

## REVIEWS

# Measures Used to Assess the Impact of Interventions to Reduce Low-Value Care: a Systematic Review



Jennifer K. Maratt, MD, MS<sup>1,2</sup>, Eve A. Kerr, MD, MPH<sup>1,2,3</sup>, Mandi L. Klamerus, MPH<sup>3</sup>, Shannon E. Lohman, BS<sup>4</sup>, Whit Froehlich, BA<sup>5</sup>, R. Sacha Bhatia, MD, MBA<sup>6</sup>, and Sameer D. Saini, MD, MS<sup>1,2,3</sup>

<sup>1</sup>Department of Internal Medicine and Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, USA; <sup>2</sup>Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, MI, USA; <sup>3</sup>Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI, USA; <sup>4</sup>Wayne State University School of Medicine, Detroit, MI, USA; <sup>5</sup>University of Michigan Medical School, Ann Arbor, MI, USA; <sup>6</sup>Department of Internal Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, USA.

**IMPORTANCE:** Studies of interventions to reduce low-value care are increasingly common. However, little is known about how the effects of such interventions are measured.

**OBJECTIVE:** To characterize measures used to assess interventions to reduce low-value care.

**EVIDENCE REVIEW:** We searched PubMed and Web of Science to identify studies published between 2010 and 2016 that examined the effects of interventions to reduce low-value care. We also searched [ClinicalTrials.gov](http://ClinicalTrials.gov) to identify ongoing studies. We extracted data on characteristics of studies, interventions, and measures. We then developed a framework to classify measures into the following categories: utilization (e.g., number of tests ordered), outcome (e.g., mortality), appropriateness (e.g., overuse of antibiotics), patient-reported (e.g., satisfaction), provider-reported (e.g., satisfaction), patient-provider interaction (e.g., informed decision-making elements), value, and cost. We also determined whether each measure was designed to assess unintended consequences.

**FINDINGS:** A total of 1805 studies were identified, of which 101 published and 16 ongoing studies were included. Of published studies ( $N=101$ ), 68% included at least one measure of utilization, 41% of an outcome, 52% of appropriateness, 36% of cost, 8% patient-reported, and 3% provider-reported. Funded studies were more likely to use patient-reported measures (17% vs 0%). Of ongoing studies (registered trials) ( $N=16$ ), 69% included at least one measure of utilization, 75% of an outcome, 50% of appropriateness, 19% of cost, 50% patient-reported, 13% provider-reported, and 6% patient-provider interaction.

**Prior Presentations** This work was presented in draft form on February 1, 2017, at the AcademyHealth “Establishing Priorities to Develop Patient-Centered Metrics for Low-Value Care” meeting in Washington, D.C., and on September 5, 2017, at the AcademyHealth “Setting Priorities for Low-Value Care Measurement with Patients at the Center” webinar. Drs. Saini and Kerr received honoraria for these presentations.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11606-019-05069-5>) contains supplementary material, which is available to authorized users.

Received October 10, 2018

Revised January 2, 2019

Accepted April 23, 2019

Published online June 27, 2019

Of published studies, 34% included at least one measure of an unintended consequence as compared to 63% of ongoing studies.

**CONCLUSIONS AND RELEVANCE:** Most published studies focused on reductions in utilization rather than on clinically meaningful measures (e.g., improvements in appropriateness, patient-reported outcomes) or unintended consequences. Investigators should systematically incorporate more clinically meaningful measures into their study designs, and sponsors should develop standardized guidance for the evaluation of interventions to reduce low-value care.

**KEY WORDS:** measures; low-value care; interventions.

J Gen Intern Med 34(9):1857–64

DOI: 10.1007/s11606-019-05069-5

© Society of General Internal Medicine (This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply) 2019

## INTRODUCTION

Low-value care has been characterized as services that provide little to no benefit to patients, have potential to cause harm, incur unnecessary costs to patients, or waste limited healthcare resources.<sup>1–3</sup> Efforts to minimize the use of low-value care are increasingly important to the public, providers, healthcare systems, and payers. The Choosing Wisely® campaign, in particular, has focused on identifying potentially unnecessary treatments, tests, and services that patients and physicians should question.<sup>4</sup> In parallel, we have seen rapid growth in studies of interventions that target low-value care, and sponsors are now increasingly supporting funding in this area.<sup>5</sup>

Prior studies have examined the types, effectiveness, and quality of interventions to reduce the use of low-value services and the validity of specific measures used to assess low-value care.<sup>1,6,7</sup> However, we have limited knowledge of how the effects of interventions to reduce unnecessary services are being *measured* (i.e., what types of intervention outcomes are being assessed?). To ensure that we are measuring outcomes that are clinically meaningful, and

because complex interventions can have far-reaching and often unintended consequences, it is important that efforts to reduce low-value care are assessed in comprehensive and systematic ways.<sup>8,9</sup>

The purpose of this review was to characterize and examine what types of measures are routinely employed in studies of interventions designed to reduce low-value services. We hypothesized that existing studies largely focus on utilization rather than more clinically meaningful measures, such as appropriateness. We also hypothesized that unintended consequences (such as underuse of appropriate services) are not being systematically assessed, and that patient-reported measures are used infrequently. By presenting a framework for measuring the effects of interventions to decrease low-value services and elucidating the current status of intervention assessment, we sought to highlight current gaps and inform researchers and sponsors about what types of clinically meaningful measures should be incorporated in future studies.

