Obesity

Construction and Use of Body Weight Measures from Administrative Data in a Large National Health System: A Systematic Review

Ann Annis [1,2], *Michelle B. Freitag*[1], *Richard R. Evans* [1], *Wyndy L. Wiitala*[1], *Jennifer Burns*[1], *Susan D. Raffa*[3,4], *Stephanie A. Spohr*[3], *and Laura J. Damschroder* [1]

Objective: Administrative data are increasingly used in research and evaluation yet lack standardized guidelines for constructing measures using these data. Body weight measures from administrative data serve critical functions of monitoring patient health, evaluating interventions, and informing research. This study aimed to describe the algorithms used by researchers to construct and use weight measures.

Methods: A structured, systematic literature review of studies that constructed body weight measures from the Veterans Health Administration was conducted. Key information regarding time frames and time windows of data collection, measure calculations, data cleaning, treatment of missing and outlier weight values, and validation processes was collected.

Results: We identified 39 studies out of 492 nonduplicated records for inclusion. Studies parameterized weight outcomes as change in weight from baseline to follow-up (62%), weight trajectory over time (21%), proportion of participants meeting weight threshold (46%), or multiple methods (28%). Most (90%) reported total time in follow-up and number of time points. Fewer reported time windows (54%), outlier values (51%), missing values (34%), or validation strategies (15%).

Conclusions: A high variability in the operationalization of weight measures was found. Improving methods to construct clinical measures will support transparency and replicability in approaches, guide interpretation of findings, and facilitate comparisons across studies.

Obesity (2020) **28**, 1205-1214.

Study Importance

What is already known?

- Body weight measurements are routinely collected from patients during clinical encounters and regularly documented in electronic health records systems; however, the construction and use of weight measures for program evaluation and research vary substantially.
- Although previous systematic reviews have examined associations between weight-management programs and patient outcomes, we did not identify any prior review that summarized approaches used by researchers to operationalize and analyze measures of body weight from administrative data.

What does this review add?

- Among 39 studies reviewed, we found high variability in the reported construction and use of body weight measures, with many studies lacking key information needed to compare findings or replicate analyses.

How might these results change the direction of research?

- The availability of consistent, replicable methods of weight measurement is critical for longitudinal monitoring of patient health.
- Improving methodologies for constructing weight measures will support more robust evidence building, transparency in reporting, and replicable science.

Introduction

Background

Body weight measurements are routinely collected from patients during clinical encounters and documented in electronic health record (EHR) systems, and they are readily accessible by health care providers to manage patients' health. Collectively, body

[1] Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, Michigan, USA. Correspondence: Ann Annis (annisann@msu.edu) [2] College of Nursing, Michigan State University, East Lansing, Michigan, USA [3] National Center for Health Promotion and Disease Prevention, Veterans Health Administration, Durham, North Carolina, USA [4] Department of Psychiatry & Behavioral Sciences, Duke University School of Medicine, Durham, North Carolina, USA.

See Commentary, pg 1174.

Received: 17 November 2019; **Accepted:** 11 February 2020; **Published online 1 June 2020. doi:10.1002/oby.22790**

weight measures documented in EHRs can be utilized for several critical functions, including monitoring patient health outcomes, evaluating health-related programs and interventions, assessing quality of care, and research. The construction and utilization of body weight measures from EHRs vary substantially in reported literature.

The Veterans Health Administration (VHA) is the largest integrated health system in the United States and is one of the earliest developers and users of a unified EHR starting in the early 1990s (1). Data from the EHR are extracted nightly and uploaded into the Department of Veterans Affairs Corporate Data Warehouse (CDW) (2). The availability of these relational data tables provides administrators, managers, and researchers opportunities to conduct evaluations of population health, programs, or treatments.

VHA developed the MOVE! Weight Management Program for Veterans (MOVE!), a robust program to address the growing prevalence of obesity among veterans. A recent systematic review examined studies that reported on the quality and effectiveness of MOVE!, often using weight-related outcome measures in analyses (3). However, little attention has been paid to how weight measures are determined using EHR data. The absence of such guidelines limits the ability to compare findings across studies and over time. Given the growing emphasis on the use of pragmatic trials (4) that rely on administrative data and the need to generate reproducible and transparent analytic methods (5), researchers are increasingly challenged to make their work accessible and replicable. To better understand the diversity of weight measurement methods, we undertook a review of weight measurement processes documented in published literature.

Objective

The objective of this systematic review was to describe the range of body weight measurement algorithms documented in studies using VHA data. Findings from this review will advance the ability to evaluate and test weight algorithms. This work was guided by VHA policy pertaining to nonresearch activities (6). Institutional review board approval was not required because we relied on publicly available materials.

Methods

Research questions

We conducted a systematic literature review to identify the various methodologies researchers utilize to construct body weight measures in VHA studies. Our review followed recommendations in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (7). We were guided by the following two questions: (1) how do researchers define, operationalize, and analyze body weight measures in VHA studies, and (2) what is the range of variation in defined body weight measures across studies?

Literature search strategy and sources

An exploratory literature search was conducted in February 2018, and a large collection of relevant articles was referred to the team by VHA colleagues. These two sources (exploratory search and referrals) served as a preliminary review to focus our objectives, define inclusion and exclusion criteria for the studies, and develop tools to abstract key information from articles. A structured literature search was subsequently conducted in October 2018 using the research database of the National Center for Biotechnology Information/PubMed.

Our focus was on use of body weight measures derived from VHA data systems; thus, we limited our review to studies that used data generated by VHA in the course of caring for veteran patients. Our search strategy specified the following terms: ("veterans health"[MeSH term] OR "veterans"[MeSH term] OR "hospitals, veterans"[MeSH term]) AND ("body weight" OR "body mass index" OR "obesity"). The search excluded animal research, conference proceedings, and systematic reviews.

Eligibility criteria

We selected articles published in peer-reviewed scholarly journals between 2008 and 2018. Articles were considered eligible for inclusion if the study reported use of body weight measures as an outcome (dependent variable) in analyses and if weight measures relied on existing electronic data originally collected for nonresearch purposes.

