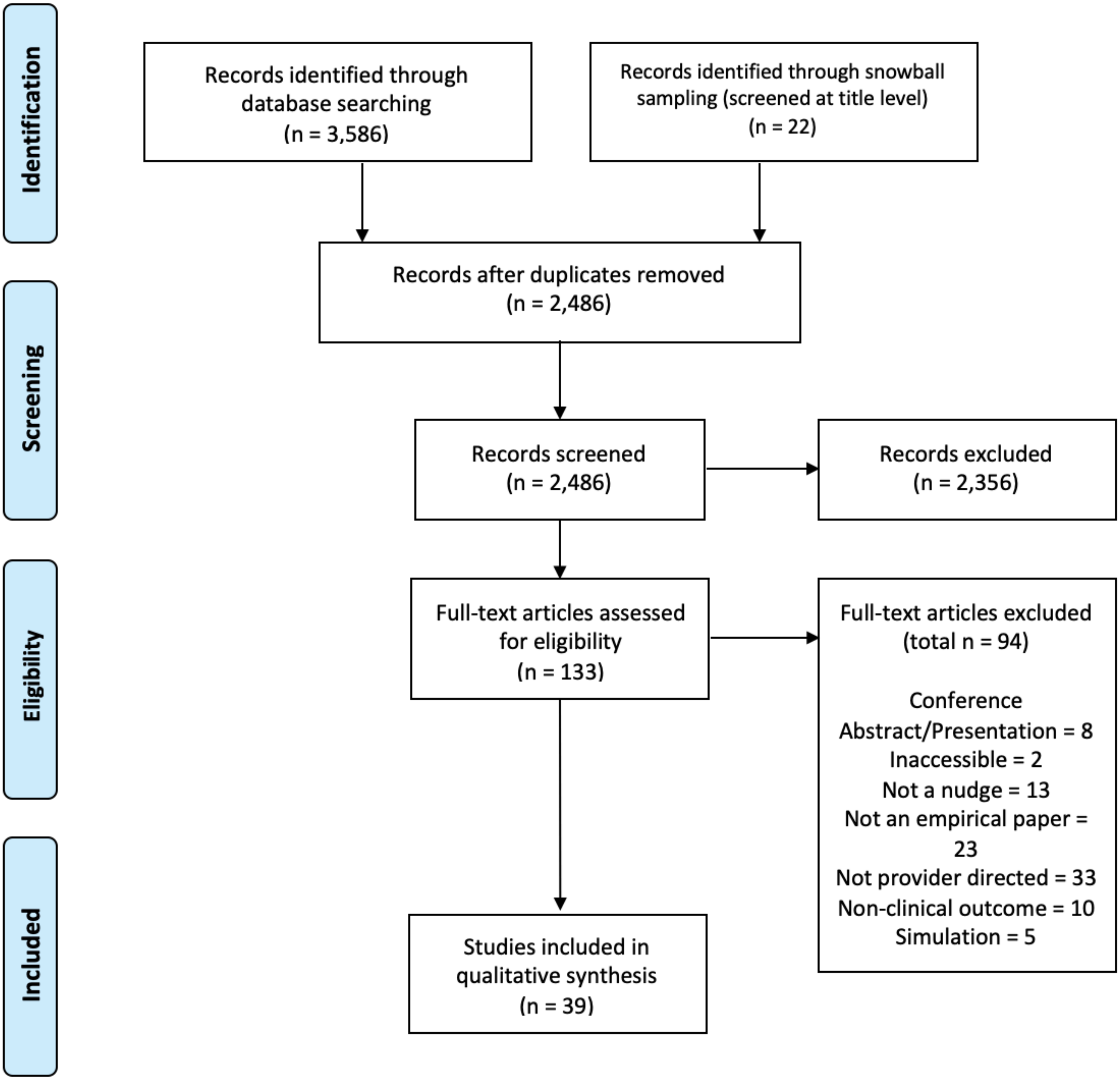
For peer review only

Winickoff et al. (1984) USA (71)	First 2 Studies: − , 3rd Study: +	First 2 Studies: − Third Study: +	First 2 Studies: − , 3rd Study: +	?	+	+
Zivin et al (2019) USA (72)	−	−	−	?	+	?
Zwank et al. (2017) USA (73)	−	−	−	?	+	?

Note. ? indicates unclear risk of bias, + indicates low risk of bias, and − indicates high risk of bias. See (72) for a full description of the Cochrane Risk of Bias tool.

For peer review only


01
02
03
04
05
Guide Choice Through Default Options
Nudges include: creating automated laboratory orders; reducing standard opioid prescriptions to smaller doses; making the generic medication the default
Enable Choice
Nudges include: reducing effort by putting a seat in the ED for clinicians to spend more time with patients; increasing effort to prescribe brand name medication; presenting choices in the electronic health record
Prompt Implementation Intentions
Nudges include: prompting individuals to think through their concrete plan for how they will implement the intervention; ask individuals to precommit to an action
Frame Information
Nudges include: social comparison feedback using leaders or descriptive norm as the reference point; setting up accountable justification for actions that do not adhere to guidelines; audit and feedback
Provide Information
Nudges include: regular emails about the clinical guidelines, text message reminders about evidence-based practices; posters around the office best practice advisory alerts in the electronic health record.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 3,586)
Records identified through snowball sampling (screened at title level) (n = 22)
Records after duplicates removed (n = 2,486)
Records screened (n = 2,486)
Records excluded (n = 2,356)
Full-text articles assessed for eligibility (n = 133)
Full-text articles excluded (total n = 94)
Conference Abstract/Presentation = 8
Inaccessible = 2
Not a nudge = 13
Not an empirical paper = 23
Not provider directed = 33
Non-clinical outcome = 10
Simulation = 5
Studies included in qualitative synthesis (n = 39)

Appendix A

Systematic Search Strategy

The methodology for the search was designed based on standards for systematic reviews (32), in consultation with a medical librarian, as well as with two experts from the field of healthcare behavioral economics. The databases used were: EconLit, Embase, EBSCO Megafile, PsycINFO, PubMed, Scopus, and Web of Science.

Search terms included combinations, plurals, and various conjugations of the words relating to identified nudge strategies. The search string and strategy from (6) was used as a basis for search terms, but adjusted to reflect the more specific clinician-directed aim of this research question. All peer-reviewed empirical studies published prior to the completion of our search phase (i.e., – 4/2020) were eligible for this review.

Following retrieval of all records, duplicates were removed using Zotero (www.zotero.org), and via manual inspection. Article screening involved two stages. First, all records were screened at the title and abstract level by a team of four coders (the first-author and three research assistants) using the web-based application for systematic reviews, Rayyan (https://rayyan.qcri.org). Criteria in this first-pass screening were more inclusive—that is, all interventions directed at clinicians were included and examined further. To establish reliability, the first-author and the three coders screened the same 20 articles and then reviewed their screening decisions together. Any disagreements were resolved by consensus. This process was repeated three additional times until 80 articles were screened by all four coders and sufficient reliability was established. Reliability was excellent (fleiss' κ = 0.96). For the remainder of the screening process, screening was done independently by all four coders; the team met weekly to discuss any articles that they were uncertain about including or excluding. This screening process

For peer review only

was followed by a full text examination to finally determine inclusion or exclusion according to more stringent inclusion and exclusion criteria (see Table 1). This screening process was done as a team and determinations of article inclusion were decided collaboratively.

Search Terms

The following search terms were used in the systematic search. All searches were conducted in the title field.

EBSCO Megafile

TI (nudg* OR choice architect OR choice architecture OR behavioral intervention OR behavioural intervention OR behavioral economic OR behavioral economics OR behavioral insight OR behavioural insight OR active choice OR default OR default bias OR default option OR opt-out OR opt-in OR prompted choice OR commitment device OR accountable justification OR peer comparison OR pre-commitment) AND TI (physician OR health OR clinician OR clinic OR provider* OR electronic health record OR health record OR doctor OR nurse OR physician assistant OR medical assistant OR electronic medical record OR medical record OR medical OR outpatient OR inpatient OR hospital OR resident)

EconLit

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND TI (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or

For peer review only

electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Embase

(nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

PsycInfo

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND TI (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

PubMed

TI (nudg* OR choice architect OR choice architecture OR behavioral intervention OR behavioural intervention OR behavioral economic OR behavioral economics OR behavioral insight OR behavioural insight OR active choice OR default OR default bias OR default option OR opt-out OR opt-in OR prompted choice OR commitment device OR accountable justification OR peer comparison OR pre-commitment) AND TI(physician OR health OR clinician OR clinic OR provider* OR electronic health record OR health record OR doctor OR nurse OR physician assistant OR medical assistant OR electronic medical record OR medical record OR medical OR outpatient OR inpatient OR hospital OR resident)

Scopus

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Web of Science

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or

For peer review only

electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Table A1. Search Dates and Yields

Database	**Date**	**Yield**
EBSCO Megafile	4/22/2020	482
EconLit	4/22/2020	28
Embase	4/22/2020	1,240
PsycInfo	4/22/2020	384
PubMed	4/22/2020	292
Scopus	4/22/2020	30
Web of Science	4/22/2020	1,130
Total		3,586

For peer review only

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	4-5
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	7-9
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	9
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	10
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	10
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	12
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplemental A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	12-13
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	12
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	13
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	14
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	14
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	15-16

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	14
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	15-16
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	15-16
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	17
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	36-55
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	17-18
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	17
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-19
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	20-21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22-24
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	3

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A systematic review of clinician-directed nudges in healthcare contexts

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-048801.R1
Article Type:	Original research
Date Submitted by the Author:	22-Apr-2021

Complete List of Authors:	Last, Briana; University of Pennsylvania, Psychology Buttenheim, AM; University of Pennsylvania School of Nursing, Department of Family and Community Health; University of Pennsylvania, Center for Health Incentives and Behavioral Economics (CHIBE) Timon, Carter; University of Pennsylvania, College of Liberal and Professional Studies Mitra, Nandita; University of Pennsylvania Perelman School of Medicine, Department of Biostatistics, Epidemiology & Informatics Beidas, Rinad; University of Pennsylvania Perelman School of Medicine, Department of Psychiatry; University of Pennsylvania, Penn Implementation Science Center at the Leonard Davis Institute of Health Economics (PISCE@LDI)

Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health economics
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

A systematic review of clinician-directed nudges in healthcare contexts

Briana S. Last[1], Alison M. Buttenheim,[2,3,4] Carter E. Timon,[5] Nandita Mitra[6], & Rinad S. Beidas[3,4,7-9]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA

[2]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA

[3]Center for Health Incentives and Behavioral Economics (CHIBE), University of Pennsylvania, Philadelphia, PA

[4]Penn Implementation Science Center at the Leonard Davis Institute of Health Economics (PISCE@LDI), University of Pennsylvania, Philadelphia, PA

[5]College of Liberal and Professional Studies, University of Pennsylvania, Philadelphia, PA

[6] Department of Biostatistics, Epidemiology & Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[7]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

For peer review only

[8]Department of Medical Ethics and Health Policy, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

[9]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

For peer review only

Author Note

The authors would like to thank Mitesh Patel, Anne Larrivee, Melanie Cedrone, Pamela Navrot, and Amarachi Nasa-Okolie for their assistance in the project.

Correspondence concerning this article should be addressed to Briana S. Last, Department of Psychology, Stephen A. Levin Building, 425 S University Ave, Philadelphia, PA 19104.

Contact: brishiri@sas.upenn.edu

For peer review only

Abstract

Objective: Nudges are interventions that alter the way options are presented, enabling individuals to more easily select the optimal option. Health systems and researchers have tested nudges to shape clinician decision-making with the aim of improving healthcare service delivery. We aimed to systematically study the use and effectiveness of nudges designed to improve clinicians' decisions in healthcare settings.

Design: A systematic review was conducted to collect and consolidate results from studies testing nudges and to determine whether nudges directed at improving clinical decisions in healthcare settings across clinician types were effective. We systematically searched seven databases (EBSCO Megafile, EconLit, Embase, PsycInfo, PubMed, Scopus and Web of Science) and used a snowball sampling technique to identify peer-reviewed published studies available between 1 January 1989 and 22 April 2020. Eligible studies were critically appraised and narratively synthesized. We categorized nudges according to a taxonomy derived from the Nuffield Council on Bioethics. Included studies were appraised using the Cochrane Risk of Bias Assessment Tool.

Results: We screened 3,608 studies and 39 studies met our criteria. The majority of studies (90%) were conducted in the United States and 36% were randomized controlled trials. The most commonly studied nudge intervention (46%) framed information for clinicians, often through peer comparison feedback. Nudges that guided clinical decisions through default options or by enabling choice were also frequently studied (31%). Information framing, default, and enabling choice nudges were effective, whereas the effectiveness of other nudge types was mixed. Given the inclusion of non-experimental designs, only a small portion of studies were at minimal risk of bias (33%) across all Cochrane criteria.

For peer review only

Conclusions: Nudges that frame information, change default options, or enable choice are frequently studied and show promise in improving clinical decision-making. Future work should examine how nudges compare to non-nudge interventions (e.g., policy interventions) in improving healthcare.

For peer review only

Strengths and limitations of this study

- This systematic review synthesizes the growing research applying nudges in healthcare contexts to improve clinical decision-making.
- The review uses both systematic search strategies and a snowball sampling approach, the latter of which is useful for identifying relatively novel literature.
- Meta-analysis was not possible due to heterogeneity in methods and outcomes.
- The systematic review was not designed to synthesize research wherein study authors did not identify the intervention as a nudge.