## METHODS

We performed a systematic review to identify studies that evaluated interventions to reduce low-value care. We adhered to the standards outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement in conducting and reporting our systematic review ([Online Appendix A](#)).

### Search Strategy

To identify potentially relevant studies, we searched PubMed, Web of Science, and [ClinicalTrials.gov](#). Our search strategy was based on a previously published systematic review by Colla et al., of interventions used to decrease low-value care, which included publications through early 2015.<sup>6</sup> We expanded the terms from this prior review with input from a medical librarian to ensure inclusion of all possibly relevant studies ([Online Appendix B](#)). In brief, we searched for articles that contained any of the key terms “health services misuse,” “unnecessary procedures,” “low value,” “waste,” “overu\*,” or “wasteful,” in addition to any word belonging to one of nine sets of topical search term sets. These search terms sets contained multiple terms and synonyms related to possible intervention designs to reduce use of low-value care: (1) cost sharing and value-based purchasing, (2) patient education and decision-making, (3) quality indicators and reporting, (4) physician performance incentives, (5) utilization management strategies, (6) financial risk sharing and physician reimbursement, (7) clinical decision support, (8) provider education, and (9) provider feedback and peer reporting. We focused on studies published or initiated after the 2010 introduction of the “Top Five List” (a precursor to Choosing Wisely®),<sup>10</sup> restricting our search from January 1, 2010, to December 31, 2016. The search was not restricted by language.

### Study Selection Criteria

Three authors (SL, MK, JM) reviewed titles and abstracts identified by the search strategy to select potentially relevant studies ([Fig. 1](#)). Studies were included if they presented primary research that examined the effect of an intervention to reduce low-value care. As previously mentioned, low-value care was defined as services that provide little to no benefit to patients, have the potential to cause harm, incur unnecessary costs to patients, or waste limited healthcare resources.<sup>1–3</sup> Studies were excluded if they examined cost reduction only. For studies that met inclusion criteria, full-text articles of published studies (from PubMed and Web of Science) and protocols for ongoing studies (from [ClinicalTrials.gov](#)) were retrieved and reviewed as described below. Our partners at AcademyHealth also reviewed our search strategy and recommended inclusion of two additional studies that were not identified in our initial literature search.<sup>11,12</sup> After full-text review of these articles to ensure that they met inclusion criteria, they were also included in this systematic review.

### Development of Measure Categorization Framework

We developed a rubric to categorize measure type. We first reviewed recommendations and literature from measurement organizations including National Quality Forum<sup>13</sup> and the American College of Physicians,<sup>3</sup> recent systematic reviews<sup>1,6,14</sup> of studies on low-value care, and recently published work<sup>8,9</sup> on proposed approaches to measuring overuse and low-value care. From this body of literature, we developed a comprehensive framework to capture the wide range of effects—both intended and unintended—of interventions to reduce use of low-value care. While we recognized that many studies might simply focus on reductions in utilization, we deliberately sought to incorporate categories for measures that are clinically meaningful and patient-centered. For example, a study designed to decrease the use of antibiotic prescriptions for viral infections could measure, in addition to the reduction in overall use, whether prescribing was reduced for the right patients (appropriateness), whether patients in the intervention arm had fewer allergic reactions (outcomes), patient satisfaction with the encounter (patient-reported experiences), and whether there were any increases in hospital admissions for patients who did not receive antibiotics (unintended outcomes). Additional examples of measure categorization may be found in [Online Appendix F](#). We further categorized each measure according to whether it assessed unintended consequences of the intervention (e.g., underuse of services, substitution of services, patient-reported experiences, provider-reported experiences, or patient-provider interactions). The resulting measure categorization framework is presented in [Table 1](#).