An article was excluded when (1) the authors did not report measurement and/or use of a body weight measure, (2) it was not original research (e.g., opinion piece, study protocol), (3) it did not focus on veteran populations, (4) weight measures were self-reported, (5) VHA administrative data were not used, or (6) weight was not an outcome in analyses.

Data collection

We extracted study design details, participant inclusion and exclusion criteria, sample size, years of data extraction, aim or purpose of the weight measure(s), and data source of the relevant weight measure from the identified studies. When study design was not explicitly reported by the authors, determination was guided by the literature (8). We also documented whether a weight measure was used as a criterion for sample selection (e.g., study included only individuals with overweight).

Weight measure construction, analysis, and utilization were recorded as well as any described algorithms. Key elements included (1) total time in the study follow-up time period, (2) number of unique time points considered for analyses, (3) time intervals (the length of time between each data measurement time point; e.g., baseline to 6 months of follow-up), (4) time windows (the length of time surrounding a time point in which observations are included; e.g., a 1-month window before/after baseline measurement), (5) missing data (number or percentage), (6) treatment of missing data, (7) definition of outliers or biologically implausible values, (8) treatment of outliers, (9) units of measurement, (10) weight measure calculation (continuous, dichotomous, or categorical variable), and (11) validity assessment of weight data.

Analyses and synthesis of results

Two team members (MF and AA) completed an initial exploratory review and summarized findings. All coauthors provided input on development of the analytic plan for the full review. Eligible abstracts were screened (AA and MF). An in-depth full-text review of articles was conducted for those meeting initial inclusion criteria (AA and MF).

To ensure validity and inter-rater reliability of the review process, two rounds of verification based on a 10% random sample of studies meeting inclusion criteria and a 10% sample of those excluded were reviewed by at least one other team member (LD, JB, RE, WW). The reviewers extracted key data elements, which were compared with the initial abstraction. The larger team then discussed extracted data, clarified key elements, and further refined the scope of review. Decisions pertaining to sampling criteria, as well as collection and analysis of key data from the studies, were based on team consensus.

Upon completing the review of all studies, a third verification round (using the same process from the first two rounds) was conducted, yielding 92% inter-rater agreement on included/excluded studies. Final decisions were made based on team consensus. Abstracted data from included articles were aggregated by content-specific categories across studies for each of the key elements described. Results were synthesized and summarized into a matrix. Quality of the studies was assessed using selected items from the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (9).

Results

Literature review sample

We identified 492 nonduplicated records (Figure 1) and eliminated 346 articles via abstract screening. Full-text review was completed for 146 articles plus 6 additional studies identified from citations and not previously included. After full-text review, we excluded 107 articles, the majority of which were eliminated because of primary data collection for weight measures ($n=37$) and/or weight was not an outcome ($n=44$). The final sample included 39 studies (Table 1). The majority (87%) described an observational, retrospective cohort study design (Tables 1-2). Most studies assessed weight outcomes to evaluate weight management programs (41%) (10-25); the remaining studies evaluated postbariatric surgery weight changes (25-35). Sample sizes ranged from 40 (29) to 4,990,424 (36), with most studies (62%) including more than 1,000 patients.

Data sources

Several VHA data sources for weight measures were listed by researchers. Most stated they used national or regional data sources, including the VHA CDW, National Patient Care Database, or other unspecified VHA administrative data sets (49%) (10,16-17,19-21,27,36-45). Two additional studies (5%) in this group reported using a Veterans Integrated Services Network database with unclear linkage to the national CDW (22,46). Other studies (44%) described using a facility EHR or related systems (Veterans Health Information Systems and Technology Architecture or Computerized Patient Record System) (11,13-15,24-26,28-35,47,48). The remaining studies were not specific about source or how data were extracted, listing MOVE! program data (15%) (10,12,14,16,18,23) or VHA data (8%) (10,14,16).

Time factors

The time frames from which the studies obtained weight measures spanned from 1997 to 2017 (Table 1). With the exception of one cross-sectional study (36), all studies established a baseline time period

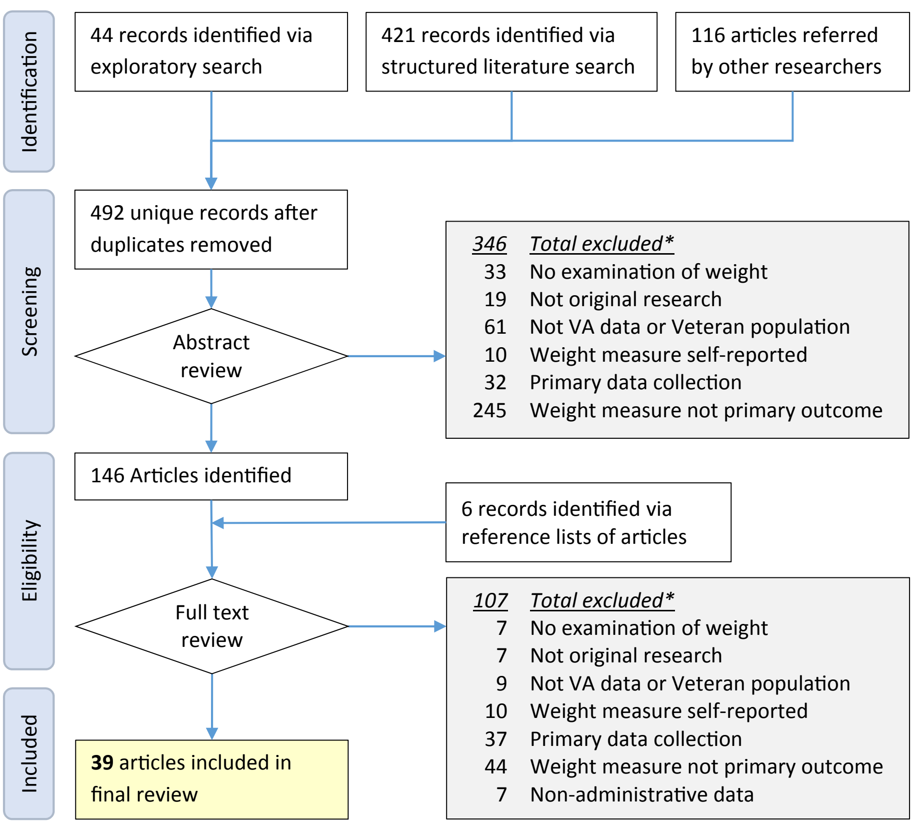

Figure 1 Flow diagram of literature review. *Articles can meet multiple exclusion criteria; thus, categories will not sum to total.