For peer review only

A systematic review of clinician-directed nudges in healthcare contexts

Rationale

Research from economics, cognitive science, and social psychology have converged on the finding that human rationality is "bounded" [1]. The intractability of certain decision problems, constraints on human cognition, and scarcity of time and resources lead individuals to employ mental shortcuts to make decisions. These mental shortcuts, often called heuristics, are strategies that overlook certain information in a problem with the goal of making decisions more quickly than more deliberative methods [2]. While heuristics can often be more accurate than more complex mental strategies, they can also lead to errors and suboptimal decisions [2,3]. Researchers have discovered interventions to harness the predictable ways in which human judgment is biased to improve decisions. These interventions, known as "nudges," reshape the "choice architecture," or the way options are presented to decision-makers to optimize choices [4]. Nudges have been applied to retirement savings, organ donation, consumer health and wellness, and climate catastrophe mitigation demonstrating robust effects [5–8].

As with retirement savings and dietary choices, clinical decision-making—clinicians' process of determining the best strategy to prevent and intervene on clinical matters—is complex and error-prone. Clinicians often use heuristics when making diagnostic and treatment decisions [9–11]. For example, clinicians are influenced by whether treatment outcomes are framed as losses or gains (e.g., doctors prefer a riskier treatment when the outcome is framed in terms of lives lost rather than lives saved) [12]. Heuristics can lead to medical errors [13]. In the face of complex medical decisions, clinicians tend to choose the default treatment option (despite clinical guidelines) or conduct clinical examinations that confirm their priors [14,15].

For peer review only

Choice architecture influences clinicians' behavior regardless of whether clinicians are conscious of it, creating opportunities for nudges [16]. Clinical decisions are increasingly made within digital environments such as electronic health record (EHR) systems [17]. More than 90% of US hospitals now use an EHR [18,19]. Researchers have explored the potential to use these ubiquitous electronic support systems to shape clinical decisions through nudges. They have subtly modified the EHR choice architecture by changing the default options for opioid prescription quantities or by requiring physicians to provide free-text justifications for antibiotic prescriptions [16]. Even when nudges are not implemented in the EHR, researchers extract aggregate data from the EHR, suggesting its increasing role in the study of clinical decision-making [20].

As health systems and researchers have embraced nudges in recent years, there is growing interest in understanding which nudges are most effective to improve clinical decision-making. Taxonomizing nudges is advantageous because many nudges explicitly target heuristics, revealing the mechanism of behavior change [21]. If nudges that leverage people's tendency to adhere to social norms are consistently more effective than nudges that exploit clinicians' default bias, then future nudges can be designed with this insight. Two systematic reviews were recently conducted to evaluate the effectiveness of healthcare nudges. Though both reviews demonstrate promise for the effectiveness of nudges, they offer somewhat conflicting evidence on the most studied and most effective nudge types, suggesting that an additional review may be useful [22,23]. Our review offers complementary and non-overlapping insights on the study of nudges in healthcare settings for the following reasons: (1) we do not exclusively study physicians as our target population as in [23], instead we include all healthcare workers; and (2) we do not restrict

For peer review only

our research to randomized controlled trials reported in the Cochrane Library of systematic reviews [22].

Our review also makes use of a nudge taxonomy derived from the widely cited Nuffield Council on Bioethics intervention ladder wherein interventions increase in potency and constrain choice with each new rung [24,25]. Interventions on the bottom of the ladder tend to be more passive, offering decision makers information and reminders. Interventions in the middle of the ladder leverage psychological insights to motivate decision-makers either through social influence or by encouraging planning. At the top of the ladder, interventions are more assertive and reduce decisions to a limited set of choices or by creating default options. The Nudge Ladder categorizes nudges by the psychological mechanisms by which they operate, the degree to which they maintain autonomy, and have the additional advantage of aligning with existing public health and quality improvement literature that make use of the Nuffield Council ladder [4,26]. The Nudge Ladder offers insights on the heuristics most relevant to the clinical decision-making process and can support health systems in selecting and applying nudges to improve clinical decision-making.

Objective

We systematically evaluated nudge interventions directed at clinicians in healthcare settings to determine the types of nudges that are most studied and most effective in improving clinical decision-making compared with other nudges, non-nudge interventions, or usual care. All quantitative study designs were included in our review.

Methods

Protocol and Registration

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Before initiating this review, we searched the international database PROSPERO to avoid duplication. After establishing that no such review was underway, we prospectively registered our review (https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=123349).

Eligibility Criteria

Types of Participants

We included only empirical studies published in peer-reviewed journals studying nudges directed at clinicians working in healthcare settings. Clinicians were defined as workers who provide healthcare to patients in a hospital, skilled nursing facility, or clinic. Examples of clinicians include physicians, nurses, medical assistants, physician assistants, clinical psychologists, clinical social workers, and lay health workers. Studies that exclusively nudged patients were not included.

Types of Intervention

Nudges were defined as “any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives” [4]. Alterations to choice architecture included changes to the information provided to the clinician (e.g., translating information, displaying information, presenting social benchmarks), altering the decision structure of the provider (e.g., modifying default options, changing choice-related effort, changing the number or types of options, or changing decision consequences) and providing decision aids (e.g., offering reminders or commitment devices) [27]. The study authors did not need to identify the intervention as a nudge to be considered for study inclusion, however given the systematic search string, which includes several behavioral economics terms (see Appendix A), studies that did not self-identify as behavioral economic interventions were unlikely to be included.

For peer review only

Interventions that required sustained education or training were not considered nudges. No options could be forbidden and there could be no financial incentives [28]. Though some financial incentives for clinicians may be considered nudges, most studies on financial incentives for clinicians involve significant compensation or "pay for performance"—of which there is already an existing literature [29].

Nudges guided clinicians to make improved clinical decisions, including (but not limited to) increasing the uptake of evidence-based practices (EBPs), adherence to health system or policy guidelines, and reducing healthcare service costs. EBPs refer to clinical techniques and interventions that integrate the best available research evidence, clinical expertise, and patient preferences and characteristics [30]. Study authors had to provide the evidentiary rationale for the nudge.

We did not include studies that analyzed the sustainability of nudges in the same study setting and/or sample of providers. In order to analyze studies with independent samples, we included the primary paper and not follow-up papers.

Types of Studies

All study designs were included that had a control or baseline comparator—the control or baseline could be usual care or another intervention (nudge or non-nudge). For studies with parallel intervention groups, we did not require that allocation of interventions be randomized (i.e., quasi-experimental studies were included). Exclusively qualitative studies were not included. See Table 1 for Eligibility Criteria.

Search

Snowball Sampling

For peer review only

The initial search strategy was based on a snowball sampling method [31] using the references from a published commentary on the uses of nudges in healthcare contexts [16]. Reviews identified during the preliminary stage of the systematic search process were also used to snowball articles, though these largely resulted in duplicates. Articles were reviewed at the title level to immediately identify those to be excluded. Those tentatively included were reviewed at the abstract level, followed by the full text for those meeting criteria. Following completion of screening of records retrieved via snowball, a systematic search of several databases was completed.

Information Sources & Systematic Search

The methodology for the search was designed based on standards for systematic reviews [32], in consultation with a medical librarian, as well as with two experts from the field of healthcare behavioral economics. The databases used were: EconLit, Embase, EBSCO Megafile, PsycINFO, PubMed, Scopus, and Web of Science.

Search terms included combinations, plurals, and various conjugations of the words relating to identified nudge interventions. The search string and strategy from [6] was used as a basis for search terms, but adjusted to reflect our research question (see Table 1). All peer-reviewed empirical studies published prior to the completion of our search phase (i.e., – 4/2020) were eligible for this review. See Appendix A for the search strings.

Data Collection Process

Following retrieval of all records, duplicates were removed using Zotero (www.zotero.org) and via manual inspection. Article screening involved two stages. First, all records were screened at the title and abstract level by a team of four coders (BSL, CET, and two research assistants) using the web-based application for systematic reviews, Rayyan

For peer review only

(https://rayyan.qcri.org). Criteria in this first-pass screening were inclusive—that is, all interventions directed at clinicians were included. To establish reliability, the coders screened the same 20 articles and then reviewed their screening decisions together. Any disagreements were resolved by consensus. This process was repeated three additional times until 80 articles were screened by all four coders and sufficient reliability was established. Reliability was excellent (Fleiss' $\kappa = 0.96$). For the remainder of the screening process, screening was done independently by all four coders; the team met weekly to discuss edge cases. This screening process was followed by a full text examination to determine eligibility according to more stringent inclusion and exclusion criteria (see Table 1). This screening process was done as a team and determinations of article inclusion were decided by consensus.

Patient and Public Involvement

Patients and the public were not involved in the design, conduct, or reporting of this research.

Data Items

Study characteristics and outcomes were extracted and tabulated systematically per recommendations for systematic reviews [32]. These data included: (1) study characteristics — author names, healthcare setting, study design, country, date of publication, details of the intervention, justification for the nudge, sample size, primary outcomes, main findings, and whether the effect was statistically significant; (2) nudge type; and (3) risk of bias assessment.

BSL and RSB trained the coding team (four Master's students in a Behavioral and Decision Sciences program) in data extraction. The team coded articles (n=16) together to ensure consensus. RSB reviewed a random sample (n=5) of the final articles to ensure reliability with systematic review reporting standards. BSL subsequently coded the remaining articles (n=18).

For peer review only

Outcomes

We only included studies that included objective measures of clinician behavior in real healthcare contexts. Studies that measured clinicians' choices in vignette or simulation studies were not included. Results could be presented as either continuous (e.g., number of opioid pills prescribed) or binary (e.g., whether physicians ordered influenza vaccinations). Outcomes were measured either directly (e.g., antibiotic prescribing rates) or indirectly (e.g., using cost to estimate changes in antibiotic prescriptions). Participants could not report on their own behavior because clinicians' self-report can be inaccurate [33]. Both absolute measurements and change relative to baseline were accepted.

Risk of Bias in Individual Studies

We evaluated whether the studies included in the systematic review were at risk for bias, using the Cochrane Risk of Bias Tool [32,34]. BSL trained CET and they assessed articles (n=2) together to ensure consensus. CET independently coded (n =12) articles and BSL coded the remaining articles (n = 27). The team met weekly to discuss articles that they were uncertain about and resolved discrepancies by consensus.

Data Synthesis

In order to examine which types of nudges were most studied and most effective, we calculated the number and percentage of studies using each nudge intervention according to the Nudge Ladder (see Figure 1). We reported the effect and statistical significance of the effect when a primary outcome was clearly identified in the study. If no primary outcome was identified by study authors, we determined a primary outcome based on the main research question. For studies that reported multi-component nudges—i.e., interventions that combine several nudges together—we reported the total effect of the intervention. For multicomponent

For peer review only

nudge interventions, we coded them according to the Nudge Ladder with all of the nudge types that apply. For studies with multiple nudge treatment groups, we reported the effect of each treatment arm separately. Only nudge interventions were compared to the control arms.

Due to the differences in the exposure, behavioral outcomes, and study designs interventions could not be directly compared with one another quantitively using effect sizes [35]. Hence, meta-analysis of nudge effects was infeasible. To synthesize the results, we used a vote counting method based on the direction of the effect for each study, an acceptable method for synthesis when meta-analysis is not possible [32]. If a simple majority of nudges were effective in a nudge category, the category was deemed effective.

Results

Study Selection

The systematic database search identified 3,586 entries, which were combined with another 22 articles of interest identified by the snowball sampling method, totaling 3,608 articles (see Appendix A for yield). After deduplication of records from the respective databases and snowball sampling techniques, 2,486 article records remained. Of the 2,486 articles, 2,486 articles from the systematic search and snowball method were retrievable and screened in the first stage of title and abstract screening, which reduced the total number of full-text screens to 133 unique articles. Of the 133 articles that were full-text screened, 39 articles [20,36–73] met inclusion criteria and the data from these were extracted and evaluated in this review (see PRISMA Diagram in Figure 2).

Study Characteristics

The characteristics of the included studies are summarized in Table 2. The majority (*n* = 35, 90%) of studies were conducted in the USA; two (5%) were conducted in the United Kingdom, one (3%) in Belgium, and one (3%) in Switzerland. Studies were set in a variety of healthcare

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

contexts (e.g., outpatient clinics, primary care practices, emergency departments, etc.) and targeted a variety of clinical decisions (e.g., opioid prescriptions, preventative cancer screening, checking vital signs of hospitalized patients). Nudges were directed at a variety of medical professionals (including physicians, nurses, medical assistants, and providers with a license to prescribe medication). Many (*n* = 20, 51%) of the studies did not report the sample size of clinicians interacting with the nudges. Instead, the studies tended to report the sample size in terms of how many patients were affected by the nudge or the number of prescription or lab orders under study. Fourteen (36%) studies were RCTs; 23 studies (59%) were pre-post designs; one study (3%) was a controlled interrupted time series design; and one study (3%) was a quasi-experimental randomized design. In terms of cluster RCTs, four studies (10%) were parallel cluster RCTs and three studies were stepped wedge cluster RCTs (8%). Most studies (*n* = 32, 82%) employed a control group/comparator that consisted of usual care or no intervention. One study (3%) used a minimal educational intervention, another study (3%) examining peer comparison letters used a placebo letter and five studies (13%) employed a factorial design in which multiple combined interventions were tested against individual interventions separately.