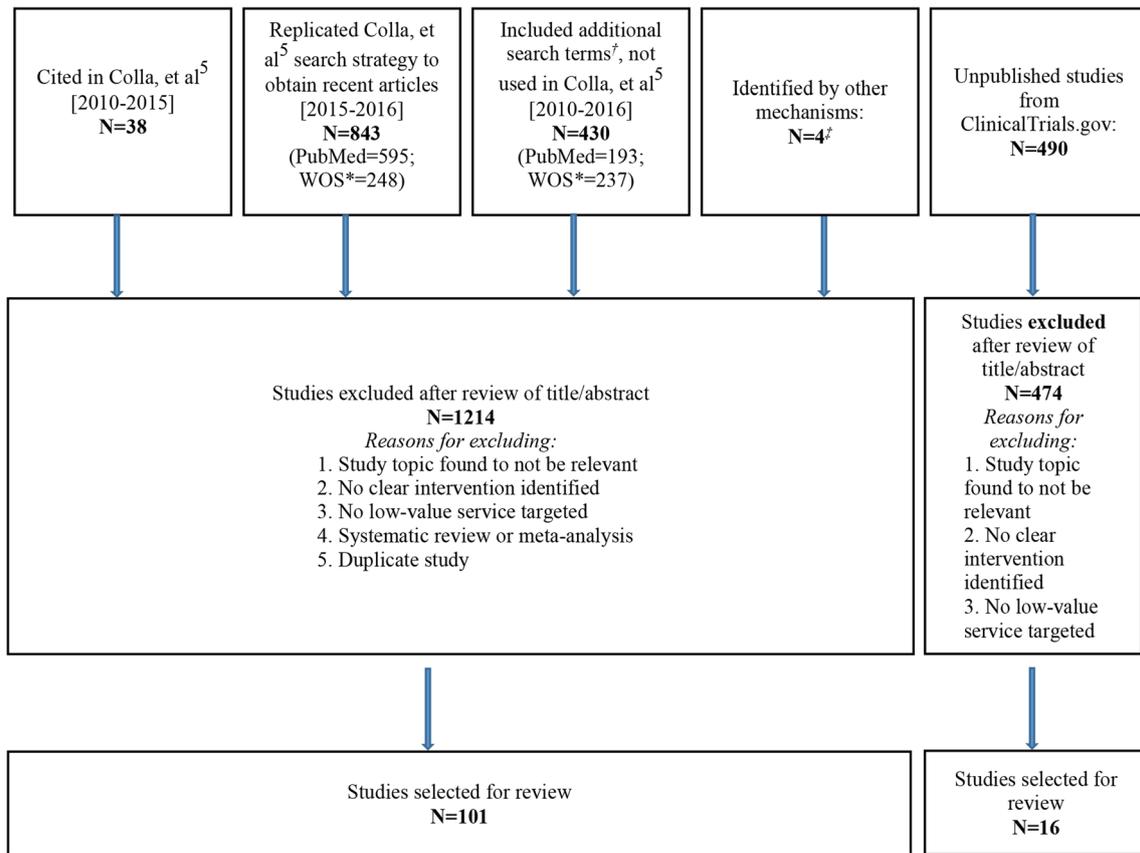


Figure 1 Flow diagram of search results. \*WOS Web of Science. †De-prescribing, de-intensification, unnecessary (lab, test, imaging, utilization), performance measurement, and behavioral economics. ‡2 articles identified as published studies in [clinicaltrials.gov](http://clinicaltrials.gov), 2 articles identified by AcademyHealth from the Choosing Wisely Campaign.

**Data Extraction and Analysis: Measures**

Four physicians (EK, SS, JM, SB) reviewed full-text articles to extract data using a standardized data extraction form. We extracted measure specifications (count, scale, or proportion), measure type (using the framework described above), and whether or not the measure assessed unintended consequences (Table 1). Unintended consequences were categorized as “definite” if they were explicitly stated as such (or a synonymous term or concept clearly identified it as such) in the methods section of the study, or “possible” if they were not specifically stated as such by the study authors, but measurement of an unintended consequence was inferred by reviewers. For example, if a study designed to decrease use of antibiotics for viral infections also examined number of subsequent emergency department (ED) visits, we coded ED visits as a measure of unintended consequences even if the authors did not specifically label it as such. Duplicate review was performed on all measures. Discrepancies were resolved by discussion to reach consensus.

**Data Extraction and Analysis: Study Characteristics**

Three authors (SL, MK, WF) also independently extracted data from each full-text article that met inclusion criteria, to

identify study characteristics. We extracted the following study characteristics: publication year, setting (e.g., inpatient, outpatient), study population, clinical function (screening, prevention, treatment, or diagnostic testing and surveillance), type of service(s) which were intended to be reduced (medications, labs/pathology, imaging, procedures, cost, other), who initiated the intervention (payer, delivery system, other), whether randomization was used, and whether a control group was employed. We also extracted information regarding the intervention target (e.g., patient, provider, system) and mechanism (e.g., cost sharing, pay for performance, insurance payment policy). Finally, we extracted funding source, if any, from all published studies. Any uncertainty about study characteristics was resolved through discussion to reach consensus. As prior systematic reviews performed quality assessments on many of the studies included in our review, we did not include a quality assessment and chose to focus primarily on the characteristics of measures within these studies.