TABLE 1 Characteristics of studies included in systematic review ($n = 39$)

Author	Study design	Study years	Sample size	Population	Aim/purpose
Adams (2012)	Retrospective cohort	2003-2008	61	Veterans with obesity who received bariatric surgery, age < 65 years	Describe prevalence of tobacco and substance use and associated factors that predict weight loss among bariatric patients
Arterburn (2013)	Retrospective cohort	2000-2006	516	Veterans with obesity who received bariatric surgery, age not specified	Identify determinants of weight loss post bariatric surgery
Baker (2015)	Retrospective cohort	2003-?	1,474	Veterans with rheumatoid arthritis, age > 18 years	Identify factors associated with long-term changes in BMI
Batch (2018)	Retrospective cohort	2008-2013	62,882	MOVE! participants, age not specified	Determine patterns of change in weight among MOVE! participants
Bounthavong (2014)	Retrospective cohort	2006-2010	51,977	Veterans with type 2 diabetes, age ≥ 18 years	Evaluate differences in BMI among veterans with type 2 diabetes mellitus
Braun (2016)	Retrospective cohort	2007-2008	1,659	MOVE! participants, age < 70 years	Evaluate degree of weight loss among MOVE! participants
Breland (2017)	Cross-sectional	1999-2016	4,990,424	Veterans with a primary care visit, age not specified	Describe obesity prevalence among subpopulations of VHA patients
Buta (2018)	Retrospective cohort	2001-2010	248,089	OEF/OIF/OND veterans, age not specified	Examine whether posttraumatic stress disorder is associated with change in BMI
Chan (2017)	Longitudinal	2004-2014	237,577	MOVE! participants, age not specified	Examine the relationship between MOVE! participation and weight outcomes
Copeland (2012)	Retrospective cohort	2001-2006	254,051	Veterans with obesity, age not specified	Assess receipt of obesity care by patients with and without mental illness
Garvin (2015)	Retrospective cohort	2008-2010	216	MOVE! participants, age < 90 years	Examine the effect of high-intensity MOVE! on achieving at least 5% weight reduction.
Garvin & Hardy (2016)	Retrospective cohort	2008-2010	375	MOVE! participants, age not specified	Determine whether early weight reduction in MOVE! predicted later participation or achievement of weight reduction goals
Garvin & Marion (2015)	Retrospective cohort	2008-2010	404	MOVE! participants, age < 90 years	Explore background and program characteristics associated with a 5% weight reduction in MOVE!
Goodrich (2016)	Retrospective cohort	2007-2013	11,188	MOVE! participants with serious mental illness diagnosis, age 18-69 years	Investigate whether there were sex differences in weight loss and program participation in MOVE!
Grabarczyk (2017)	Retrospective cohort	2012-2016	66,035	Veterans with overweight or obesity, age not specified	Compare the effectiveness of weight management medications
Hauser (2010)	Retrospective cohort	2003-2006	70	Veterans with obesity who received bariatric surgery, age 29-66 years (range, not inclusion criteria)	Evaluate the long-term outcomes following gastric bypass
Hoerster (2014)	Retrospective cohort	2007-2012	20,819	MOVE! participants, age not specified	Compare the effectiveness of MOVE! among veterans with posttraumatic stress disorder, other mental health conditions, and no mental health diagnoses
Huerta (2008)	Retrospective cohort	2004-2006	40	Veterans with obesity who received bariatric surgery, age not specified	Assess whether preoperative weight loss resulted in favorable outcomes after gastric bypass
Huizinga (2010)	Retrospective cohort	2001-2007	18,205	Veterans who had received an oral antidiabetic drug, age ≥ 18 years	Compare the effect of oral diabetes therapy regimens on BMI

TABLE 1 (continued).

Author	Study design	Study years	Sample size	Population	Aim/purpose
Ikossi (2010)	Retrospective cohort	2001-2007	24 cohort and 78 control	Veterans with obesity who received bariatric surgery, with and without posttraumatic stress disorder, age not specified	Evaluate the effect of posttraumatic stress disorder on excess body weight loss after gastric bypass
Jackson (2015)	Retrospective cohort	2005-2012	238,540 MOVE! participants and 1,606,257 matched nonparticipants	MOVE! participants and nonparticipants, age < 70 years	Investigate whether MOVE! participation is associated with reduced incidence of diabetes
Janney (2016)	Retrospective cohort	2008-2012	84,770	MOVE! participants, age 18-69 years	Investigate the influence of sleep disordered breathing on weight loss
Kahwati (2011)	Case control	2008-2010	31,904 MOVE! participants and 71,967 nonparticipants	MOVE! participants and nonparticipants, age not specified	Explore variation in MOVE! program implementation to identify facility structure, policies, and processes associated with larger patient weight loss outcomes
Kazerooni (2016)	Retrospective cohort	2000-2014	767	Veterans who were new topiramate users, age not specified	Determine whether topiramate use results in significant weight loss.
Littman (2012)	Cross-sectional and longitudinal	2005-2008	3,192 MOVE! participants and 73,407 nonparticipants	MOVE! participants and nonparticipants, age 18-69 years	Assess the reach and effectiveness of MOVE!
Littman (2015)	Retrospective cohort	2008-2012	207,131	Random sample of veterans with a MOVE! visit, age < 70 years	Determine whether obesity screening and MOVE! participation and outcomes are equitable for patients with serious mental illness and depressive disorder compared with those without
Maciejewski (2016)	Retrospective cohort	1999-2004	1,787 cohort and 5,305 controls	Veterans with obesity who received bariatric surgery and nonsurgical matched controls, age not specified	Examine weight change in a large, multisite, clinical cohort of veterans who received gastric bypass compared with nonsurgical matched patients
Maguen (2013)	Retrospective cohort and longitudinal	2001-2011	496,722	Veterans with a clinical visit after the end of their last deployment, age not specified	Explore the relationship between BMI and posttraumatic stress disorder and evaluate trajectories of BMI change
Noel (2012)	Retrospective cohort and longitudinal	2002-2006	223,246	Veterans with obesity, age ≥ 18 years	Determine whether obesity-related education or a weight management program was associated with declines in BMI
Pandey (2018)	Retrospective cohort	2004-2014	213,985	Women veterans, age ≥ 18 years	Analyze the association of military sexual trauma and posttraumatic stress disorder with obesity
Romanova (2013)	Retrospective cohort	2006-2009	382	MOVE! participants, age not specified	Evaluate the effectiveness of MOVE! in achieving weight loss
Rosenberger (2011)	Retrospective cohort	2001-2009	16,656	Veterans with sufficient BMI values, age not specified	Determine BMI trajectories and sociodemographic factors associated with BMI trajectory
Rutledge (2011)	Retrospective cohort	1998-2007	60	Veterans with obesity who received bariatric surgery, age ≤ 65 years	Assess relationship between patient's number of psychiatric conditions and weight loss surgery outcomes
Rutledge (2012)	Longitudinal	1999-2007	55	Veterans with obesity who received bariatric surgery, age ≤ 65 years	Examine longitudinal psychiatric treatment changes and changes in weight and weight-related comorbidities after bariatric surgery