Of the 39 studies included in the review, 48 nudges were tested. Some studies contained multiple sub-studies, study arms or treatment groups, which were coded and analyzed separately (see Table 3). Given that some interventions (*n* = 5) were multicomponent (i.e., combinations of multiple nudges) these studies were analyzed separately using the Nudge Ladder (see Table 4).

Analyzing the single component nudges using the Nudge ladder, 6 nudges involved guiding choice through default options (e.g., changing the default opioid prescription quantity in the EHR); 9 nudges involved enabling choice (e.g., electronic prompts to accept or cancel orders for influenza vaccination); 22 nudges involved framing information (e.g., peer comparison letters to the clinicians in the top 50th percentile of antipsychotic prescriptions); two nudges involved prompting implementation intentions (e.g., displaying clinicians' pre-commitment letters in their own examination rooms) and four nudges involved providing information (e.g., an EHR reminder to clinicians when their patients were due for immunizations). Five studies involved multicomponent nudges, with four studies involving a combination of two nudges and one study involving a combination of three nudges (see Table 4).

Risk of Bias of Included Studies

Most studies were at high risk for selection bias including random sequence generation (*n* = 25) and allocation concealment (*n* = 25). Attrition bias was low risk based on incomplete outcome data (*n* = 31). A large number of trials were judged as unclear for selective reporting (*n* = 21). In terms of blinding of participants, most studies were high risk (*n*=25) and in terms of blinding outcome assessment, 25 studies were judged as having unclear risk of bias. Overall, 13 studies (33%) were considered low risk of bias across all criteria (see Table 5).

Synthesis of Results

With significance defined as ($p<0.05$), 33 of the 48 nudges (73%) significantly improved clinical decisions, suggesting that nudges are generally effective. According to the Nudge Ladder, all 6 (100%) of the nudges that involved changing the default option to guide decision-making were significantly related to clinician behavior change. Seven of the 9 (78%) nudges that enabled choice led to significant change in clinician behavior. Fourteen of the 22 (64%) nudges that involved framing information changed behavior significantly, suggesting their effectiveness. One of the two (50%) nudges that prompted implementation intentions was significantly effective and the other was not. None of the four (0%) nudges that provided information to clinicians resulted in statistically significant results. The five studies (100%) that combined nudges in multicomponent interventions were all effective.

Guiding choice through default options or enabling choice through an "active opt-out" model (i.e., active choice) were the most effective interventions in changing clinician behavior. These nudges also tended to result in the largest effect sizes. Nudges that framed information—the plurality of nudges under study—tended to also change clinician behavior. The other types of nudges were inconclusive or had more insignificant findings than significant findings.

Discussion

Summary of Evidence

This systematic review of 39 studies found that a variety of nudge interventions have been tested to improve clinical decisions. Thirty-three of the 48 (73%) clinician-directed nudges significantly improved clinical practice in the hypothesized direction. Nudges that changed default options or enabled choice were the most effective and nudges framing information for clinicians were also largely effective. Conversely, nudges that provided information to the clinician through reminders and prompting implementation intentions did not conclusively lead to significant changes in clinician behavior.

One strength of the taxonomy organizing this review is the ability to explicate why certain nudges are more effective and the mechanism by which they operate. Drawing on the Nudge Ladder, evidence suggests that less aggressive healthcare nudges lower on the ladder such as providing information and prompting intentions may be less effective than more aggressive nudges that are higher on the ladder such as changing the default options. This accords with nudge research in other areas outside of healthcare [74]. For example, one study comparing various types of nudges that increase the salience of information (e.g., including providing reminders, leveraging social norms, and framing information) with defaults found that only default nudges were effective at changing consumer pro-environmental behavior [8]. One large RCT of calorie labeling in restaurants found that posting caloric benchmarks (an informational nudge) paradoxically increased caloric intake for consumers [75].

The theoretical reasons for why less potent nudges (i.e., nudges at the bottom of Nudge Ladder) often fail are well established. People have a limited capacity to process information, so providing more data to decision-makers can be distracting or cognitively loading [76]. The timing of information is also essential—information is beneficial if it is top-of-mind during the decision [77]. Some of the social comparison nudges in this review provided information at opportune times, others did not [43]. Additionally, information improves decisions only if existing heuristics encourage errors. Often the information individuals receive may not be new to them. Worse still, informational nudges can have negative unintended consequences. For example, alert fatigue describes when clinicians are so inundated by alerts that they become desensitized and either miss or postpone their responses to them [78]. Finally, often reminders and information frames can be insufficiently descriptive in the course of action they suggest, rendering them futile. Given how much of clinicians' time is spent with the EHR, health system decision supports must be effective and not self-undermining.

More potent nudges (i.e., nudges at the top of the Nudge Ladder) are successful because they act on several key heuristics [79]. Defaults leverage inertia wherein overriding the default requires an active decision [80]. When people are busy and their attention scarce, they tend to rely on the status quo [81]. Moreover, people often see the default option as signaling an injunctive norm [82]. They see the default choice as the recommended choice and don't want to actively override this option unless they are very confident in their private decision. It is not surprising that

For peer review only

our study found that defaults were effective. It is also not surprising that nudges leveraging peer comparison tended to also be effective at shaping clinician behavior—clinicians who received messages that their behavior was abnormal compared to their peers, received a signal that helped them update their behavior.

Overall, results align with the conclusions of one [23] of the two recent systematic reviews of nudges tested in healthcare settings [22,23]. Differences in findings may be explained by different search strategies. One of these systematic reviews exclusively searched RCTs included in the Cochrane Library of systematic reviews and found that priming nudges—nudges that provide cues to participants—were the most studied and most effective nudges [22]. In that review, priming encompassed heterogenous interventions that span cues that elude conscious awareness, audit-and-feedback, and clinician reminders—to name a few—which may account for why study authors found those nudges to be the most numerous. The findings from our review conform with the results of the more traditional systematic review, conducted using a systematic search of several databases [23]. The latter review, like this one, found that default and social comparison nudges were the most frequently studied and most effective nudges. However, study authors focused their review on physician behavior, and our review is more expansive by studying all healthcare workers.

Limitations

Many of the studies in this review included at least some education (i.e., a non-nudge intervention) such as a reminder of the clinical guidelines. Because many studies (59%) were pre-post designs, they could not use these brief trainings in a control arm to evaluate the independent effect of the nudge. Therefore, we cannot decisively conclude whether nudges alone are responsible for the changes in clinician behavior. Similarly, many of the studies (51%) did not report the number of clinicians involved in the study (often reporting the sample in terms of how many patients or lab orders were affected by the nudge). Though unlikely, many of the effects could presumably be driven by a small portion of clinicians.

There was considerable variability in how researchers operationalized their primary outcome of interest. The effect of nudges may be contingent upon the behavior under study. One study [71] examining changes in opioid prescriptions led to a change in the number of 15-pill prescriptions (i.e., the change in "default" orders) but not in the total quantity of opioid pills prescribed, whereas other studies resulted in changes in the total number of opioid pills ordered after an EHR default change [83]. Establishing common metrics would enable direct comparison across studies and would allow us to conclusively determine if the nudge was effective overall at improving clinical decisions.

The considerable number of included papers reporting a statistically insignificant result decreases the usual concern over publication bias, which would skew the results towards desirable and more statistically significant outcomes. The majority of studies (n =21, 54%) were at unclear risk of selective reporting of outcomes (See Table 5). Moving forward, the field would benefit from reporting of all experimentation, whether its results are successful, unsuccessful, significant, or insignificant. Though not a majority, a large portion of studies (n = 12, 31%) were conducted by the same research team in the same health system. To validate that clinician-directed nudges are effective in other settings, other researchers should conduct nudge studies.

Though the nudge taxonomy used in the current review offered a way to classify the nudges described in the studies included, it was not developed empirically. The Nudge Ladder was developed based on a theoretical understanding of public health interventions. It is important to understand whether the conceptual distinctions made between nudge types are scientifically reliable and valid.

Future Research

Behavioral economics recognizes that nudges are "implicit social interactions" between the decision maker and the choice architect [84]. When faced with a nudge, people evaluate the motivations and values of the choice architect as well as how their decision will be understood by the choice architect and others. People tend to adhere to the default option when the choice architect is trusted, well-intentioned, and expert. Several non-healthcare default studies backfired when consumers distrusted the choice architect or felt they were nudged to spend more money [85]. Clinicians may reject nudges when they perceive health systems' preferences to conflict with their patients' interests. Research should attend to how engaged clinicians are in the implementation process and how they make inferences about the motivations and values of the choice architect when interacting with nudges using qualitative methods and surveys.

Nudges are also dependent on how decision makers believe they will be perceived. For example, around 40% of adults seeking care for upper respiratory tract infections want antibiotics and general practitioners report that patient expectations are a major reason for prescribing antibiotics [86,87]. Nudges that attempt to curtail antibiotic prescribing behavior may shape clinicians' behaviors in unexpected ways given clinicians' desire to demonstrate to their patients that they are taking serious action. Subtle features of how nudges are implemented may also influence clinicians' perceptions of the choice architect, heighten awareness of how their own actions may be perceived, and may undermine the nudge. Investigations of the clinicians' choice environment and clinicians' perspectives using qualitative and survey methods are crucial to the success of nudges.

Future research should also explore how clinician-directed nudges interact with one another in clinicians' choice environments. In our review, all multicomponent nudge studies (*n* = 5) were effective. However, it is also possible that nudges may crowd each other out when several different clinical decisions are targeted. In addition to alert fatigue, clinicians may experience nudge fatigue and begin to ignore decision support embedded in the EHR. Research should seek to understand how to develop nudges that can work synergistically with one another. Health systems and scientists can work together to understand which guidelines to prioritize and to develop decision support systems within their electronic interfaces that guide providers to make better clinical decisions.

Little work has been done on the sustainability of nudges beyond the study period, with some notable exceptions [88]. Particularly for nudges that require continued intervention on the part of the choice architects (e.g., peer comparison interventions), it's necessary to also understand their cost-effectiveness. Finally, understanding how nudges can be implemented across health systems is essential given that many of the studies included in this review were conducted in one health system.

Conclusion

This study adds to the growing literature on the study and effectiveness of nudges in healthcare contexts and can guide health systems in their choices of the types of nudges they should implement to improve clinical practice. The review describes how nudges have been employed in healthcare contexts and the evidence for their effectiveness across clinician behaviors, demonstrating potential for nudges, particularly nudges that change default settings or frame information for clinicians. More research is warranted to examine how nudges scale and their global effect on improving clinical decisions in complex healthcare environments.

Authors' Contributions

BSL conceived of and designed the research study; acquired and analyzed the data; interpreted the data; drafted the manuscript and substantially revised it. AMB helped design the research study; analyzed the data; interpreted the data; and substantially revised the manuscript. CET analyzed the data; interpreted the data; and substantially revised the manuscript. NM interpreted the data and substantially revised the manuscript. RSB helped conceive of and design the research study; interpreted the data; and substantially revised the manuscript. All authors approved the submitted version; have agreed to be accountable for the contributions; attest to the accuracy and integrity of the work, even aspects for which the authors were not personally involved.

Competing Interests

BSL, AMB, CET, and NM declare no financial or non-financial competing interests. Dr. Rinad Beidas reports royalties from Oxford University Press, has received consulting fees from the Camden Coalition of Healthcare Providers, currently consults for United Behavioral Health, and sits on the scientific advisory committee for Optum Behavioral Health.

Patient Consent for Publication

Not required.

Ethics Approval

Given the nature of systematic reviews, no human participant research was conducted for this original research contribution. Thus, the systematic review was not deemed subject to ethical approval and no human participants were involved in this study.

Funding

For peer review only

Funding for this study was provided by grants from the National Institute of Mental Health (P50 MH 113840, Beidas, Buttenheim, Mandell, MPI) and National Cancer Institute (P50 CA 244960). Briana S. Last also receives funding support from the National Science Foundation Graduate Research Fellowship Program (DGE-1321851).

Availability of Data and Materials

Given the nature of systematic reviews, the dataset generated and analyzed for the current study is already available. All studies analyzed for the present review are referenced for readers.

For peer review only

References

1 Simon HA. *Models of bounded rationality: Empirically grounded economic reason*. Cambridge, MA: : MIT press 1997.

2 Gigerenzer G, Gaissmaier W. Heuristic decision making. *Annual review of psychology* 2011;**62**:451–82. doi:https://doi.org/10.1146/annurev-psych-120709-145346

3 Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;**185**:1124. doi:10.1126/science.185.4157.1124

4 Thaler RH, Sunstein CR. *Nudge: Improving decisions about health, wealth, and happiness*. Revised and Expanded Edition. New York, NY: : Penguin Books 2009.

5 Szaszi B, Palinkas A, Palfi B, *et al.* A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making* 2018;**31**:355–66.

6 Arno A, Thomas S. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. *BMC Public Health* 2016;**16**:676. doi:10.1186/s12889-016-3272-x

7 Harbers MC, Beulens JW, Rutters F, *et al.* The effects of nudges on purchases, food choice, and energy intake or content of purchases in real-life food purchasing environments: a systematic review and evidence synthesis. *Nutrition Journal* 2020;**19**:1–27.

8 Momsen K, Stoerk T. From intention to action: Can nudges help consumers to choose renewable energy? *Energy Policy* 2014;**74**:376–82.