**RESULTS**

**Results from Published Studies**

*Characteristics of Studies.* We identified 1311 original manuscripts from PubMed and Web of Science. Including

**Table 1 Framework for Measure Categorization**

Measure specifications
1. Count
2. Scale
3. Proportion
a. Numerator
b. Denominator
c. Exclusions
4. Other
Measure type
1. Appropriateness
a. Overuse
b. Underuse
c. Misuse
d. Other
2. Utilization/ordering
3. Outcomes (including intermediate)
4. Patient reported
a. Patient-reported outcome measures (PROM)
b. Patient-reported experience measures (PREM)
c. Patient preferences
5. Provider-reported experience
6. Patient-provider interaction
7. Value (outcome/cost)
8. Cost
9. Other
Measure of unintended consequences type (if applicable)
1. Substitution of an alternative low-value service
2. Underuse of the service being intervened upon
3. Underuse of related services
4. Patient-reported experiences
5. Provider-reported experiences
6. Patient-provider interaction
7. Patient selection
8. Care location shift
9. Harm (outcome)
10. Reimbursement
11. Other

the additional two manuscripts that were identified by our partners in AcademyHealth and two that were identified from completed studies in [ClinicalTrials.gov](http://ClinicalTrials.gov), our search strategy resulted in 1315 original manuscripts (Fig. 1). After

**Table 2 Characteristics of Published<sup>a</sup> and Ongoing<sup>b</sup> Studies**

Study characteristics	Number of published studies (N, % of total articles [101])	Number of ongoing studies (N, % of total articles [16])
Used randomization	19 (19%)	13 (81%)
Employed a control group	30 (30%)	12 (75%)
Setting		
Inpatient only	42 (42%)	7 (44%)
Outpatient only	32 (32%)	5 (31%)
Other (e.g., ED, multiple)	27 (27%)	4 (25%)
Clinical function		
Screening	12 (12%)	1 (6%)
Prevention	7 (7%)	1 (6%)
Treatment	47 (47%)	14 (88%)
Diagnostic testing or surveillance	54 (53%)	4 (25%)
Low-value service targeted by intervention <sup>†</sup>		
Medications	32 (32%)	12 (75%)
Labs/pathology	34 (34%)	2 (13%)
Cost	15 (15%)	0
Imaging	26 (26%)	2 (13%)
Procedures	14 (14%)	1 (6%)
Other	6 (6%)	2 (13%)

<sup>a</sup>Ongoing studies include those from [ClinicalTrials.gov](http://ClinicalTrials.gov)

<sup>b</sup>Published studies include those from PubMed and Web of Science

<sup>†</sup>Studies may include more than one

reviewing these by title and abstract, we excluded 1214 due to irrelevant study topic, no identifiable intervention and/or low-value care targeted, manuscript comprising either a systematic review or meta-analysis, or duplicate study. In total, 101 manuscripts met inclusion criteria.

These 101 studies were conducted in a variety of practice settings, including inpatient only (42%), outpatient only (32%), and other or multiple settings (27%). Few studies (30%) employed an external control group, and only 19% used randomization. Clinical functions included screening (12%), prevention (7%), treatment (47%), and diagnostic testing or surveillance (53%), with some studies covering more than one domain. The low-value services targeted in these studies included laboratory testing (including pathology) (34%), medication use (32%), imaging tests (26%), cost of care (15%), and medical or surgical procedures (14%) (Table 2). Forty-eight studies (48%) were externally funded (e.g., through grants). Twenty-five of these 48 studies (52%) received federal funding (US or non-US). A complete list of studies included in this review is provided in [Online Appendix C](#), and additional details for each study are provided in [Online Appendix D](#).

**Characteristics of Interventions.** Most interventions targeted providers, with 60% of studies using education or guidelines, 44% using clinical decision support tools, 42% using feedback mechanisms (e.g., report cards), and 3% using pay-for-performance. A small number of studies (13%) comprised interventions that focused on patients, of which the majority were education-related. Finally, a minority of studies (5%) included interventions that targeted payers (e.g., changes to insurance or payment policy).

**Characteristics of Measures Within Studies.** Most studies used at least one measure of utilization or ordering, while fewer used measures of appropriateness, outcomes, or cost.

**Table 3 Summary of Measures from Published<sup>a</sup> and Ongoing<sup>b</sup> Studies**

Type of measure	Published studies	Ongoing studies
	Studies with at least one of following measure types (N, % of total articles [101])	Studies with at least one of following measure types (N, % of total [16])
Appropriateness	53 (52%)	8 (50%)
Utilization/ordering	69 (68%)	11 (69%)
Outcome	41 (41%)	12 (75%)
Patient-reported	8 (8%)	8 (50%)
Provider-reported	3 (3%)	2 (13%)
Patient-provider interaction	1 (1%)	1 (6%)
Value	1 (1%)	0
Cost	36 (36%)	3 (19%)
Other	3 (3%)	0
Measures of unintended consequences	34 (34%)	10 (63%)

<sup>a</sup>Published studies include those from PubMed and Web of Science.

<sup>b</sup>Ongoing studies include those from [ClinicalTrials.gov](http://ClinicalTrials.gov)

Patient reports (including PREMs and PROMs), provider reports, measures of the patient-provider interaction, and measures of value (e.g., cost-effectiveness) were used infrequently (Table 3). A summary of measures from each published study and specific examples of measures from selected studies can be found in [Online Appendix E](#) and [Online Appendix F](#), respectively.