TABLE 1 (continued).

Author	Study design	Study years	Sample size	Population	Aim/purpose
Shi (2011)	Retrospective cohort	2006-2008	13,293	Veterans with type 2 diabetes mellitus, age not specified	Investigate the impact of publicized safety warnings for thiazolidinediones on glycemic outcomes in patients with diabetes mellitus
Skancke (2018)	Retrospective cohort	2010-2017	100	Veterans with obesity who received bariatric surgery, age not specified	Benchmark the outcomes of gastrectomy based on recent guidelines
Win (2014)	Retrospective cohort	2004-2010	79	Veterans with obesity who received bariatric surgery, age not specified	Explore the relationship between the number of nutrition visits and change in BMI after gastric bypass surgery
Xiao (2017)	Retrospective cohort	1998-2014	1,077	Veterans with a new diagnosis of lymphoma, age not specified	Describe long-term and short-term weight change trends, identify factors associated with weight gain, and determine whether weight change during chemotherapy affects time to next treatment and survival
Yan (2008)	Retrospective cohort	1997-2002	59	Veterans with obesity who received bariatric surgery, age not specified	Determine success of gastric bypass as measured by changes in excess weight loss

MOVE!, Weight Management Program for Veterans; OEF, Operation Enduring Freedom; OIF, Operation Iraqi Freedom; OND, Operation New Dawn.

TABLE 2 Characteristics of studies included in systematic review (*n* = 39)

	n	%
Study design[a]		
Retrospective cohort	34	87.2%
Longitudinal	5	12.8%
Cross-sectional	2	5.1%
Case control	1	2.6%
MOVE! / TeleMOVE! Study		
Yes	16	41.0%
Sample size		
< 100 patients	7	17.9%
100-999 patients	8	20.5%
1,000-49,999	9	23.1%
50,000-99,999	5	12.8%
≥ 100,000	10	25.6%
Description of data source[a]		
National/regional: CDW, NPCD, VISN, VHA administrative data sets	19	48.7%
Local: facility data, EHR, VISTA, CPRS	17	43.6%
Program-related or not specific: MOVE! data, VHA data	6	15.4%
Total time in follow-up time period		
< 12 months	5	12.8%
1 year	11	28.2%
2-4 years	12	30.8%
≥ 5 years	7	17.9%
Not reported	1	2.6%
Number of time points in data collection		
1-2	14	35.9%
3-4	10	25.6%
5-6	4	10.3%
7-9	3	7.7%
≥ 10	4	10.3%
All available (multiple) weight values within time period	6	15.4%
Not reported	1	2.6%
Time intervals reported[b]	35	89.7%
Time windows reported[c]	18	46.2%

[a]Sum more than total because of studies reporting multiple categories.
[b]Time intervals are amount of time between each data collection time point, for example, baseline, 6 months after, and 12 months after baseline.
[c]Time windows correspond to amount of time surrounding data collection time point in which observations are included. For example, 6-month time point may include observations within 1 month prior or 1 month after 6-month time point.
CDW, Corporate Data Warehouse; CPRS, Computerized Patient Record System; EHR, electronic health record; MOVE!, Weight Management Program for Veterans, NPCD, National Patient Care Database; TeleMOVE!, Home Telehealth Weight Management Program for Veterans; VHA, Veterans Health Administration; VISN, Veterans Integrated Services Networks; VISTA, Veterans Health Information Systems and Technology Architecture.

to be compared with a follow-up period. The baseline weight measure was typically assigned based on the date of an index event, such as a first MOVE! visit, date of bariatric surgery, or clinic visit. The majority (90%) of studies reported the total time in the follow-up period (90%), most spanning a period of 1 year (28%) (11-12,16,18,20,22-23,27,30,40,41) or 2 to 4 years (31%) (14-15,19,24,26,29,33-34,38,42,45) (Table 2).

Similarly, almost all studies (90%) reported the number of time points from which data were collected. Most often, studies focused on one or two time points (36%) (10-13,21,27,29-30,34,36,38-40,46). The duration of time between the assessed time points (time intervals) varied across studies but most often included intervals of 6 months (e.g., 6, 12, or 18 months after baseline).