9 Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Medical Decision Making* 2015;**35**:539–57. doi:https://doi.org/10.1177/0272989X14547740

For peer review only

10 Croskerry P. A universal model of diagnostic reasoning. *Acad Med* 2009;**84**:1022–8. doi:10.1097/ACM.0b013e3181ace703

11 Graber M. Diagnostic Errors in Medicine: A Case of Neglect. *Joint Commission Journal on Quality and Patient Safety* 2005;**31**:106–13. doi:10.1016/S1553-7250(05)31015-4

12 Almashat S, Ayotte B, Edelstein B, *et al.* Framing effect debiasing in medical decision making. *Patient education and counseling* 2008;**71**:102–7.

13 Saposnik G, Redelmeier D, Ruff CC, *et al.* Cognitive biases associated with medical decisions: a systematic review. *BMC medical informatics and decision making* 2016;**16**:138.

14 Mendel R, Traut-Mattausch E, Jonas E, *et al.* Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine* 2011;**41**:2651–9. doi:https://doi.org/10.1017/S0033291711000808

15 Redelmeier DA, Shafir E. Medical decision making in situations that offer multiple alternatives. *Jama* 1995;**273**:302–5.

16 Patel MS, Volpp KG, Asch DA. Nudge Units to Improve the Delivery of Health Care. *The New England journal of medicine* 2018;**378**:214. doi:https://doi.org/10.1056/NEJMp1712984

17 Sittig DF, Singh H. *Electronic health records and national patient-safety goals*. Mass Medical Soc 2012.

18 Chaiyachati KH, Shea JA, Asch DA, *et al.* Assessment of Inpatient Time Allocation Among First-Year Internal Medicine Residents Using Time-Motion Observations. *JAMA Internal Medicine* 2019.

19 Hsiao C-J, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012. 2012.

For peer review only

20 Sacarny A, Barnett ML, Le J, *et al.* Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial. *JAMA Psychiatry* 2018;**75**:1003–11. doi:10.1001/jamapsychiatry.2018.1867

21 Lewis CC, Klasnja P, Powell B, *et al.* From classification to causality: advancing understanding of mechanisms of change in implementation science. *Frontiers in public health* 2018;**6**:136.

22 Yoong SL, Hall A, Stacey F, *et al.* Nudge strategies to improve healthcare providers' implementation of evidence-based guidelines, policies and practices: a systematic review of trials included within Cochrane systematic reviews. *Implementation Science* 2020;**15**:1–30.

23 Wang SY, Groene O. The effectiveness of behavioral economics-informed interventions on physician behavioral change: A systematic literature review. *Plos one* 2020;**15**:e0234149.

24 Council NB. Public health: ethical issues. *Nuffield Council on Bioethics* 2007.

25 Patel MS. Nudges for influenza vaccination. *Nature Human Behaviour* 2018;**2**:720.

26 Hillier-Brown FC, Summerbell CD, Moore HJ, *et al.* The impact of interventions to promote healthier ready-to-eat meals (to eat in, to take away or to be delivered) sold by specific food outlets open to the general public: a systematic review. *Obesity Reviews* 2017;**18**:227–46.

27 Münscher R, Vetter M, Scheuerle T. A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making* 2016;**29**:511–24.

28 Sunstein CR. Nudges vs. shoves. *Harv L Rev F* 2013;**127**:210.

29 Flodgren G, Eccles MP, Shepperd S, *et al.* An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane database of systematic reviews* 2011.

30 Sackett DL, Rosenberg WM, Gray JM, *et al. Evidence based medicine: what it is and what it isn't*. British Medical Journal Publishing Group 1996.

For peer review only

31 Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* 1981;**10**:141–63.

32 Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. 2008.

33 Davis DA, Mazmanian PE, Fordis M, *et al.* Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Jama* 2006;**296**:1094–102.

34 Higgins JP, Altman DG, Gøtzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj* 2011;**343**:d5928.

35 Nugent WR. The (Non) Comparability of the Correlation Effect Size Across Different Measurement Procedures: A Challenge to Meta-Analysis as a Tool for Identifying "Evidence Based Practices." *Journal of evidence-based social work* 2011;**8**:253–74.

36 Allen J.M., Dunn R., Bush J. Effect of prescriber peer comparison reports on fluoroquinolone use across a 16-facility community hospital system. *JACCP JAm Coll Clin Pharm* 2019;**2**:502–8. doi:10.1002/jac5.1106

37 Andereck J.W., Reuter Q.R., Allen K.C., *et al.* A Quality Improvement Initiative Featuring Peer-Comparison Prescribing Feedback Reduces Emergency Department Opioid Prescribing. *Jt Comm J Qual Patient Saf* 2019;**45**:669–79. doi:10.1016/j.jcjq.2019.07.008

38 Arora VM, Mochado N, Anderson SL, *et al.* Effectiveness of SIESTA on Objective and Subjective Metrics of Nighttime Hospital Sleep Disruptors. *J Hosp Med* 2019;**14**:38–41. doi:10.12788/jhm.3091

39 Bourdeaux CP, Davies KJ, Thomas MJ, *et al.* Using 'nudge'principles for order set design: a before and after evaluation of an electronic prescribing template in critical care. *BMJ Qual Saf* 2014;**23**:382–8.

For peer review only

40 Buntinx F, Knottnerus JA, Essed GG, *et al.* Long-term effect of feedback and peer comparison on the sampling quality of cervical smears--a randomized controlled trial. *Eur J Cancer Prev* 1995;**4**:153–7.

41 Chiu AS, Jean RA, Hoag JR, *et al.* Association of Lowering Default Pill Counts in Electronic Medical Record Systems With Postoperative Opioid Prescribing. *JAMA Surg* 2018;**153**:1012–9. doi:10.1001/jamasurg.2018.2083

42 Delgado MK, Shofer FS, Patel MS, *et al.* Association between Electronic Medical Record Implementation of Default Opioid Prescription Quantities and Prescribing Behavior in Two Emergency Departments. *Journal of general internal medicine* 2018;:1–3.

43 Hemkens LG, Saccilotto R, Reyes SL, *et al.* Personalized prescription feedback using routinely collected data to reduce antibiotic use in primary care: a randomized clinical trial. *JAMA internal medicine* 2017;**177**:176–83.

44 Hempel D, Pivetta E, Kimberly HH. Personalized peer-comparison feedback and its effect on emergency medicine resident ultrasound scan numbers. *Crit Ultrasound J* 2014;**6**:1. doi:10.1186/2036-7902-6-1

45 Hsiang EY, Mehta SJ, Small DS, *et al.* Association of an Active Choice Intervention in the Electronic Health Record Directed to Medical Assistants With Clinician Ordering and Patient Completion of Breast and Colorectal Cancer Screening Tests. *Jama Network Open* 2019;**2**:e1915619. doi:10.1001/jamanetworkopen.2019.15619

46 Kim RH, Day SC, Small DS, *et al.* Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination. *JAMA Netw Open* 2018;**1**:e181770. doi:10.1001/jamanetworkopen.2018.1770

For peer review only

47 Kullgren JT, Krupka E, Schachter A, *et al.* Precommitting to choose wisely about low-value services: a stepped wedge cluster randomised trial. *BMJ Qual Saf* 2018;**27**:355–64.

48 Lewis S, Young B, Thurley P, *et al.* Evaluation of a nudge intervention providing simple feedback to clinicians of the consequence of radiation exposure on demand for computed tomography: a controlled study. *Clinical Medicine* 2019;**19**:290–3. doi:10.7861/clinmedicine.19-4-290

49 Meeker D, Knight TK, Friedberg MW, *et al.* Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. *JAMA Intern Med* 2014;**174**:425–31. doi:10.1001/jamainternmed.2013.14191

50 Meeker D, Linder JA, Fox CR, *et al.* Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. *Jama* 2016;**315**:562–70.

51 Nguyen C.T., Davis K.A. Evaluating the impact of peer comparison on vancomycin dose order verification among pharmacists. *JACCP JAm Coll Clin Pharm* 2019;**2**:137–42. doi:10.1002/jac5.1046

52 O'Reilly-Shah VN, Easton GS, Jabaley CS, *et al.* Variable effectiveness of stepwise implementation of nudge-type interventions to improve provider compliance with intraoperative low tidal volume ventilation. *BMJ Qual Saf* 2018;**27**:1008–18. doi:10.1136/bmjqs-2017-007684

53 Olson J, Hollenbeak C, Donaldson K, *et al.* Default settings of computerized physician order entry system order sets drive ordering habits. *J Pathol Inform* 2015;**6**:16. doi:10.4103/2153-3539.153916

For peer review only

54 Orloski CJ, Tabakin ER, Shofer FS, *et al.* Grab a Seat! Nudging Providers to Sit Improves the Patient Experience in the Emergency Department. *Journal of Patient Experience* 2019;**6**:110–6. doi:10.1177/2374373518778862

55 Parrino TA. The nonvalue of retrospective peer comparison feedback in containing hospital antibiotic costs. *Am J Med* 1989;**86**:442–8.

56 Patel MS, Volpp KG, Small DS, *et al.* Using active choice within the electronic health record to increase influenza vaccination rates. *Journal of general internal medicine* 2017;**32**:790–5.

57 Patel MS, Volpp KG, Small DS, *et al.* Using active choice within the electronic health record to increase physician ordering and patient completion of high-value cancer screening tests. *Healthc (Amst)* 2016;**4**:340–5. doi:10.1016/j.hjdsi.2016.04.005

58 Patel MS, Day S, Small DS, *et al.* Using Default Options Within the Electronic Health Record to Increase the Prescribing of Generic-Equivalent Medications. *Annals of Internal Medicine* 2014;**161**:S44–52.

59 Patel MS, Kurtzman GW, Kannan S, *et al.* Effect of an Automated Patient Dashboard Using Active Choice and Peer Comparison Performance Feedback to Physicians on Statin Prescribing The PRESCRIBE Cluster Randomized Clinical Trial. *Jama Network Open* 2018;**1**:e180818. doi:10.1001/jamanetworkopen.2018.0818

60 Persell SD, Doctor JN, Friedberg MW, *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

61 Ryskina K, Dine CJ, Gitelman Y, *et al.* Effect of social comparison feedback on laboratory test ordering for hospitalized patients: a randomized controlled trial. *Journal of general internal medicine* 2018;**33**:1639–45.

For peer review only

62 Sedrak MS, Myers JS, Small DS, *et al.* Effect of a price transparency intervention in the electronic health record on clinician ordering of inpatient laboratory tests: the PRICE randomized clinical trial. *JAMA internal medicine* 2017;**177**:939–45.

63 Sharma S, Guttmann D, Small DS, *et al.* Effect of Introducing a Default Order in the Electronic Medical Record on Unnecessary Daily Imaging During Palliative Radiotherapy for Adults With Cancer: A Stepped-Wedge Cluster Randomized Clinical Trial. *Jama Oncology* 2019;**5**:1220–2. doi:10.1001/jamaoncol.2019.1432

64 Shively N.R., Buehrle D.J., Wagener M.M., *et al.* Improved antibiotic prescribing within a veterans affairs primary care system through a multifaceted intervention centered on peer comparison of overall antibiotic prescribing rates. *Antimicrob Agents Chemother* 2020;**64**. doi:10.1128/AAC.00928-19

65 Srinivasan M., Huntman J., Nelson M., *et al.* Use of peer comparison, provider education, and electronic medical record triggers to increase influenza vaccination rates in hospitalized children. *Hosp Pediatr* 2020;**10**:76–83. doi:10.1542/hpeds.2019-0076

66 Suffoletto B., Landau A. Nudging Emergency Care Providers to Reduce Opioid Prescribing Using Peer Norm Comparison Feedback: A Pilot Randomized Trial. *Pain Med* Published Online First: 2019. doi:10.1093/pm/pnz314

67 Szilagyi PG, Serwint JR, Humiston SG, *et al.* Effect of provider prompts on adolescent immunization rates: a randomized trial. *Academic pediatrics* 2015;**15**:149–57.

68 Trent SA, Havranek EP, Ginde AA, *et al.* Effect of Audit and Feedback on Physician Adherence to Clinical Practice Guidelines for Pneumonia and Sepsis. *Am J Med Qual* 2018;:1062860618796947. doi:10.1177/1062860618796947

For peer review only

69 Wigder HN, Ballis SFC, Lazar L, *et al.* Successful implementation of a guideline by peer comparisons, education, and positive physician feedback. *Journal of Emergency Medicine* 1999;**17**:807–10. doi:10.1016/S0736-4679(99)00087-6

70 Winickoff RN, Coltin KL, Morgan MM, *et al.* Improving physician performance through peer comparison feedback. *Med Care* 1984;**22**:527–34.

71 Zivin K, White JO, Chao S, *et al.* Implementing Electronic Health Record Default Settings to Reduce Opioid Overprescribing: A Pilot Study. *Pain Med* 2019;**20**:103–12. doi:10.1093/pm/pnx304

72 Zwank M.D., Kennedy S.M., Stuck L.H., *et al.* Removing default dispense quantity from opioid prescriptions in the electronic medical record. *Am J Emerg Med* 2017;**35**:1567–9. doi:10.1016/j.ajem.2017.04.002

73 Patel M, Day SC, Halpern SD, *et al.* Generic medication prescribing rates after health system-wide redesign of default options within the electronic health record. *JAMA Internal Medicine* 2016;**176**:847–8. doi:10.1001/jamainternmed.2016.1691

74 Sunstein CR. Nudges that fail. *Behavioural Public Policy* 2017;**1**:4–25.

75 Downs JS, Wisdom J, Wansink B, *et al.* Supplementing menu labeling with calorie recommendations to test for facilitation effects. *American Journal of Public Health* 2013;**103**:1604–9.