**Utilization Measures.** Sixty-nine studies (68%) included a measure of utilization of care, including use of antibiotics/medications, laboratory testing, imaging tests, medical procedures, consultation of specialty services, screening tests or procedures, and transfusion of blood products. These were largely collected as overall proportions and rates of service utilization across units or clinics, often determined pre- and post-intervention, without considering whether the service was indicated in specific circumstances. For example, due to the rising rate of antibiotic resistance among the neonatal population, Nitsch-Osuch and colleagues studied the impact of a hospital antibiotic policy on overall antibiotic use in the neonatal intensive care unit (ICU).<sup>15</sup> Similarly, given the overuse of laboratory testing, Procop and colleagues compared the impact of two clinical decision support tools (“Hard Stop,” which required a call to the laboratory to provide justification for ordering a duplicate test, and “Smart Alert,” which was simply a notification to the ordering provider that a duplicate test was ordered) by measuring the number of duplicate laboratory tests placed.<sup>16</sup>

**Appropriateness Measures.** Fifty-three studies (52%) included a measure of appropriateness. These measures assessed appropriate and inappropriate use, including overuse of a variety of services and procedures among specific patients for whom the service was not indicated. For example, several studies examined overuse of medications among patients not requiring treatment, such as antibiotics prescribed for patients with viral upper respiratory infections or asymptomatic bacteriuria, or acid suppressive therapy among low-risk inpatients.<sup>17–19</sup> Other studies examined inappropriate use of other treatments (e.g., transfusions of red blood cells in patients with hematocrit greater than 21% or fresh frozen plasma among patients who did not have a prolonged INR and/or active bleeding), imaging modalities (e.g., MRI for back pain in patients without red flags, or neuroimaging for uncomplicated headaches), and medical procedures (e.g., rate of cesarean sections in low-risk births).<sup>20–23</sup> Finally, appropriateness measures also included rates of adherence to guidelines. For example, Kost and colleagues assessed clinical decisions made by primary care physicians before and after the implementation of Choosing Wisely® in common scenarios such as antibiotic prescribing for acute sinusitis or use of dual-energy x-ray absorptiometry (DEXA) for bone loss among low-risk women.<sup>24</sup>

**Outcome Measures.** Forty-one studies (41%) included an outcome measure. These included length of stay, hospitalization, ICU admission, hospital readmission, and mortality. For example, Algaze and colleagues found that a computerized order entry rule reduced laboratory utilization but did not affect hospital length of stay or mortality among pediatric cardiovascular ICU patients.<sup>25</sup> Outcome measures also included adherence to recommendations, delayed care, and treatment failures. For example, outpatient parenteral antibiotic therapy has been shown to be overused among patients with a variety of infectious diseases despite national guidelines.<sup>26</sup> Due to ongoing overuse of such therapy, Conant and colleagues investigated the impact of mandatory infectious disease approval for outpatient parenteral antibiotic therapy. The authors found a low rate of treatment failures among patients for whom authorization of outpatient therapy was denied.<sup>27</sup>

**Cost-Related Measures.** Thirty-six studies (36%) assessed the impact of interventions on costs of care. Measures included cost related to medications, inpatient admissions, laboratory tests, procedures, specialty services, imaging, and blood product transfusions. For example, in one study, global payment contracts between payer and provider organizations were found to decrease spending for cardiovascular services and imaging.<sup>28</sup> In another study, a computerized integrated antibiotic authorization system reduced expenditure related to antibiotic overuse.<sup>29</sup>

**Patient-Reported Measures.** Eight studies (8%) included patient-reported measures, such as PREMs and PROMs. Examples of patient-reported measures included patient satisfaction and quality of life.<sup>30,31</sup> Funded studies ( $N = 48$ ) were more likely to use patient-reported measures than unfunded studies (17% versus 0%).

**Other Measures.** Three studies (3%) included provider-reported measures, including physicians’ intention to follow practice guidelines, physicians’ feedback and experiences related to an intervention, and self-reported comfort level with procedures such as hysteroscopy.<sup>32–35</sup> Only one study included a measure of value (cost-effectiveness of alternative coronary heart disease diagnostic testing strategies)<sup>30</sup>, and one study included a measure of patient-provider interaction.<sup>22</sup>

**Measures of Unintended Consequences Within Studies.** A total of 34 studies included at least one measure of an unintended consequence. Across these 34 studies, 75 out of the total 349 (21%) measures that we identified in our review assessed unintended consequences including measures of appropriateness, utilization, outcomes, and patient-reported experiences. Fifteen studies included a total of 30 measures that we categorized as “definite” measures of unintended consequences (i.e., stated explicitly in the study), and the

**Table 4 Measures of Unintended Consequences from Published<sup>a</sup> and Ongoing<sup>b</sup> Studies**

Type of measure	From published studies (N, % of total [75])	From ongoing studies (N, % of total [22])
Appropriateness	2 (3%)	0
Utilization/ordering	5 (7%)	2 (9%)
Outcome	65 (87%)	17 (77%)
Patient-reported	3 (4%)	3 (14%)
Provider-reported	0	0
Patient-provider interaction	0	0
Value	0	0
Cost	0	0
Other	0	0

<sup>a</sup>Published studies include those from PubMed and Web of Science.