Just under half (46%) of studies described the time windows used to determine point-in-time weights, with wide variability in the width of the window as follows: 2 weeks (37,45), 30 days or 1 month (12,20-21,23-24,41,45), 2 months (12,16,18,20-21,23), 3 months (17,24,41,45), 5 to 7 months (10,19,22,41), 10 to 15 months (11,22,33,36,40), or longer (31). Several studies reported using more than one time window, often using different windows to define baseline weights versus follow-up weights (12,20-24,41,45). Typically, when multiple weights were found within the defined window, the weight that was recorded closest to the relevant time point was retained for use in analyses. Six studies did not use windows but instead used all available weight values to assess weight trajectories (11-12,31,42-43,47), one of which calculated a quarterly median value of weights over time (43).

Measure construction

There were wide differences in how researchers defined and constructed outcomes (Table 3). Studies used BMI (49%) or weight in pounds (41%) or kilograms (13%). Parameterization of weight outcomes included assessing change in weight from baseline to a follow-up time point (62%) (11,16-18,20-26,28-35,37,40-41,45,46) or weight trajectory over time (23%) (11-12,27,31,38,42-43,47,48) as continuous variables. Several studies (46%) created a binary indicator to classify participants meeting a specific weight threshold (e.g., at least 5% weight loss) and then calculated the proportion of participants meeting this threshold (10,12-15,17-18,20-23,27,31,36,39,41,44,45). Eleven (28%) studies used more than one approach. For example, Maciejewski et al. (31) determined the percentage change in weight at follow-up as compared with baseline, predicted the percentage of weight change at several follow-up time periods, and assessed whether patients lost a specified percentage of their baseline weight.

TABLE 3 Construction of weight measures among studies included in systematic review ($n = 39$)

Characteristic[a]	*n*	%
Unit of measurement		
Weight in pounds	16	41.0%
Weight in kilograms	5	12.8%
BMI	19	48.7%
Use of weight variable in analyses		
Change in weight compared with baseline	24	61.5%
Weight trajectory over time	9	23.1%
Proportion of participants meeting weight loss/gain threshold	18	46.2%
Missing weight measures		
Number or percentage of missing values reported	14	35.9%
Not reported	23	59.0%
Not applicable (e.g., because of sampling)	2	5.1%
Treatment of missing weight measures		
Exclusion of individual	15	38.5%
Exclusion of observation	4	10.3%
Not reported	22	56.4%
Outlier definition		
Absolute values	12	30.8%
Relative, calculated, and/or data derived values	8	20.5%
Not reported	25	64.1%
Treatment of outliers		
Exclusion of individual	2	5.1%
Exclusion of observation	12	30.8%
Not reported	25	64.1%
Validation strategies reported	6	15.4%

[a]Studies may report more than one category; thus, categories are not mutually exclusive.

Missing values

Only 14 (36%) studies reported the number or percentage of missing weight values (12,16,20-23,26-27,31,36-37,41,43,47) (Table 3). Missing values were handled by excluding the patient from the analyses (39%) (12,16,19,21-23,27,31,36-37,41,43-44,47,48) or excluding only the relevant missing weight value (10%) (20,26,37,43). Two studies in particular described using a combination of both methods (5%) (37,43). Notably, among studies that reported exclusion at the level of the patient, several assessed their cohorts longitudinally over multiple years with repeated measures. Thus, it is likely that the exclusion occurred when all the weight values were absent for an individual participant rather than excluding a participant based on any single value that was missing for a particular visit. The reported methods often did not allow us to distinguish between exclusions made because all values were missing versus exclusions because one or more values were missing. Finally, two studies described using all available weight values over a specified time period (44,48).

Outlier values

Most studies (64%) did not identify outlier values (Table 3). Of those that did, some (31%) used specific cutoff values based on lower and upper thresholds (12,16,18,20-23,36,40,42-43,47). For example, Maguen et al. (42) and Noel et al. (43) identified 70 lb and 700 lb as lower and upper thresholds, respectively. Fewer studies (21%) used a more complex process for determining outliers based on statistical properties of the analytic sample (10,12,18,20-22,31,36). For example, Chan et al. (12) computed mean weight by individual and defined outliers as individual values more than 3 SDs from an individual's mean. Similarly, Maciejewski et al. (31) calculated rolling SD of consecutive weight values in groups of three and then used these as thresholds for determining outliers. Breland et al. (36) determined outlier values by defining specific upper- and lower-bound threshold values of the sequential ratios of current weight to previous weight and current weight to next weight. Six studies used both simple thresholds and more complex statistical approaches (12,18,20-22,36). Outliers were addressed by excluding the participant from the analyses (5%) or excluding only the outlier value from analyses (31%).

Validity

Validation of weight measures was rarely reported (15% of studies). One study (27) conducted a sensitivity analysis that varied the threshold value for weight change over time, while another (43) performed sensitivity analysis of a 5% weight change threshold. Additional validation strategies included alternative analyses by adjusting statistical models (20), comparing results to other similar studies (16), comparing participants with one weight measure to those with multiple weight measures (42), and examining within-subject changes in the trajectory of the weight measure (47).

Study quality

Study quality and risk of bias was evaluated in terms of the construction of the weight outcome measure and the reported number of relevant attributes (e.g., time factors, outlier values). Most studies (56%) described in some level of detail at least half of the selected attributes. However, only 10 (26%) studies met an excellent level of quality for all relevant measure attributes. Many (46%) had a substantial lack of reported attributes for the weight measure, which hinders an adequate assessment of the validity of findings.

Discussion

We found high variability in the reported construction and use of body weight measures among VHA studies, with many lacking key information needed to compare findings or replicate analyses. Although most studies (90%) reported the amount of time in the follow-up period and the number of time points assessed, just over half reported time windows (54%) or identified outlier values (51%), and fewer reported missing weight values (34%) or validation strategies (15%). Interestingly, 11 of the 39 reviewed studies constructed more than one weight measure as outcome variables in analyses, while 8 studies identified more than one time window, and 6 studies used multiple definitions for outlier values. Thus, the variation in measurement specification was not only present across studies but within studies as well. Often, studies lacked a rationale or justification for including multiple sets of defined measurement criteria.