76 Marois R, Ivanoff J. Capacity limits of information processing in the brain. *Trends in cognitive sciences* 2005;**9**:296–305.

77 Karlan D, McConnell M, Mullainathan S, *et al.* Getting to the top of mind: How reminders increase saving. *Management Science* 2016;**62**:3393–411.

For peer review only

78 Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. *AACN advanced critical care* 2013;**24**:378–86.

79 Jachimowicz JM, Duncan S, Weber EU, *et al.* When and why defaults influence decisions: a meta-analysis of default effects. *Behavioural Public Policy* 2019;**3**:159–86. doi:https://doi.org/10.1017/bpp.2018.43

80 Madrian BC, Shea DF. The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics* 2001;**116**:1149–87.

81 Kahneman D, Knetsch JL, Thaler RH. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives* 1991;**5**:193–206.

82 Everett JA, Caviola L, Kahane G, *et al.* Doing good by doing nothing? The role of social norms in explaining default effects in altruistic contexts. *European Journal of Social Psychology* 2015;**45**:230–41.

83 Zwank MD, Kennedy SM, Stuck LH, *et al.* Default versus open text narcotic prescription writing in the emergency department electronic medical record. *Acad Emerg Med* 2014;**21**:S87–8. doi:10.1111/acem.12365

84 Krijnen JM, Tannenbaum D, Fox CR. Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy* 2017;**3**:i–18.

85 Brown CL, Krishna A. The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of consumer research* 2004;**31**:529–39.

86 Fletcher-Lartey S, Yee M, Gaarslev C, *et al.* Why do general practitioners prescribe antibiotics for upper respiratory tract infections to meet patient expectations: a mixed methods study. *BMJ open* 2016;**6**:e012244–e012244. doi:10.1136/bmjopen-2016-012244

For peer review only

87 Linder JA, Singer DE. Desire for antibiotics and antibiotic prescribing for adults with upper respiratory tract infections. *Journal of general internal medicine* 2003;**18**:795–801. doi:10.1046/j.1525-1497.2003.21101.x

88 Olshan D, Rareshide CAL, Patel MS. Longer-Term Durability of Using Default Options in the Electronic Health Record to Increase Generic Prescribing Rates. *J Gen Intern Med* Published Online First: 2018. doi:10.1007/s11606-018-4719-9

For peer review only

Table 1. Eligibility Criteria

Inclusion Criteria	Full-text empirical journal articles
	English language
	Published in a Peer-Reviewed Journal
	The studies in the paper empirically investigated one or more behavioral intervention techniques that were considered nudges or were connected to the choice architecture literature by the original authors. These interventions are all clinician-directed (e.g., nurses, doctors, residents, medical assistants), not patient-directed.
	The studies in the paper had behavioral outcome variables, not preferences or attitudes (e.g., prescribing behavior).
Exclusion Criteria	Abstracts unavailable in the first-pass screen
	Review articles, conference abstracts, textbooks, chapters, and conference papers.
	Studies without a control group or baseline comparator
	The studies in the paper applied interventions that restrict the freedom of choice of the target population, included significant economic incentives, ongoing education, complex decision support systems, or consultation.

For peer review only

Figure 1. Ladder of nudge interventions.

Note. Adapted from [24,25].

For peer review only

Figure 2. PRISMA flow diagram

For peer review only

Table 2. Study Characteristics

Authors (Year, Country)	Setting	Design	Intervention	Justification	Sample size	Outcomes Measured	Main findings	Significance
Allen, Dunn, & Bush (2019) USA [36]	Health system (16 community hospitals across 8 counties)	Prospective, pre-post design	Quarterly peer comparison reports were sent to eligible prescribers (by email, fax, or in-person, etc.). Eligible prescribers (who accounted for 75%-80% of total prescribed “antibiotic days”) were unaware they were high-volume antibiotic prescribers.	Reduce antibiotic prescriptions of fluoroquinolones due to their broad spectrum of activity, known adverse event profile, and availability of other less toxic therapeutic options	Internal medicine; hospitalists; family medicine (n = 189). Critical care; pulmonology (n = 67). Infectious diseases (n = 60)	Primary study outcome was fluoroquinolone days of therapy/1000 patient days (DOT/1000 PD). A day of therapy was defined as at least one dose of a fluoroquinolone in a 24-hour period, per each facility's medication administration records.	Antibiotic use declined 29% (baseline: 83.9 DOT/1000 PD, range: 59.3-118.7; intervention: 58.3 DOT/1000 PD, range: 37.1-76.7). Primary outcome (fluoroquinolone days of therapy/1000 patient days) declined for all facilities included in the study.	$p<0.001$
Andereck et al. (2019) USA [37]	Large urban academic Emergency Department (ED)	Prospective pre-post design (QI initiative)	Quarterly feedback by e-mail. Prescribers could compare their rates to peers on a de-identified chart of their peers. Formal education and training complimented the peer intervention (e.g., a brief “pharmacy fact ”	Unnecessary prescribing patterns have contributed to the opioid epidemic.	Preintervention period, 35,636 ED visits were discharged. *M* = 44 attending physicians, 30 senior resident physicians, and 33 junior resident physicians and advanced practice providers per block met inclusion.	The primary outcome of this evaluation was the overall ED discharge opioid prescribing rate. Prescribing rate was defined as the proportion of discharged patient encounters with an opioid prescription for the department in a specific	Departmental opioid prescribing rates during the evaluation period declined; Preintervention period rate: 8.6% (95% CI: 8.3–8.9) vs post: 5.8% (95% CI: 5.5–6.1)	$p<0.001$

			with each email and a pharmacist lecture)		Postintervention period, a total of 18,830 ED visits were discharged. *M* = 40 attending physicians, 30 senior residents, and 35 junior residents and advanced practice providers per block met inclusion threshold	scheduling block.		
Arora et al. (2019) USA [38]	Two general medicine inpatient units	Prospective, cross-sectional pre-post design	Changing the EHR, creating a default to monitor patient's vital signs; Customized office signs for nurses educating them about best "sleep-friendly" vitals monitoring practices; pocket-cards with information; 20-minute education session.	Sleep is important for patient recovery but patients struggle to sleep in hospitals, which is related to poor outcomes.	*n* = ? providers 1,083 general-medicine patients, 1,669 EHR general medicine orders	Changes in the mean percentage of "sleep-friendly" (i.e., non-nocturnal) orders for checking vital signs and venous thromboembolism prophylaxis compared to baseline.	Increases in the mean percentage of sleep-friendly orders rose for both: no vital sign: 3% to 22%, sleep-promoting VTE prophylaxis: 12% to 28%.	p<0.001
Bourdeaux et al. (2014) UK [39]	Inpatient Intensive Care Unit	Retrospective Pre-post design	Prescription template with preprescribed drugs and fluids Doctors choose to use the	Chlorhexidine mouthwash reduces ventilator associated pneumonia in critically ill	*n* = ? providers 2231 ventilated patients were eligible for	Changes in the delivery of chlorhexidine mouthwash and HES to patients	Percentage of patients prescribed chlorhexidine increased (35.1%). The	1- p<0.001 2- p< 0.001

For peer review only

			template upon admission.	patients. It is cheap and acceptable. Hydroxyethyl starch (HES) is an intravenous fluid that helps circulation.	chlorhexidine, 591 pre- and 1640 post-intervention. 6199 patients were eligible for HES intervention, 2177, pre- and 4022 post	in the intensive care unit.	mean volume of HES infused per patient fell and the percentage of patients receiving HES fell (-51.0%).	
Buntinx et al. (1993) Belgium [40]	Department of Pathology	Randomized controlled trial (RCT)	Interventions, four groups. Some arms had feedback and then advice. One arm had peer comparison	Cervical screening can help prevent cancer.	183 doctors	Percentage of smears lacking endocervical cells	Smears lacking endocervical cells decreased in the groups receiving monthly peer comparison overviews compared to groups not receiving this type of feedback. OR = 0.75, 95% CI (0.58 – 0.96)	p<0.05
Chiu et al. (2018) USA [41]	Health System (5 hospitals)	Prospective pre-post design	Changing the EHR, lowered the default number of pills on electronic opioid prescriptions from 30 to 12 after procedure.	Postprocedural analgesia prescriptions have contributed to the opioid epidemic	*n* = ? providers 1447 procedures before default change and 1463 procedures after the default change	Changes in prescription rates, the median number of opioid pills prescribed per operation.	Decreases in the median number of opioid pills prescribed -5.22 (CI: -6.12 - -4.32)	p<0.01
Delgado et al. (2018) USA [42]	Two emergency departments	Prospective pre-post design	Changing the EHR, lowered the default number of pills	Reliance on prescription opioids for postprocedural	*n* = ? providers 3264 prescriptions	Increase in 10 pill prescriptions relative to control 4 weeks	Increase in proportion of prescriptions for 10 tablets	p<0.001

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

			on electronic opioid prescriptions to 10 pills.	analgesia has contributed to the opioid epidemic	were written across the two EDs	after implementation; changes in the mean number of Oxy/Apap tablets prescribed per week.	27.8%, 95% CI 17.4–37.5. No change in the mean number of Oxy/APAP tablets prescribed per week.	
Hemkens et al. (2017) Switzerland [43]	Nationwide	Pragmatic RCT	Personalized antibiotic prescription feedback by mail and an online dashboard and a letter on antibiotic prescribing guidelines	Clinicians often inappropriately prescribe antibiotics for acute respiratory tract infections	2,900 primary care physicians	Changes in defined daily doses of any antibiotic to any patient per 100 consultations in first year, intention-to-treat, relative to control.	No change in prescribing behavior: between-group difference, 0.81%; 95% CI, −2.56 - 4.30.	N.S.
Hempel et al. (2014) USA [44]	Emergency department	Prospective pre-post design	Peer comparison feedback on emergency medicine resident ultrasound scan numbers.	Clinician-performed ultrasounds are part of emergency medicine residency curricula; there is a need for effective teaching.	44 emergency medicine residents	Changes in number of scans done per shift in the three months after intervention (relative to baseline)	Increase in number of scans performed (number of ultrasound exams per shift increased from 0.39 scans/shift to 0.61 scans/shift).	$p<0.05$
Hsiang et al. (2019) USA [45]	Health System (25 primary care practices)	Retrospective difference-in-differences approach (intervention vs control practices during post-intervention year compared to the 2	Active choice of a best-practice alert for medical assistants. During vitals check, the electronic health record (EHR) prompted medical assistants to	US Preventive Services Task Force guidelines for breast and colorectal cancer screening	*n* = ? providers 26,269 women eligible for breast cancer screening, 43,647 men eligible for colorectal cancer screening	Primary outcome was ordering of the screening test during a visit (primary care) compared to control groups relative to 2 pre-intervention years	Breast cancer screening tests (22.2 % point increase, 95% CI, 17.2-27.6) and colorectal cancer screening test increased (13.7% point	$p<0.001$

For peer review only

		preintervention years)	accept/cancel a cancer screening order. If accepted, a pending order was made for the clinician to review and sign during the patient visit.				increase, 95% CI, 8.0-18.9).	
Kim et al. (2018) USA [46]	11 Primary Care Practices	Prospective, cross-sectional pre-post design (Differences-in-differences)	Changing the EHR, an "active choice" intervention using a best practice alert directed to medical assistants—prompt to accept or cancel a flu vaccine order. If accepted, the order was made for the physician to review and sign during the patient visit.	Center for Disease Control recommends universal influenza vaccination	*n = ?* providers 96, 291 patients	Changes in flu vaccination rates compared with control practices over time.	Increase in flu vaccination rates (9.5 % point increase in vaccination rates (95% CI, 4.1-14.3).	p<0.001
Kullgren et al. (2018) USA [47]	6 adult primary care practices	12-month stepped wedge cluster RCT, randomization by clinic	Clinicians precommited to "Choosing Wisely" choices against low-value orders. They received 1–6 months of point-of-care precommitment reminders, patient education	Clinicians often order costly and inappropriate tests as well as inappropriately prescribe antibiotics for acute respiratory tract infections	45 primary care physicians and advanced practice providers	Primary outcome was the difference between control and intervention period percentages of visits with potentially low-value orders.	No change in in the percentage of visits with potentially low-value orders overall, for headaches or for acute sinusitis (−1.4%, 95%CI −2.9 - 0.1).	N.S.