<sup>b</sup>Ongoing studies include those from [ClinicalTrials.gov](https://clinicaltrials.gov)

remaining 19 studies included a total of 45 measures that we categorized as “possible” measures of unintended consequences (i.e., inferred by the reviewer).

Of the 75 measures, the majority (87%) assessed outcomes (e.g., length of stay or mortality). The remaining assessed utilization or ordering (7%), appropriateness (3%), and patient-reported outcomes and experiences (4%). No studies used measures of provider-reported experiences, patient-provider interactions, value, or costs to assess unintended consequences (Table 4).

### Results from Ongoing Studies

In addition to the studies identified from PubMed and Web of Science, we identified 490 potentially relevant studies from [ClinicalTrials.gov](https://clinicaltrials.gov) (Fig. 1) of which 16 met inclusion criteria for review (Online Appendix C). These studies were conducted in a variety of practice settings including inpatient only (44%), outpatient only (31%), and other or multiple settings (25%). Seventy-five percent of the studies employed a control group and 81% of the studies used randomization. Clinical functions included screening (6%), prevention (6%), treatment (88%), and diagnostic testing or surveillance (25%), with some studies covering more than one domain. Low-value services that were targeted include medications (75%), laboratory testing (including pathology) (13%), imaging (13%), and procedures (6%) (Table 2). All of the studies used interventions targeting providers, and 4 studies used interventions targeting patients. None of these ongoing studies targeted payers.

Twelve studies (75%) included at least one outcome measure, and 11 (69%) included at least one utilization measure. These were followed by 8 studies (50%) with an appropriateness measure, 8 studies (50%) with a patient-reported measure, 3 (19%) with a cost measure, 2 (13%) with a provider-reported measure, and 1 (6%) with a patient-provider interaction measure (Table 3). None of the studies included a measure of value.

Twenty-two out of 103 measures (21%) assessed unintended consequences, with 10 studies (63%) measuring at least one unintended consequence. Three were “definite” measures of

unintended consequences while 19 were “possible.” Of the 22 measures, the majority (77%) assessed unintended consequences related to outcomes. The remainder assessed utilization or ordering (9%) or were patient-reported measures (14%). No studies used measures of appropriateness, cost, provider-reports, patient-provider interaction, or value to assess unintended consequences (Table 4).

### DISCUSSION

In this systematic review, we found that across 101 studies of interventions to reduce low-value care, more than two-thirds focused on rates of utilization. About half of the studies examined changes in appropriateness of services, a more clinically meaningful measure for studies intending to reduce low-value care. Studies rarely used patient-reported measures, provider-reported measures, or measures of patient-provider interactions. Funded studies were more likely to use patient-reported measures (17% vs 0%). Finally, only one-third of studies included a measure of unintended consequences. Even then, many of these unintended consequences assessed readily available but rare outcomes such as mortality.

Among the published intervention studies, less than one-third used randomization or a control group. Perhaps unsurprisingly, ongoing clinical trials registered in [ClinicalTrials.gov](https://clinicaltrials.gov) were more likely to use randomization (81% vs 19%) or an external control group (75% vs 30%) than published manuscripts (which included not only clinical trials, but also quality improvement studies). Like the published literature, most of the ongoing trials used measures of utilization (69%), but a greater proportion used outcome measures (75%) and patient-reported measures (50%). Provider-reported and patient-provider interaction measures were used rarely. Unlike published studies, over half (63%) of the ongoing trials included measures of unintended consequences. However, the sample size for ongoing studies was small ( $N = 16$ ), limiting our ability to draw robust conclusions.

These findings highlight critical gaps in the way we measure the effects of interventions to reduce low-value care. As we have discussed, such interventions are often complex, comprising multiple components that are tested in active healthcare delivery contexts which include a broad array of stakeholders. As a result, these interventions can have unexpected and (in some instances) unintended effects on clinical processes and outcomes as well as patient and provider experiences and outcomes. It is therefore imperative that the evaluations of these interventions consider broader sets of measures to assess the interventions’ effects, both positive and negative.