Though our search was not limited to studies of weight change for veterans with overweight or obesity, most (41%) were evaluations of MOVE!, 28% evaluated veterans after bariatric surgery, and 10% focused on receipt of obesity-related care within VHA. Thus, the implications of our findings are especially relevant for assessments of patient weight in the context of weight loss interventions. Moreover, EHR and associated data warehouses within integrated health systems are generating larger and more comprehensive data sources that include many health-related measures beyond weight (e.g., blood pressure, cholesterol). The issues highlighted in this systematic review are likely to be prevalent for these additional measures as well. Importantly, the availability of consistent, replicable methods of measurement over time are particularly valuable in longitudinal monitoring of health status, an approach frequently used in population-based studies.

Despite the frequent use of weight measures in research studies and program evaluations, our findings highlight the absence of standardized guidelines for defining, constructing, utilizing, and reporting weight measurement approaches, pointing to the need for more consistent methodologies. This is a key issue, especially in light of increasing use of administrative data for program evaluation and the rise of pragmatic clinical trials (4) that often rely on these data sources. In order to foster the use of valid and reliable measurement approaches among clinical trials and evaluations, the data infrastructure to support such research should be strengthened by standardized data extraction and analytic processes (49). The development of standardized measurement approaches for weight data should consider employing suggested methods for handling missing data (50,51) and outliers (52,53). In a previous review of reported treatments for missing data from 169 articles published in the journal *Prevention Science*, researchers found that deletion-based techniques (i.e., listwise [exclusion of individuals] and pairwise [exclusion of observations] deletion) were the most commonly reported methods for handling missing data, used by 30% of the studies (54). Moreover, based on how the studies reported missing data, the authors concluded that this proportion was likely higher and closer to 37% (54), similar to our finding of 44% of studies. Deletion-based techniques, however, are considered to be poor solutions for addressing missing data (55), as there are more sophisticated methods available.

Multiple imputation in particular is becoming increasingly used and recommended as a way to decrease bias because of missing data (54,56,57). The review mentioned above identified the use of imputation strategies by 21% of studies (54). In contrast, none of the 39 studies included in this review used imputation methods, although in many cases, we were unable to distinguish whether missing data techniques were not employed or not reported. Certainly, there are circumstances whereby other methods of handling missing data may be more appropriate than imputation (58). However, the lack of detail reported by studies that pertain to the treatment of missing data is a common problem, even among those using imputation methods (50-51,56). Regardless of the method selected, studies would benefit from additional reporting clarity and transparency about how missing data were addressed.

Parallel to the issue of missing data, the majority (64%) of studies we reviewed did not state how outlier weight values were defined or handled. This is consistent with a previous review that found that 41% of 42 large epidemiologic studies did not address biologically implausible values, while an additional 26% reported insufficient information regarding implausible values (59). Similar to our approach, the authors categorized biologically implausible values into the following three groups: externally defined limits (i.e., comparisons to values obtained outside of the study, or “absolute” values), internally defined limits (i.e., defined based on the values of the study sample, or “relative” values), and using a combination of both methods (59).

Among the studies we reviewed that did report outliers, there was wide variability in outlier definitions in both the “absolute” and “relative” groups. For example, multiple cutoff values were used by our sample to define plausible weight values, ranging from 50 to 91 lb for a minimum weight and from 600 to 823 lb for a maximum weight value. Several variations of BMI cutoff values were also reported. Similarly, the studies that used relative approaches also differed in their strategies to identify outlier values, such as using calculated participant mean weights, SDs, weight ratios, and/or weight change. Interestingly, six of the studies used a combination of approaches. Numerous statistical approaches to address outlier or implausible values have been described elsewhere, along with recommendations for selecting and utilizing an appropriate method (52,53). Yet, the finding that only 14 (36%) distinct studies in our sample described any information related to outliers indicates, again, that there are opportunities to improve reporting of basic elements of data analyses.

Lastly, our sample of studies was decidedly split between constructing the weight outcome as a continuous versus categorical or binary variable. In fact, a substantial number of our reviewed studies (46%) created a binary indicator to designate whether or not a study participant met a weight threshold. As evident from the variability in the construction of weight outcomes, there does not appear to be a consistent practice for defining weight measures or defining a clinically meaningful change in weight. Although several studies considered meaningful weight change as a loss in excess weight of 5% or more, rationale for this threshold was not reported. Lack of consensus for operationalizing weight outcomes limits the ability to compare and contrast findings across studies. Moreover, categorizing or dichotomizing continuous variables can lead to a loss of information and biased results (60). Further exploration of weight measures will require careful consideration of what constitutes clinically meaningful weight change and how measurement approaches can be constructed to optimally detect these changes while limiting the risk of bias.

Robust science requires transparency and replicability, which allow others to validate and potentially extend findings. Researchers are increasingly requested by funding agencies and peer-reviewed journals to share data and provide clear descriptions of analytic methods. Several standards for reporting research studies have been published by expert working groups to guide such documentation (61-64). These guidelines should extend to the need for clear articulation of methods used to construct and analyze health system–generated data.

This study had some limitations. Our review of weight algorithms was limited to VHA administrative data sources and studies of Veteran populations. VHA has long emphasized the importance of assessing and documenting weight for screening, preventing, and treating obesity (21,65). Additionally, large relational databases of clinical and administrative data fed by one of the earliest integrated EHR systems in the world provide a robust data infrastructure to support quality improvement, program evaluation, and research studies. Thus, our findings may not be generalizable to studies outside VHA using non-VHA data sources. We did not verify the data sources reported by researchers and noted variation in the terminology they used to describe these sources. Thus, it is possible some reported sources were misclassified. Our specification of search terms and/or exclusion criteria may have led to missed opportunities to further understand approaches to derive weight-related measures. Lastly, the lack of reporting on a key data element in an article does not necessarily indicate absence but rather may simply reflect space constraints for published research and lack of priority for detailed documentation.