For peer review only

			handouts, and weekly emails.					
Lewis et al. (2019) UK [48]	Acute medical hospital	Controlled interrupted time series design.	Message at the bottom of all inpatient and outpatient paper and electronic computerized tomography (CT) reports, highlighting patients at risk after exposure to ionising radiation and asks the provider if they informed the patient.	CT scans are known to expose individuals to radiation, which can increase cancer risk.	n = ? providers	Immediate change in level or a gradual trend change in CT counts in electronic reports compared to control hospital.	Significant reduction in CT scans (-4.6%, 95% CI (-7.4 — -1.7).	p = 0.002
Meeker et al., (2014) USA [49]	5 primary care clinics	RCT, randomization by clinician	Poster-sized commitment letters in clinicians' personal examination rooms for 12 weeks. These letters displayed clinician photographs, signatures, and commitment to not inappropriately prescribe antibiotics for acute respiratory infections	Clinicians often inappropriately prescribe antibiotics for acute respiratory tract infections despite guidelines and several clinical interventions	14 clinicians (11 physicians and 3 nurse practitioners) 954 eligible adult patients	Differences in antibiotic prescribing rates for antibiotic-inappropriate acute respiratory infection diagnoses at baseline and during intervention periods.	Decrease in inappropriate antibiotic prescribing rate compared to control (difference in difference - 19.7%, 95% CI (-33.4 — -5.8)	$p<0.05$
Meeker et al. (2016) USA [50]	47 primary care practices in 2 different	2 × 2 × 2 factorial RCT (Practices	1- Changes in EHR, "suggested alternatives"	Clinicians often inappropriately prescribe	248 clinicians	Changes in rates of inappropriate antibiotic	1- No significant change in	1 – NS; 2- $p<0.001$; 3- $p<0.001$.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	health systems	received 0, 1, 2, or 3 interventions)	presented electronic order sets with nonantibiotic treatments 2- Changes in EHR, "accountable justification" clinicians enter free-text justifications for prescribing antibiotics 3- Peer comparison emails about how clinicians' antibiotic prescribing rates compare to lowest inappropriate prescribers	antibiotics for acute respiratory tract infections	(14, 753 visits at baseline and 16, 959 during intervention period)	prescribing behavior compared to baseline	inappropriate antibiotic prescriptions; difference in difference: -5%, 95% CI (-7.8-0.1) 2- Decrease in inappropriate antibiotic prescriptions; difference in difference: -7%, 95% CI(-9.1 — -2.9) 3- Decrease in inappropriate antibiotic prescriptions; difference-in-difference: -5.2%, 95 CI (-6.9 — -1.6)	No statistically significant interactions between intervention s
Nguyen & Davis (2019) USA [51]	One multi-specialty academic medical center	Single center, prospective, quasi-experimental pre-post design	Peer comparison reports of the percentage of appropriately verified vancomycin orders for each pharmacist. In phase I, reports were blinded. In phase II, reports were unblinded. Intervention phases were compared to a	Pharmacist "order verification" prevents medical errors, which are harmful to patients. Vancomycin is a commonly prescribed drug for hospitalized patients.	*n* = ? providers 1,625 vancomycin orders were included for evaluation (537 orders in the control group, 549 orders in intervention phase I, and 539 orders in	Appropriate vancomycin dose order verification, Appropriate dose was determined by the institution's guidelines.	Appropriately verified vancomycin orders significantly increased in the phase II (unblinded) compared with the control group (OR = 1.79; 95% CI (1.36-2.34)	$p < 0.001$

			pre-intervention control.		intervention phase II).			
O'Reilly-Shah et al. (2018) USA [52]	Department of Anesthesiology in a large health system (two academic hospitals, two private practice hospitals and two academic surgery centers)	Retrospective pre-post design (stepwise cluster implementation in 5 facilities)	1- Audit and feedback on provider level and department-level compliance with lung-protective ventilation (LPV) for attending physicians. 2- Audit and feedback for advance practice providers and residents 2- Changes to the EHR, default setting on anaesthesia machines for tidal volume was decreased from 700 mL to 400 mL.	There is a need to improve compliance with anesthesiology surgical quality metrics	*n = ?* providers 5 facilities, Total surgical case count (n) = 14,793 unique patients (n) = 12,785. 5 facilities.	Rates of compliance with low tidal wave ventilation compared to baseline	Attending physician dashboards increased compliance odds 41% (OR 1.41, 95% CI 1.17 - 1.69). Adding advanced practice provider and resident dashboards increased compliance odds 93% (OR 1.93, 95% CI 1.52 - 2.46). Changing ventilator defaults led to 376% increase in compliance odds OR 3.76, 95% CI 3.1 - 4.57.	1- p = 0.002 2- p<0.001 3- p<0.001
Olson et al. (2015) USA [53]	Clinical pathology, hematology, and oncology departments in a health system	Prospective Pre-post design (multiple baseline)	Changes in the EHR default order sets for posttransfusion hematocrits and platelet counts changed from "optional" to "preselected." Platelet count default settings	Need to improve the monitoring of posttransfusion outcomes	> 500 residents and fellows. 7578 orders for red blood cell transfusion, 3285 total orders for platelet transfusion	Rates of lab test ordering for post-transfusion counts after default change and post default change	Increase in hemocrit and platelet posttransfusion count orders after default for order was set to "pre-selected" (8.3% to 57.5% change). After switch back to	p < 0.001

For peer review only

			later changed back to "optional"				"optional", significant decrease in orders	
Orloski et al. (2019) USA[54]	2 urban, academic emergency departments	Prospective, controlled pre–post trial	Placed institution-branded folding seats in the emergency department and an educational campaign on good communication. Only the intervention ED received folding seats.	Patient satisfaction is important	*n* = ? providers 2,827 patients were surveyed	Primary outcome was the impact of provider sitting on patient satisfaction. Secondary outcome was provider sitting frequency	Sitting at any point during an emergency department encounter increased patient satisfaction across all measures (polite: 67% vs 59%, cared: 64% vs 54%, listened: 60% vs 52%, informed: 57% vs 47%, time: 56% vs 45%. Odds of provider sitting increased 30% when a seat was in the room, OR = 1.3, 95% CI (1.1-1.5)	p<0.0001
Parrino (1989) USA [55]	One tertiary referral hospital	Prospective pre-post design	Monthly peer comparison letters sent to two groups (surgical and nonsurgical physicians) who were in the top 50 percentiles of prescribers for	Antibiotics are often inappropriately prescribed and can be expensive	202 physicians, surgical (n = 83) and nonsurgical (n = 119)	Changes in expenditures (total dollars) on antibiotics per physician (mean difference from quarter 3 to quarter 4 compared to control group	No significant change in total dollars spent on antibiotics (mean difference: $797.50 vs $1355.33)	N.S.

For peer review only

			antibiotic expenditures			before and after feedback)		
Patel et al. (2014) USA[58]	One general internal medicine and one family medicine practice	Retrospective cross-sectional pre-post design	Modify EHR default from showing brand and generic medications to displaying only generics at first, with the ability to opt out.	Generic medications are less expensive than brand-name medications and are of comparable quality	Internal and family medicine attending physicians (IM, n = 38; FM, n = 17) and residents (IM, n = 166; FM, n = 34)	Monthly prescriptions of brand-name and generic equivalent for: beta-blockers, statins, and proton-pump inhibitors compared to control.	Increase in generic prescribing behavior for all three medications; 5.4 % points, 95% CI, (2.2 — 8.7)	p<0.001
Patel et al. (2016) USA [57]	Three internal medicine practices	Prospective cross-sectional pre-post design (Differences in differences)	Changing the EHR through "active choice" using a best practice alert for medical assistants and physicians, prompting them to accept/cancel an order for a colonoscopy, mammography, or both. Physician needed to review and sign order during visit.	Guidelines suggest that increasing early cancer detection can be done through regular screening practices	*n = ?* providers One intervention practice, 2 controls. 7560 patients eligible for colonoscopy with 14,546 clinic visits and 8,337 patients eligible for mammography with 14,410 clinic visits.	Percentage of patients eligible for screening who received a cancer screening order	Increase in mammography (12.4% points, 95% CI: 8.7–16.2) and colonoscopy orders (11.8% points, 95% CI: 8.0–15.6).	p<0.001
Patel et al. (2016) USA [73]	All specialties across a Health System	Pre-post design, difference-in-differences approach	"Active choice" in the EHR. An opt-out "checkbox" that said "dispense as written" was added to the prescription EHR screen, and	Generic medications are linked to higher adherence to medication regimens and better clinical outcomes	*n* = ? providers Pre-intervention data: 811,561 eligible prescription sets during 10-month	Generic prescribing rates for 10 medical conditions i.e., 10 drugs	The overall generic prescribing rate increased significantly (75.3% to 98.4%)	p < 0.001

For peer review only

			if unchecked the drug's generic version was prescribed.		preintervention period to 655,011 prescriptions during 7-month postintervention period			
Patel et al. (2017) USA [56]	Three Internal Medicine practices	Prospective cross-sectional, pre-post design (Differences-in-differences)	Changing the EHR through "active choice" using a best practice alert directed to medical assistants and physicians—prompting to accept/cancel an order for the flu vaccine. Physician needed to review and sign during the patient visit.	The Center for Disease Control recommends universal influenza vaccination	*n = ?* providers One intervention practice, 2 control practices. 45,926 patients	Changes in flu vaccination rates	Increase in vaccination rates (adjusted difference-in-difference: 6.6 % points; 95% CI, 5.1–8.1).	p<0.001
Patel et al (2018) USA [59]	One health system, 32 primary care practices	3-arm cluster randomized Clinical trial	1- "Active choice" and "accountable justification" Physicians received an email with number of eligible patients for statin therapy who had not been prescribed a statin and were asked to actively choose to	50% of eligible patients do not receive statins despite evidence of their efficacy	96 PCPs 4774 patients eligible but not receiving statin therapy	Percentage of eligible patients receiving statin prescription orders compared to usual care	1- No significant increase in statin prescription rates vs. usual care (adjusted difference: 4.1%, 95% *CI*, −0.8 to 13.1). 2- Increase in statin prescription compared to usual care	1- NS 2- p<0.01

For peer review only

			prescribe atorvastatin, 20 mg, once daily, atorvastatin at another dose, or another statin or not prescribe a statin and describe a reason. 2- Active choice and accountable justification and peer comparison e-mails describing how physicians compared to peers.				(adjusted difference, 5.8%; 95% CI, 0.9-13.5).	
Persell et al (2016) USA[60]	General internal medicine clinic	2 × 2 × 2 factorial RCT with 3 interventions	1- "Accountable justification" in EHR. Physicians received an alert when inappropriately prescribing an antibiotic and provided free-text justification 2- "Suggested alternatives" in EHR when physicians inappropriately prescribe antibiotics 3- Peer comparison monthly performance	Clinicians frequently prescribe antibiotics inappropriately for acute respiratory infections	*n = ?* providers 3,276 visits in the pre-intervention year and 3,099 visits in the intervention year	Rate of oral inappropriate antibiotic prescriptions for acute respiratory infection diagnoses compared to control group and baseline.	No significant decrease in inappropriate prescribing rates compared to control group. Significant decrease in inappropriate prescribing across all groups (including controls) compared to baseline: 1 –OR= 0.98, 95% CI (0.42 – 2.29)	N.S.

For peer review only

			feedback compared to lowest 10% of inappropriate prescribers				2- OR = 0.68, 95% CI (0.29-1.58) 3- OR = 0.45, 95% CI (0.18 to 1.11)	
Ryskina et al. (2018) USA[61]	Six general medicine teams in one health system	Single-blinded cluster RCT, Randomization by 2-week service block.	Peer comparison e-mails sent to physicians on general medicine teams, summarizing their routine lab test orders vs. the service average that week	Routine laboratory tests for hospitalized patients can be wasteful and are overused	6 attending physicians, 114 interns and residents	Number of routine laboratory orders placed by each physician per patient day.	No significant changes in number of laboratory orders by each physician (-0.14 tests per patient-day vs. control group, 95% CI − 0.56 —0.27).	N.S.
Sacarny et al (2018) USA [20]	Highest volume primary care prescribers of quetiapine in 2013 and 2014, whose patients have Medicare	RCT (intent to treat) placebo-control parallel-group design, balanced randomization (1:1) to control group (placebo letter) and treatment group (peer comparison letter).	Mailed peer comparison letters saying that prescriber's quetiapine prescribing was under review and was high relative to same-state peers, which was concerning and could be medically unjustified.	Antipsychotic agents like quetiapine fumarate are often overprescribed when not clinically indicated/supported with the potential to cause patient harm.	5,055 PCPs, 231 general practitioners, 2428 were in family medicine, and 2396 were in internal medicine.	Total quetiapine days prescribed by physicians from the intervention start to 9 months in intervention vs control.	Decrease in quetiapine days per prescriber in treatment vs control arm; -11.1%, 95% CI (-13.1 — -9.2)	$p<0.001$
Sedrak et al (2017) USA [62]	Three hospitals in one health system	RCT comparing a 1-year nudge to a 1-year pre-nudge period, accounting for	Intervention lab tests showed Medicare allowable fees at the time of order in the EHR and control lab tests	A significant number (30%) of laboratory tests in the U.S. may be wasteful. Increasing price transparency at	*n = ?* providers 60 diagnostic laboratory tests, 30 most frequently	Frequency of tests ordered per patient-day. Secondary outcome was the number of tests done per patient-	No significant changes in number of tests ordered between intervention and control	N.S.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

		time and patient features Randomizatio n at test-level	did not show prices.	the time of lab order entry may influence provider decisions and decrease wasteful tests	ordered and 30 most expensive. 142, 921 hospital admissions, 98,529 patients	day and the Medicare fees.	group (0.05 tests ordered per patient-day; 95% CI, −0.002 — 0.09)	
Sharma et al. (2019) USA [63]	One health System	Stepped-wedge cluster randomized clinical trial	Change EHR through a default imaging order for no daily imaging during palliative radiotherapy, which physicians could opt-out from by specifying another imaging frequency	Guidelines suggest that imaging using radiography or computed tomography on a daily basis is unnecessary for patients undergoing palliative radiotherapy. Daily imaging can be costly and increase treatment duration for patients.	21 radiation oncologists 1019 patients who received 1188 palliative radiotherapy courses (n = 747 at university practice; n = 441 at community practices) to bone, soft tissue, brain, or various sites	Primary outcome was binary outcome (whether radiotherapy courses with daily imaging were ordered). Daily imaging course was defined as imaging during ≥80% of palliative therapy treatments.	Default led to a significant reduction in daily imaging adjusted OR = 0.43; 95% CI, 0.24-0.77; adjusted difference in % points, −18.6; 95% CI, −34.1 — −2.1	p=0.004
Shively et al. (2020) USA [64]	Veterans' Affairs Health System (7 primary care practices)	Prospective pre-post design	Peer comparison feedback—an educational session for all primary care providers and monthly e-mails with their antibiotic prescribing rate, their colleague's rates, and the system's goal rates.	Clinicians frequently inappropriately prescribe antibiotics despite guidelines.	Baseline = 65 primary care professionals (PCPs) serving 40,734 patients, 28,402 office visits Intervention = 73 PCPs serving 41,191 patients, 32,982 office visits	Monthly mean rate of antibiotic prescribing rates. Secondary outcomes were inappropriate antibiotic prescribing rates and appropriate antibiotic prescribing rates	Mean rate of monthly antibiotic prescriptions significantly reduced 35.6%. Unnecessary antibiotic prescribing decreased 33.9% and the appropriate antibiotic rates increased 50.8%.	p<0.001