To this end, we present a comprehensive and systematic framework to assess measures that researchers may apply to their studies to reduce low-value care. In particular, researchers should incorporate more patient-centered measures to ensure that the right services are being reduced in the right patients, that patient-provider relationships are assessed, and

that outcomes are improved (or do not worsen). Measures that focus on appropriateness, patient reports, and clinically meaningful outcomes that the intervention is adequately powered to assess are particularly important. In addition, unintended consequences need to be assessed routinely to ensure that interventions are reducing low-value care without promoting harm. In our review, unintended consequences were assessed in only a small proportion of published studies. Moreover, in both published and ongoing studies, most were measures of rare outcomes (e.g., mortality) for which the study was likely underpowered. Studies infrequently assessed unintended effects of processes of care and intermediate outcomes, such as underuse of recommended services, care location shift, increasing use of an alternative test or treatment (substitution), damage to the patient-physician relationship, or patient or provider dissatisfaction.<sup>7,8</sup>

Strengths of our study include a comprehensive literature search of both published and ongoing studies and development of a framework to characterize measures. Despite efforts to clearly define measure categories, variation in classification among the reviewers was expected. However, we did conduct a duplicate review of all measures for both published and ongoing studies with discrepancies resolved through discussion.

There are several limitations to our study. First, we did not perform a quality assessment of the studies themselves. However, a quality assessment that included many of our studies was recently published by Colla and colleagues.<sup>6</sup> Second, despite a comprehensive literature search, there remains a possibility that we may have missed relevant studies, but it is likely that the number of omissions was small. Finally, data from the ongoing, unpublished studies should be interpreted with caution given the small number and the limited details that were available in the study protocols. In addition, the measures that were reported in [ClinicalTrials.gov](http://ClinicalTrials.gov) may or may not ultimately be used in the published evaluation of these studies.

In conclusion, our findings confirm our hypotheses that (1) existing studies largely focus on utilization rather than more clinically meaningful measures, such as appropriateness, (2) unintended consequences are often not systematically assessed, and (3) patient-reported measures are used infrequently. Study designers and evaluators should explicitly incorporate more clinically meaningful and patient-centered measures into study designs. Additionally, researchers and funding agencies should develop standardized guidance for study designs of interventions to reduce low-value care. Finally, editors and reviewers should request that patient-centered measures be included in evaluations (or their absence be clearly mentioned in limitations). As we seek to develop and test increasingly complex interventions to reduce low-value care, we must be sure to comprehensively assess the effects of these interventions on the totality of clinical care and the patient experience.

**Acknowledgments:** The authors would like to thank Dr. Brahmajee Nallamothu, MD, MPH, and Dr. Jacob Kurlander, MD, MS, for their valuable input in drafting this manuscript.

**Corresponding Author:** Jennifer K. Maratt, MD, MS; Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, MI, USA (e-mail: jkrai@umich.edu).

**Funders** This review was commissioned by AcademyHealth and the American Board of Internal Medicine Foundation as part of a PCORF-funded initiative to identify and prioritize future directions related to measuring low-value care.

#### Compliance with Ethical Standards:

**Conflict of Interest:** The authors declare that they do not have a conflict of interest.

## REFERENCES

1. **de Vries EF, Struijs JN, Heijink R, Hendriks RJ, Baan CA.** Are low-value care measures up to the task? A systematic review of the literature. *BMC Health Serv Res.* 2016;16:1–12.
2. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: The National Academies Press; 2001.
3. **Baker DW, Gaseem A, Reynolds PP, Gardner LA, Schneider EC.** Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med.* 2012;158:55–59.
4. American Board of Internal Medicine. *Choosing Wisely: An Initiative of the ABIM Foundation.* Available at: <http://www.choosingwisely.org> [Accessed February 12, 2019].
5. **Kerr EA, Kullgren JT, Saini SD.** Choosing wisely: how to fulfill the promise in the next 5 years. *Health Aff.* 2017;36:2012–18.
6. **Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N.** Interventions aimed at reducing use of low-value health services: a systematic review. *Med Care Res Rev.* 2016;74:507–50.
7. **Schwartz AL, Landon BE, Elshaug AG, Cherner ME, McWilliams JM.** Measuring low-value care in medicare. *JAMA Intern Med.* 2014;174:1067–76.
8. **Bhatia RS, Levinson W, Shortt S, et al.** Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care. *BMJ Qual Saf.* 2015;24:523–31.
9. Mathias JS, Baker DW. Developing quality measures to address overuse. *JAMA.* 2013;309:1897–98.
10. **Brody H.** Medicine's ethical responsibility for health care reform—the top five list. *N Engl J Med.* 2010;362:283–85.
11. **Larochelle MR, Knight AM, Pantle H, Riedel S, Trost J.** Reducing excess cardiac biomarker testing at an Academic Medical Center. *J Gen Intern Med.* 2014;29:1468–74.
12. **Pittenger K, Williams BL, Mecklenburg RS, Blackmore CC.** Improving acute respiratory infection care through nurse phone care and academic detailing of physicians. *J Am Board Fam Med.* 2015;28:195–204.
13. National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties.* Available at: [http://www.qualityforum.org/Publications/2011/01/Measure\\_Testing\\_Task\\_Force.aspx](http://www.qualityforum.org/Publications/2011/01/Measure_Testing_Task_Force.aspx). [Accessed February 12, 2019].
14. **Chan KS, Chang E, Nassery N, Chang HY, Segal JB.** The state of overuse measurement: a critical review. *Med Care Res Rev.* 2013;70:473–96.
15. **Nitsch-Osuch A, Kurpas D, Kuchar E, Zycinska K, Zielonka T, Wardyn K.** Antibiotic consumption pattern in the neonatal special care unit before and after implementation of the hospital's antibiotic policy. *Adv Exp Med Biol.* 2015;835:45–51.
16. **Procop GW, Keating C, Stagno P, et al.** Reducing duplicate testing: a comparison of two clinical decision support tools. *Am J Clin Pathol.* 2015;143:623–26.
17. **Jenkins TC, Irwin A, Coombs L, et al.** Effects of clinical pathways for common outpatient infections on antibiotic prescribing. *Am J Med.* 2013;126:327–335.e12.