Conclusion

Routinely collected clinical measures of patient health, including body weight, captured by administrative systems are important sources for outcome measures in research and program evaluation. The high variability found in measurement approaches, along with the expanding availability of large data systems, challenges researchers to identify reliable, valid, and consistent methodologies for constructing measures based on administrative data. Improving these methods will support more robust evidence building, transparency in reporting, and replicable science. Future research should examine whether specific analytic approaches in constructing weight measures can potentially add value to the validity or robustness of the analyses and/or negatively impact data collection or other logistical aspects of the study.**O**

Acknowledgments

We would like to thank our colleagues Jane Kim, Michael Goldstein, Sophia Califano, Margaret Dundon, Sophia Hurley, Eugene Oddone, and Felicia McCant for ongoing support for this work, including reviewing drafts of this manuscript and providing valuable feedback.

Funding agencies: This work was supported by the National Center for Health Promotion and Disease Prevention, VHA. The views expressed in this article are those of the authors and do not necessarily reflect the policy or position of the Department of Veterans Affairs or the US Government.

Disclosure: The authors declared no conflict of interest.

References

1. Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA–U.S. Department of Veterans Affairs national-scale HIS. *Int J Med Inform* 2003;69:135-156.
2. VA Information Resource Center. *VIReC Resource Guide: VA Corporate Data Warehouse*. Hines, IL: U.S. Dept. of Veterans Affairs; 2012.
3. Maciejewski ML, Shepherd-Banigan M, Raffa SD, Weidenbacher HJ. Systematic review of behavioral weight management program MOVE! for veterans. *Am J Prev Med* 2018;54:704-714.
4. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464-475.
5. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol* 2016;14:e1002333. doi:10.1371/journal.pbio.1002333
6. Veterans Health Administration. *VHA Operations Activities That May Constitute Research*. Washington, DC: Veterans Health Administration; 2011.
7. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006-1012.
8. Andrews J, Likis FE. Study design algorithm. *J Low Genit Tract Dis* 2015;19:364-368.
9. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22. doi:10.1186/1471-2288-10-22
10. Batch BC, Goldstein K, Yancy WS, et al. Outcome by gender in the Veterans Health Administration motivating overweight/obese veterans everywhere weight management program. *J Womens Health (Larchmt)* 2018;27:32-39.
11. Braun K, Erickson M, Utech A, List R, Garcia JM. Evaluation of veterans MOVE! program for weight loss. *J Nutr Educ Behav* 2016;48:299-303.e1.
12. Chan SH, Raffa SD. Examining the dose-response relationship in the veterans health administration's MOVE!((R)) weight management program: a nationwide observational study. *J Gen Intern Med* 2017;32(suppl 1):18-23.
13. Garvin JT. Weight reduction goal achievement with high-intensity MOVE!((R)) treatment. *Public Health Nurs* 2015;32:232-236.
14. Garvin JT, Hardy D, Xu H. Initial response to program, program participation, and weight reduction among 375 MOVE! participants, Augusta, Georgia, 2008-2010. *Prev Chronic Dis* 2016;13:E55. doi:10.5888/pcd13.150598
15. Garvin JT, Marion LN, Narsavage GL, Finnegan L. Characteristics influencing weight reduction among veterans in the MOVE!(R) Program. *West J Nurs Res* 2015;37:50-65.
16. Goodrich DE, Klingaman EA, Verchinina L, et al. Sex differences in weight loss among veterans with serious mental illness: observational study of a national weight management program. *Womens Health Issues* 2016;26:410-419.
17. Grabarczyk TR. Observational comparative effectiveness of pharmaceutical treatments for obesity within the Veterans Health Administration. *Pharmacotherapy* 2018;38:19-28.
18. Hoerster KD, Lai Z, Goodrich DE, et al. Weight loss after participation in a national VA weight management program among veterans with or without PTSD. *Psychiatr Serv* 2014;65:1385-1388.
19. Jackson SL, Long Q, Rhee MK, et al. Weight loss and incidence of diabetes with the Veterans Health Administration MOVE! lifestyle change programme: an observational study. *Lancet Diabetes Endocrinol* 2015;3:173-180.
20. Janney CA, Kilbourne AM, Germain A, et al. The influence of sleep disordered breathing on weight loss in a national weight management program. *Sleep* 2016;39:59-65.
21. Kahwati LC, Lance TX, Jones KR, Kinsinger LS. RE-AIM evaluation of the Veterans Health Administration's MOVE! Weight Management Program. *Transl Behav Med* 2011;1:551-560.
22. Littman AJ, Boyko EJ, McDonell MB, Fihn SD. Evaluation of a weight management program for veterans. *Prev Chronic Dis* 2012;9:E99. doi:10.5888/pcd9.110267
23. Littman AJ, Damschroder LJ, Verchinina L, et al. National evaluation of obesity screening and treatment among veterans with and without mental health disorders. *Gen Hosp Psychiatry* 2015;37:7-13.
24. Romanova M, Liang LJ, Deng ML, Li Z, Heber D. Effectiveness of the MOVE! multidisciplinary weight loss program for veterans in Los Angeles. *Prev Chronic Dis* 2013;10:E112. doi:10.5888/pcd10.120325
25. Skancke M, Schoolfield C, Grossman R, Kerns JC, Abel N, Brody F. Laparoscopic sleeve gastrectomy for morbid obesity at a Veterans Affairs Medical Center. *J Laparoendosc Adv Surg Tech A* 2018;28:650-655.
26. Adams CE, Gabriele JM, Baillie LE, Dubbert PM. Tobacco use and substance use disorders as predictors of postoperative weight loss 2 years after bariatric surgery. *J Behav Health Serv Res* 2012;39:462-471.
27. Arterburn D, Livingston EH, Olsen MK, et al. Predictors of initial weight loss after gastric bypass surgery in twelve Veterans Affairs Medical Centers. *Obes Res Clin Pract* 2013;7:e367-e376.
28. Hauser DL, Titchner RL, Wilson MA, Eid GM. Long-term outcomes of laparoscopic Roux-en-Y gastric bypass in US veterans. *Obes Surg* 2010;20:283-289.
29. Huerta S, Dredar S, Hayden E, et al. Preoperative weight loss decreases the operative time of gastric bypass at a Veterans Administration hospital. *Obes Surg* 2008;18:508-512.