Srinivasan et al. (2020) USA [65]	Inpatient units in a 350-bed children's hospital	Prospective Pre-post design	EHR reminders, provider education (including a quiz), and peer comparison feedback (how unit rates compared to other units in the hospital, shown on posters and sent by email)	American Academy of Pediatrics guidelines for universal, yearly influenza vaccination for all children 6 months and older	*n = ?* providers Baseline = 6,089 admitted children (6 months and older) to the medical and surgical units Intervention = 6,206 children admitted	Primary outcome was percentage of children discharged with 1 dose (or greater) of the influenza vaccine (from the hospital or before admission)	Significant increase in the percentage of discharged children with at least 1 dose of the flu vaccine (4.7-fold increase, from 10% to 46%)	p<0.001
Suffoletto & Landeau (2019) USA [66]	Emergency departments in one hospital system, 16 hospitals	A pilot RCT (randomizatio n by provider)	Audit and feedback (A&F) emails vs peer norm comparison (PC) emails to other emergency medicine providers at their hospital	Opioid epidemic is still a persistent problem; need to reduce opioid prescriptions	37 emergency medicine providers	Mean monthly opioid prescriptions by provider	Opioid prescriptions reduced non-significantly in both conditions (audit and feedback, and peer norm comparison) Mean reduction (SD) was 3.3. (9.6) for controls, 3.9 (10.5) in A&F, and 7.3 (7.8) for A&F + PC	N.S.
Szilagyi et al. (2014) USA [67]	Practices in two large research networks	RCT, randomization unit by practices in two practice-based research networks	EHR prompts/alerts at all office visits with vaccine recommendation s. Reminder sheet on the provider's desk in the exam room with	Guidelines recommend adolescent immunization for a host of diseases; yet vaccination rates are not in line with guidelines	*n* = ? providers 2 practice networks: 1 network: 5 intervention, 5 control practices; 1 network: 6 intervention, 6	Changes in adolescent immunization rates, by practice	No significant difference in immunization rates between intervention and control practices for any vaccine or combination of vaccines (e.g., adjusted OR	N.S.

For peer review only

			indicated vaccines.		control practices		for HPV vax at one site: 0.96; 95% CI 0.64–1.34), at another: adjusted OR = 1.06; 95% CI 0.68–1.88	
Trent et al. (2018) USA [68]	One medical center, an urban, safety net, Level 1 trauma center	Stepped wedge design and cluster randomization	Monthly audit and feedback emails with blinded peer comparison feedback adherence to guidelines for pneumonia and severe sepsis. Physicians also received emails about patients that got nonadherent service to review	Adherence to guidelines for pneumonia and sepsis treatment are low in emergency departments	*n* = ? providers 469 patients during entire study period	Primary outcome was guideline-adherent antibiotic choices (guidelines determined by the institution)	Adherence to antibiotic guidelines significantly increased after audit and feedback with peer comparison was introduced (adjusted OR = 1.8, 95% CI: 1.01-3.2	$p<0.05$
Wigder et al (1999) USA [69]	Emergency department in a 600-bed hospital, with a Level 1 Trauma center	Prospective, pre-post design	1- Education campaign of "Ottawa rule" 2- Physicians shown baseline data. 3- Audit and feedback. Knee injury patient charts put in physician mailboxes praising them for "Ottawa rule" adherence or	Physicians overorder X-rays when guidelines (i.e., the "Ottowa rule") recommend less invasive and cheaper ways for evaluating knee problems/injuries	27 physicians	Primary outcome was changes in patients with knee injuries who received an X-ray study. Secondary outcome was percentage of X-ray orders with abnormal results	Significant decrease (23%) in number of X-ray studies, increase (58.4%) in percentage of abnormal X-rays compared to baseline.	$p<0.001$

For peer review only

			informing of nonadherence					
Winickoff et al. (1984) USA [70]	Department of Internal Medicine at one group practice	3 Interventions: Pre-post design for first 2. 3rd intervention: RCT with crossover design (over a 1 year, crossover at 6 months)	1 - Educational meeting for clinical standard 2 – Peer comparison, meeting presenting group standard adherence pre and post the educational meeting 3 – Peer comparison feedback, monthly feedback about how physicians compare to peers at practice.	Many clinicians do not follow guidelines for colorectal screening	*n = ?* for first 2 interventions 16 physicians for RCT (3rd intervention)	Number of stool tests completed for colorectal cancer screening across groups who received peer comparison intervention.	1 – Little change in stool tests done 2- Little change in stool tests done 3- Increase in number of stool tests done (66.7% to 82.2% across groups)	1- N.S. 2- N.S. 3- $p < 0.001$
Zivin et al (2019) USA [71]	Two health systems	Prospective, pre-post design	Modify EHR default for all Schedule II opioid prescriptions to 15-pills (many EHRs had 30-day defaults previously, others had no default)	The opioid epidemic; overprescription of opioids for postprocedural pain management is a problem and out of step with guidelines	448 prescribers 6,390 opioid prescriptions	Primary outcome was changes in the proportion of opioid prescriptions for 15 pills for high frequency prescribers	Percentage of 15-pill prescriptions by high prescribers increased from 2.3% to 8.1% (chi-squared = 6.72), 15-pill opioid prescription rates increased at both sites (4.1% to 7.2% at one site, 15.9% to 37.2% at other site)	$p<0.04$

Zwank et al. (2017) USA [72]	Emergency department of a Level 1 trauma center	Retrospective pre-post design	Changing the EHR default number of pills for opioid prescriptions from 15 tablets to a number the physician had to enter themselves	The opioid epidemic; overdose deaths due to prescriptions from opioids as analgesics	*n = ?* providers 7,019 eligible prescriptions	Changes in the total opioid pill quantity per prescription	No significant change in mean number of opioid tablets per prescription Mean tablets dispensed increased from 15.31 (SD = 5.30) tablets to 15.77 (SD = 7.30).	N.S.

For peer review only

Table 3. Studies Organized According to Nudge Ladder

Nudge Ladder	Study	Significant Effect in the Hypothesized Direction?	Majority in Category Significant?
Provide Information	Meeker et al. (2016) USA [50] —Arm 1	N.S.	No
	Persell et al. (2016) USA [60] — Arm 2	N.S.	
	Sedrak et al (2017) USA [62]	N.S.	
	Szilagyi et al. (2014) USA [67]	N.S.	
Frame Information	Allen, Dunn, & Bush (2019) USA [36]	p<0.001	Yes
	Andereck et al. (2019) USA [37]	p<0.001	
	Buntinx et al. (1993) Belgium [40]	p>0.05	
	Hemkens et al. (2017) Switzerland [43]	N.S.	
	Hempel et al. (2014) USA [44]	p<0.05	
	Lewis et al. (2019) UK [48]	p = 0.002	
	Meeker et al. (2016) USA [50] – Arm 2	p<0.001	
	Meeker et al. (2016) USA [50] – Arm 3	p<0.001	
	Nguyen & Davis (2019) USA [51]	p < 0.001	
	O'Reilly-Shah et al. (2018) [52]— Arm 1	p = 0.002	
	O'Reilly-Shah et al. (2018) [52] — Arm 2	p <0.001	
	Parrino (1989) [55] USA	N.S.	
	Persell et al. (2016) USA [60] — Arm 1	N.S.	
	Persell et al. (2016) USA [60] — Arm 3	N.S.	
	Ryskina et al. (2018) USA [61]	N.S.	
	Sacarny et al (2018) USA [20]	p<0.001	
	Shively et al. (2020), USA [64]	P<0.001	
	Suffoletto & Landeau (2019) USA [66]	N.S.	
	Trent et al. (2018), USA [68]	p<0.05	
	Winickoff et al. (1984) USA [70] — Study 1	N.S.	
	Winickoff et al. (1984) USA [70] — Study 2	N.S.	
	Winickoff et al. (1984) USA [70] — Study 3	p <0.001	
Prompt Implementation Intentions	Kullgren et al. (2018) USA	N.S.	No
	Meeker et al. (2014) USA [49]	p <0.05	
Enable Choice	Bourdeaux et al. (2014) UK [39]	p<0.001 for both	Yes
	Hsiang et al. (2019) USA [45]	<0.001	
	Kim et al. (2018) USA [46]	p<0.001	
	Orloski et al. (2019) USA [54]	p<0.0001	
	Patel et al. (2016) USA [73]	p<0.001	
	Patel et al. (2016) USA [57]	p < 0.001	
	Patel et al. (2017) USA [56]	p<0.001	

For peer review only

	Patel et al. (2018) USA [59] — Arm 1	N.S.	
	Zwank et al. (2017) USA [72]	N.S.	
Guide choice through default options	Chiu et al. (2018) USA [41]	p<0.01	Yes
	Delgado et al. (2018) USA [42]	p<0.001	
	Olson et al. (2015) USA [53]	p < 0.001	
	Patel et al. (2014) USA [58]	p<0.001	
	Sharma et al. (2019) USA [63]	p=0.004	
	Zivin et al. (2019) USA [71]	p<0.04	

Note. Articles that included multiple intervention treatment groups, studies, or study arms are described.

For peer review only

Table 4. Multicomponent Intervention Studies Organized According to Nudge Ladder

Nudge Ladder	**Study**	**Significant Effect in the Hypothesized Direction?**
Provide information + Guide choice through default options	Arora et al. (2019) USA [38]	$p < 0.001$
Provide Information + Frame Information	Wigder et al. (1999) USA [69]	$p<0.001$
Enable Choice + Frame Information	Patel et al. (2018) USA [59]— Arm 2	$p < 0.001$
Frame Information + Guide choice through default options	O'Reilly-Shah et al. (2018) USA [52] — Arm 3	$p < 0.001$
Provide information + Frame Information + Enable choice	Srinivasan et al. (2020) USA [65]	$p < 0.001$

For peer review only

Table 5. Cochrane Risk of Bias Assessment Tool.

Authors (Year, Country)	**Random Sequence Generation**	**Allocation Concealment**	**Blinding (participants and personnel)**	**Blinding Outcome Assessors**	**Incomplete Outcome Data**	**Selective Reporting**
Allen, Dunn, & Bush (2019) USA [36]	−	−	−	?	+	?
Andereck et al. (2019) USA [37]	−	−	−	?	+	?
Arora et al. (2019) USA [38]	−	−	−	?	−	?
Bourdeaux et al. (2014) UK [39]	−	−	−	?	+	?
Buntinx et al. (1993) Belgium [40]	+	+	+	+	+	?
Chiu et al. (2018) USA [41]	−	−	−	?	+	+
Delgado et al. (2018) USA [42]	−	−	−	?	+	?
Hemkens et al. (2017) Switzerland [43]	+	+	+	+	+	+
Hempel et al. (2014) USA [44]	−	−	−	?	?	?
Hsiang et al. (2019) USA [45]	−	−	−	?	+	?
Kim et al. (2018) USA [46]	−	−	−	?	+	+
Kullgren et al. (2018) USA [47]	+	+	+	+	+	+
Lewis et al. (2019) UK [48]	−	−	−	?	?	+
Meeker et al., (2014) USA [49]	+	+	+	+	+	+
Meeker et al. (2016) USA [50]	+	+	+	+	+	+
Nguyen & Davis (2019) USA [51]	−	−	−	?	?	?
O'Reilly-Shah et al. (2018) USA [52]	−	−	−	?	+	+
Olson et al. (2015) USA [53]	−	−	−	?	+	?
Orloski et al. (2019) USA[54]	−	−	−	?	?	?