18. **Kelley D, Aaronson P, Poon E, McCarter YS, Bato B, Jankowski CA.** Evaluation of an antimicrobial stewardship approach to minimize overuse of antibiotics in patients with asymptomatic bacteriuria. *Infect Control Hosp Epidemiol.* 2014;35:193–95.
19. **Gupta R, Marshall J, Munoz JC, Kottoor R, Jamal MM, Vega KJ.** Decreased acid suppression therapy overuse after education and medication reconciliation. *Int J Clin Pract.* 2013;67:60–65.
20. **Butler CE, Noel S, Hibbs SP, et al.** Implementation of a clinical decision support system improves compliance with restrictive transfusion policies in hematology patients. *Transfusion.* 2015;55:1964–71.
21. **Thakkar RN, Lee KHK, Ness PM, et al.** Relative impact of a patient blood management program on utilization of all three major blood components. *Transfusion.* 2016;56:2212–20.
22. **Fenton JJ, Kravitz RL, Jerant A, et al.** Promoting patient-centered counseling to reduce use of low-value diagnostic tests. *JAMA Intern Med.* 2016;176:191–7.
23. **Chaillet N, Dumont A, Abrahamowicz M, et al.** A cluster-randomized trial to reduce cesarean delivery rates in Quebec. *N Engl J Med.* 2015;372:1710–21.
24. **Kost A, Genao I, Lee JW, Smith SR.** Clinical decisions made in primary care clinics before and after choosing wisely. *J Am Board Fam Med.* 2015;28:471–4.
25. **Algaze CA, Wood M, Pageler NM, Sharek PJ, Longhurst CA, Shin AY.** Use of a checklist and clinical decision support tool reduces laboratory use and improves cost. *Pediatrics.* 2016;137:1–7.
26. **Tice AD, Rehm SJ, Dalovisio JR, et al.** Practice guidelines for outpatient parenteral antimicrobial therapy. *Clin Infect Dis.* 2004;38:1651–72.
27. **Conant MM, Erdman SM, Osterholzer D.** Mandatory infectious diseases approval of outpatient parenteral antimicrobial therapy (OPAT): clinical and economic outcomes of averted cases. *J Antimicrob Chemother.* 2014;69:1695–1700.
28. **Song Z, Fendrick AM, Safran DG, Landon B, Chernew ME.** Global budgets and technology-intensive medical services. *Healthcare (Amst).* 2013;1:15–21.
29. **Potasman I, Naftali G, Grupper M.** Impact of a computerized integrated antibiotic authorization system. *Isr Med Assoc J.* 2012;14:415–19.
30. **Ripley DP, Brown JM, Everett CC, et al.** Rationale and design of the Clinical Evaluation of Magnetic Resonance Imaging in Coronary heart disease 2 trial (CE-MARC 2): A prospective, multicenter, randomized trial of diagnostic strategies in suspected coronary heart disease. *Am Heart J.* 2015;169:17–24.
31. **Tannenbaum C, Martin P, Tamblyn R, Benedetti A, Ahmed A.** Reduction of inappropriate benzodiazepine prescriptions among older adults through direct patient education. *JAMA Intern Med.* 2014;174:890–8.
32. **Legare F, Labrecque M, Cauchon M, Castel J, Turcotte S, Grimshaw J.** Training family physicians in shared decision-making to reduce the overuse of antibiotics in acute respiratory infections: a cluster randomized trial. *CMAJ.* 2012;184:E726–34.
33. **Lin IB, Coffin J, O'Sullivan PB.** Using theory to improve low back pain care in Australian Aboriginal primary care: a mixed method single cohort pilot study. *BMC Fam Pract.* 2016;17:1–14.
34. **Iturrate E, Jubelt L, Volpicelli F, Hochman K.** Optimize your electronic medical record to increase value: reducing laboratory overutilization. *Am J Med.* 2016;129:215–20.
35. **Leung S, Leyland N, Murji A.** Decreasing diagnostic hysteroscopy performed in the operating room: a quality improvement initiative. *J Obstet Gynaecol Can.* 2016;38:351–56.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.