30. Ikossi DG, Maldonado JR, Hernandez-Boussard T, Eisenberg D. Post-traumatic stress disorder (PTSD) is not a contraindication to gastric bypass in veterans with morbid obesity. *Surg Endosc* 2010;24:1892-1897.
31. Maciejewski ML, Arterburn DE, Van Scoyoc L, et al. Bariatric surgery and long-term durability of weight loss. *JAMA Surg* 2016;151:1046-1055.
32. Rutledge T, Braden AL, Woods G, Herbst KL, Groesz LM, Savu M. Five-year changes in psychiatric treatment status and weight-related comorbidities following bariatric surgery in a veteran population. *Obes Surg* 2012;22:1734-1741.
33. Rutledge T, Groesz LM, Savu M. Psychiatric factors and weight loss patterns following gastric bypass surgery in a veteran population. *Obes Surg* 2011;21:29-35.
34. Win AZ, Ceresa C, Schafer AL, Mak P, Stewart L. Importance of nutrition visits after gastric bypass surgery for American veterans, San Francisco, 2004-2010. *Prev Chronic Dis* 2014;11:E226. doi:10.5888/pcd11.140289
35. Yan E, Ko E, Luong V, Wang HJ, Romanova M, Li Z. Long-term changes in weight loss and obesity-related comorbidities after Roux-en-Y gastric bypass: a primary care experience. *Am J Surg* 2008;195:94-98.
36. Breland JY, Phibbs CS, Hoggatt KJ, et al. The obesity epidemic in the Veterans Health Administration: prevalence among key populations of women and men veterans. *J Gen Intern Med* 2017;32(suppl 1):11-17.
37. Baker JF, Cannon GW, Ibrahim S, Haroldsen C, Caplan L, Mikuls TR. Predictors of longterm changes in body mass index in rheumatoid arthritis. *J Rheumatol* 2015;42:920-927.
38. Bounthavong M, Tran JN, Golshan S, et al. Retrospective cohort study evaluating exenatide twice daily and long-acting insulin analogs in a Veterans Health Administration population with type 2 diabetes. *Diabetes Metab* 2014;40:284-291.
39. Copeland LA, Pugh MJ, Hicks PB, Noel PH. Use of obesity-related care by psychiatric patients. *Psychiatr Serv* 2012;63:230-236.
40. Huizinga MM, Roumie CL, Greevy RA, et al. Glycemic and weight changes after persistent use of incident oral diabetes therapy: a Veterans Administration retrospective cohort study. *Pharmacoepidemiol Drug Saf* 2010;19:1108-1112.
41. Kazerooni R, Lim J. Topiramate-associated weight loss in a veteran population. *Mil Med* 2016;181:283-286.
42. Maguen S, Madden E, Cohen B, et al. The relationship between body mass index and mental health among Iraq and Afghanistan veterans. *J Gen Intern Med* 2013;28(suppl 2):S563-S570.
43. Noel PH, Wang CP, Bollinger MJ, et al. Intensity and duration of obesity-related counseling: association with 5-Year BMI trends among obese primary care patients. *Obesity (Silver Spring)* 2012;20:773-782.
44. Pandey N, Ashfaq SN, Dauterive EW, MacCarthy AA, Copeland LA. Military sexual trauma and obesity among women veterans. *J Womens Health (Larchmt)* 2018;27:305-310.
45. Xiao DY, Luo S, O'Brian K, Liu W, Carson KR. Weight change trends and overall survival in United States veterans with follicular lymphoma treated with chemotherapy. *Leuk Lymphoma* 2017;58:851-858.
46. Shi L, Zhao Y, Szymanski K, Yau L, Fonseca V. Impact of thiazolidinedione safety warnings on medication use patterns and glycemic control among veterans with diabetes mellitus. *J Diabetes Complications* 2011;25:143-150.
47. Buta E, Masheb R, Gueorguieva R, Bathulapalli H, Brandt CA, Goulet JL. Posttraumatic stress disorder diagnosis and gender are associated with accelerated weight gain trajectories in veterans during the post-deployment period. *Eat Behav* 2018;29:8-13.
48. Rosenberger PH, Ning Y, Brandt C, Allore H, Haskell S. BMI trajectory groups in veterans of the Iraq and Afghanistan wars. *Prev Med* 2011;53:149-154.
49. Horwitz LI, Kuznetsova M, Jones SA. Creating a learning health system through rapid-cycle, randomized testing. *N Engl J Med* 2019;381:1175-1179.
50. Mackinnon A. The use and reporting of multiple imputation in medical research - a review. *J Intern Med* 2010;268:586-593.
51. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393. doi:10.1136/bmj.b2393
52. Lin CH, Hsu KC, Johnson KR, Luby M, Fann YC. Applying density-based outlier identifications using multiple datasets for validation of stroke clinical outcomes. *Int J Med Inform* 2019;132:103988. doi:10.1016/j.ijmedinf.2019.103988
53. Aggarwal CC. *Outlier Analysis*. 2nd ed. Cham, Switzerland: Springer International Publishing AG; 2017.
54. Lang KM, Little TD. Principled missing data treatments. *Prev Sci* 2018;19:284-294.
55. Wilkinson L, Task Force on Statistical Interference. Statistical methods in psychology journals. *Am Psychol* 1999;54:594-604.
56. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol* 2015;15:30. doi:10.1186/s12874-015-0022-1
57. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research–part 2: multiple imputation. *Acad Emerg Med* 2007;14:669-678.
58. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol* 2019;48:1294-1304.
59. Lawman HG, Ogden CL, Hassink S, Mallya G, Vander Veur S, Foster GD. Comparing methods for identifying biologically implausible values in height, weight, and body mass index among youth. *Am J Epidemiol* 2015;182:359-365.
60. Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. *AJNR Am J Neuroradiol* 2011;32:437-440.
61. Agha RA, Altman DG, Rosin D. The SPIRIT 2013 statement–defining standard protocol items for trials. *Int J Surg* 2015;13:288-291.
62. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495-1499.
63. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1-e34.
64. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390. doi:10.1136/bmj.a2390
65. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). *Implement Sci* 2013;8:51. doi:10.1186/1748-5908-8-51