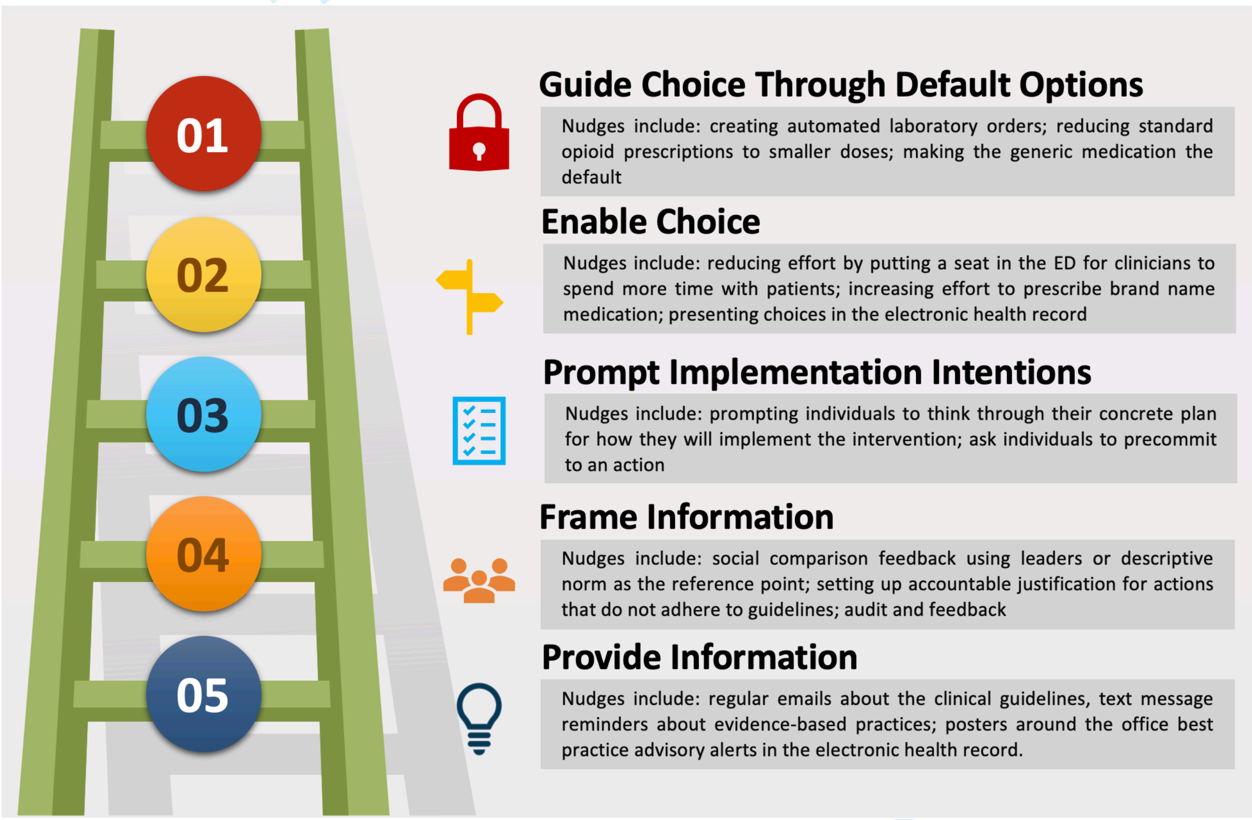
For peer review only

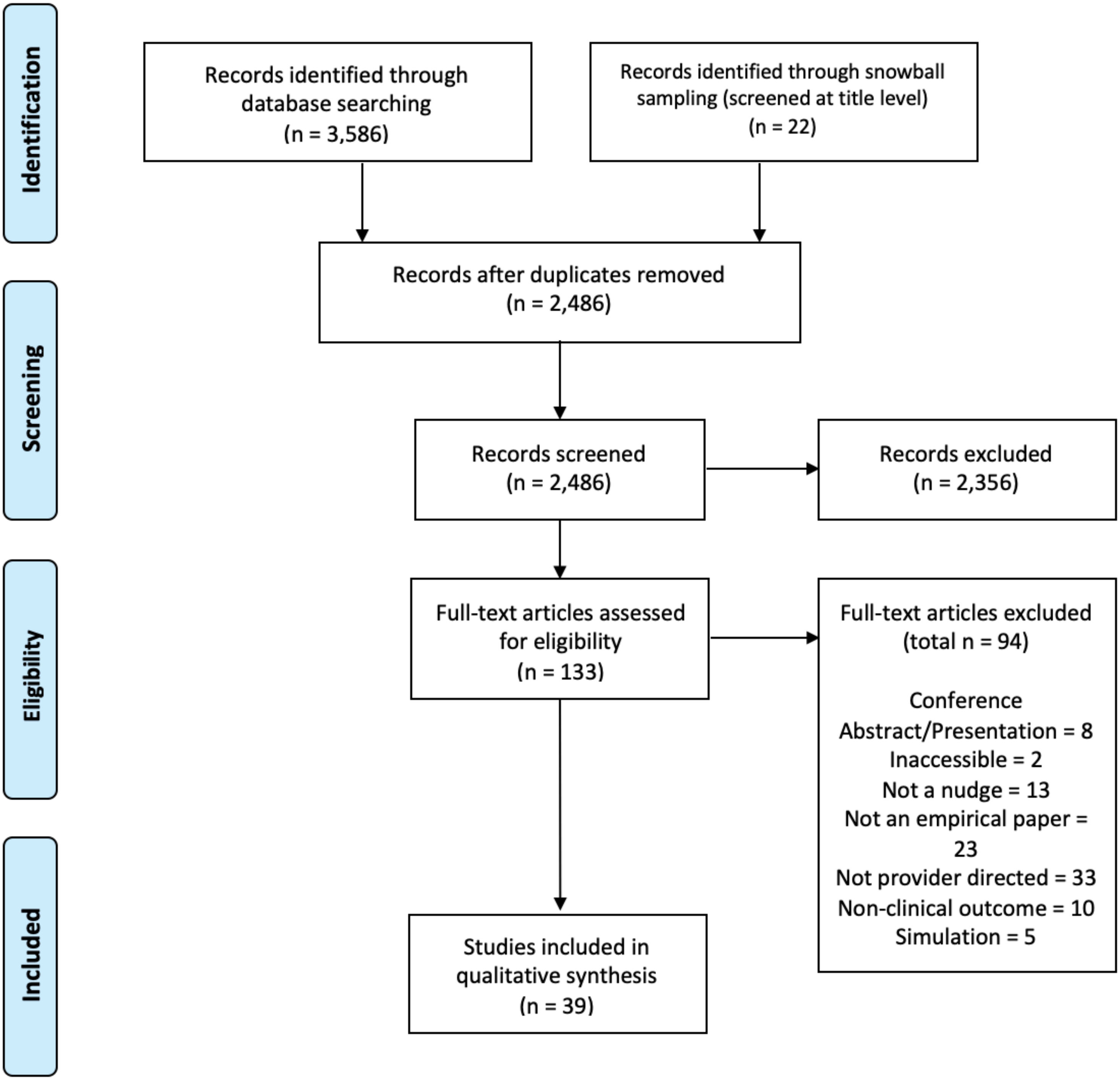
Parrino (1989) USA [55]	−	−	−	?	+	?
Patel et al. (2014) USA[58]	−	−	−	?	+	?
Patel et al. (2016) USA [73]	+	+	+	+	+	+
Patel et al. (2016) USA [57]	+	+	+	+	+	+
Patel et al. (2017) USA [56]	+	+	+	+	+	+
Patel et al (2018) USA [59]	+	+	+	+	+	+
Persell et al (2016) USA[60]	+	+	+	+	+	+
Ryskina et al. (2018) USA[61]	+	+	+	+	+	+
Sacarny et al (2018) USA [20]	+	+	+	+	+	+
Sedrak et al (2017) USA [62]	+	+	+	+	+	+
Sharma et al. (2019) USA [63]	+	+	+	+	+	+
Shively et al. (2020) USA [64]	−	−	−	?	?	?
Srinivasan et al. (2020) USA [65]	−	−	−	?	?	?
Suffoletto & Landeau (2019) USA [66]	+	+	+	+	+	+
Szilagyi et al. (2014) USA [67]	+	+	+	+	+	+
Trent et al. (2018) USA [68]	+	+	+	+	+	+
Wigder et al (1999) USA [69]	−	−	−	?	?	?
Winickoff et al. (1984) USA [70]	First 2 studies: − 3rd study: +	First 2 studies: − 3rd study: +	First 2 studies: − 3rd study: +	?	+	+
Zivin et al (2019) USA [71]	−	−	−	?	+	?
Zwank et al. (2017) USA [72]	−	−	−	?	+	?

For peer review only

Note. ⊕ indicates low risk of bias, ⊖ indicates high risk of bias, and ? indicates unclear risk of bias. See (72) for a full description of the Cochrane Risk of Bias tool.

For peer review only


01
02
03
04
05
Guide Choice Through Default Options
Nudges include: creating automated laboratory orders; reducing standard opioid prescriptions to smaller doses; making the generic medication the default
Enable Choice
Nudges include: reducing effort by putting a seat in the ED for clinicians to spend more time with patients; increasing effort to prescribe brand name medication; presenting choices in the electronic health record
Prompt Implementation Intentions
Nudges include: prompting individuals to think through their concrete plan for how they will implement the intervention; ask individuals to precommit to an action
Frame Information
Nudges include: social comparison feedback using leaders or descriptive norm as the reference point; setting up accountable justification for actions that do not adhere to guidelines; audit and feedback
Provide Information
Nudges include: regular emails about the clinical guidelines, text message reminders about evidence-based practices; posters around the office best practice advisory alerts in the electronic health record.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 3,586)
Records identified through snowball sampling (screened at title level) (n = 22)
Records after duplicates removed (n = 2,486)
Records screened (n = 2,486)
Records excluded (n = 2,356)
Full-text articles assessed for eligibility (n = 133)
Full-text articles excluded (total n = 94)
Conference Abstract/Presentation = 8
Inaccessible = 2
Not a nudge = 13
Not an empirical paper = 23
Not provider directed = 33
Non-clinical outcome = 10
Simulation = 5
Studies included in qualitative synthesis (n = 39)

Appendix A

Systematic Search Strategy

The methodology for the search was designed based on standards for systematic reviews [32], in consultation with a medical librarian, as well as with two experts from the field of healthcare behavioral economics. The databases used were: EconLit, Embase, EBSCO Megafile, PsycINFO, PubMed, Scopus, and Web of Science.

Search terms included combinations, plurals, and various conjugations of the words relating to identified nudge strategies. The search string and strategy from [6] was used as a basis for search terms, but adjusted to reflect the more specific clinician-directed aim of this research question. All peer-reviewed empirical studies published prior to the completion of our search phase (i.e., – 4/2020) were eligible for this review.

Following retrieval of all records, duplicates were removed using Zotero (www.zotero.org), and via manual inspection. Article screening involved two stages. First, all records were screened at the title and abstract level by a team of four coders (the first-author and three research assistants) using the web-based application for systematic reviews, Rayyan (https://rayyan.qcri.org). Criteria in this first-pass screening were more inclusive—that is, all interventions directed at clinicians were included and examined further. To establish reliability, the first-author and the three coders screened the same 20 articles and then reviewed their screening decisions together. Any disagreements were resolved by consensus. This process was repeated three additional times until 80 articles were screened by all four coders and sufficient reliability was established. Reliability was excellent (fleiss' $\kappa = 0.96$). For the remainder of the screening process, screening was done independently by all four coders; the team met weekly to discuss any articles that they were uncertain about including or excluding. This screening process

For peer review only

was followed by a full text examination to finally determine inclusion or exclusion according to more stringent inclusion and exclusion criteria (see Table 1). This screening process was done as a team and determinations of article inclusion were decided collaboratively.

Search Terms

The following search terms were used in the systematic search. All searches were conducted in the title field.

EBSCO Megafile

TI (nudg* OR choice architect OR choice architecture OR behavioral intervention OR behavioural intervention OR behavioral economic OR behavioral economics OR behavioral insight OR behavioural insight OR active choice OR default OR default bias OR default option OR opt-out OR opt-in OR prompted choice OR commitment device OR accountable justification OR peer comparison OR pre-commitment) AND TI (physician OR health OR clinician OR clinic OR provider* OR electronic health record OR health record OR doctor OR nurse OR physician assistant OR medical assistant OR electronic medical record OR medical record OR medical OR outpatient OR inpatient OR hospital OR resident)

EconLit

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND TI (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or

For peer review only

electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Embase

(nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

PsycInfo

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND TI (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

PubMed

TI (nudg* OR choice architect OR choice architecture OR behavioral intervention OR behavioural intervention OR behavioral economic OR behavioral economics OR behavioral insight OR behavioural insight OR active choice OR default OR default bias OR default option OR opt-out OR opt-in OR prompted choice OR commitment device OR accountable justification OR peer comparison OR pre-commitment) AND TI(physician OR health OR clinician OR clinic OR provider* OR electronic health record OR health record OR doctor OR nurse OR physician assistant OR medical assistant OR electronic medical record OR medical record OR medical OR outpatient OR inpatient OR hospital OR resident)

Scopus

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Web of Science

TI (nudg* or choice architect or choice architecture or behavioral intervention or behavioural intervention or behavioral economic or behavioral economics or behavioral insight or behavioural insight or active choice or default or default bias or default option or opt-out or opt-in or prompted choice or commitment device or accountable justification or peer comparison or pre-commitment) AND (physician or health or clinician or clinic or provider* or electronic health record or health record or doctor or nurse or physician assistant or medical assistant or

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

electronic medical record or medical record or medical or outpatient or inpatient or hospital or resident)

Table A1. Search Dates and Yields

Database	**Date**	**Yield**
EBSCO Megafile	4/22/2020	482
EconLit	4/22/2020	28
Embase	4/22/2020	1,240
PsycInfo	4/22/2020	384
PubMed	4/22/2020	292
Scopus	4/22/2020	30
Web of Science	4/22/2020	1,130
Total		3,586

For peer review only

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	4-5
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	7-9
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	9
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	10
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	10
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	12
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplemental A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	12-13
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	12
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	13
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	14
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	14
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	14-15

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	14
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	15
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	15-16
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	17
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	40-57
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	17-18
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	17
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-20
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	20-2
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22-23
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	25

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml