



Published in final edited form as:

Epidemiology. 2016 January ; 27(1): 82–90. doi:10.1097/EDE.0000000000000393.

Learning about missing data mechanisms in electronic health records-based research: a survey-based approach

Sebastien Haneuse¹, Andy Bogart², Ina Jazic¹, Emily O. Westbrook³, Denise Boudreau³, Mary Kay Theis³, Greg E. Simon³, and David Arterburn³

¹Harvard T.H. Chan School of Public Health, Boston, Massachusetts

²RAND Corporation, Santa Monica, California

³Group Health Research Institute, Seattle, Washington

Abstract

Purpose—Bias due to missing data is a major concern in electronic health record (EHR)-based research. As part of an ongoing EHR-based study of weight change among patients treated for depression, we conducted a survey to investigate determinants of missingness in the available weight information and to evaluate the missing-at-random assumption.

Methods—We identified 8,345 individuals enrolled in a large EHR-based health care system who had monotherapy treatment for depression from 04/2008-03/2010. A stratified sample of 1,153 individuals completed a detailed survey. Logistic regression was used to investigate determinants of whether or not a patient (i) had an opportunity to be weighed at treatment initiation (baseline) and (ii) had a weight measurement recorded. Parallel analyses were conducted to investigate missingness during follow-up. Throughout, inverse-probability weighting was used to adjust for the design and survey non-response. Analyses were also conducted to investigate potential recall bias.

Results—Missingness at baseline and during follow-up was significantly associated with numerous factors not routinely collected in the EHR including whether or not the patient had ever chosen not to be weighed, external weight control activities, and self-reported baseline weight. Patient attitudes about their weight and perceptions regarding the potential impact of their depression treatment on weight were not related to missingness.

Discussion—Adopting a comprehensive strategy to investigate missingness early in the research process gives researchers information necessary to evaluate key assumptions. While the survey presented focuses on outcome data, the overarching strategy can be applied to any and all data elements subject to missingness.

Introduction

Electronic health record (EHR) databases offer numerous appealing opportunities for public health research¹⁻³. Relative to data obtained from a typical prospective study, EHR-based

data contain information on a broad range of factors for large patient populations over long timeframes in real-world settings and are relatively inexpensive to obtain⁴⁻⁷. Nevertheless, since EHRs are designed to support clinical and/or billing systems, their use for research purposes requires considerable care. Among the many challenges that researchers face is the extent to which information in the EHR is complete and accurate, and whether or not sufficient information is available to control confounding bias^{6,8-12}.

We currently face these issues in an ongoing EHR-based comparative effectiveness study of treatment for depression and weight change at 2 years post-treatment initiation. The setting for the study is Group Health, a large integrated health insurance and health care delivery system which maintains an EHR (Epic Systems Corporation of Madison, WI). Consistent with prior studies, feasibility assessments during the planning phase indicated wide variation in the number and timing of weight measurements in the EHR, suggesting that a substantial number of patients would have incomplete outcome data^{13,14}. In the presence of incomplete or missing data, a naïve analysis strategy is to restrict to patients with complete data. The corresponding exclusions, however, may result in a form of bias analogous to collider or selection bias that arises in traditional (i.e. non-EHR based) studies that actively recruit patients^{15,16}. To control this form of selection bias, statistical methods for missing data such as multiple imputation¹⁷ and inverse-probability weighting¹⁸ can be used. The validity of these methods, however, relies on the so-called *missing at random (MAR)* assumption. Intuitively, MAR requires that all factors relevant to whether or not a patient has complete data are observed in the EHR. In many EHR-based settings, however, researchers may have good reason to believe that the MAR assumption does not hold. In our study, for example, a clear violation of MAR would be if a patient's weight or recent weight change was a driving force behind whether or not they had a primary care visit at which they could have been weighed, or whether or not a measurement was recorded in the EHR during a visit.

When the MAR assumption does not hold, the data are said to be *missing not at random (MNAR)*, and statistical adjustments will fail to completely resolve selection bias. Unfortunately, whether or not the data are MAR or MNAR is not empirically verifiable given the EHR data alone. In practice, researchers can perform sensitivity analyses to investigate the potential impact of the unobserved factors, although if the results are sensitive the study may be rendered inconclusive. Arguably, the only reliable strategy for evaluating the MAR assumption and establishing the validity of statistical adjustments for selection bias is to perform additional primary data collection. Such data collection may target data elements that are missing (e.g. weight in our comparative study of treatments for depression) and/or target factors hypothesized to be related to missingness (e.g. attitudes towards weight measurement in clinical contexts). With this philosophy in mind we conducted a one-time telephone survey to collect additional detailed information on the missing weight values (i.e. the response in the parent study) and reasons for incomplete data. Here we describe the design and sampling strategy used for the survey, and also report on results from an analysis of the survey data.

Methods

Study Setting

Group Health (GH) provides comprehensive health care on a pre-paid basis to approximately 600,000 individuals in Washington State and Idaho. Information on health plan enrollment, health care use including diagnoses, procedures, pharmacy dispensings, and laboratory values is routinely recorded in GH's electronic databases. In addition, a fully-integrated EHR system documents all outpatient care at GH clinics since 2005.

Study Population

We identified all adults aged 18-65 years with a diagnosis of depressive disorder (ICD-9 = 296.2x, 296.3x, 311, or 300.4) who had at least one new monotherapy treatment episode involving an antidepressant medication (defined as a dispensing for antidepressant medication without any other antidepressant medications or psychotherapy visits) or at least one monotherapy episode of psychotherapy (defined as occurrence of one or more psychotherapy visits without receipt of antidepressant medications) between 04/2008 and 03/2010. Treatment episodes were considered “new” if there was no evidence of treatment in the prior 90 days in the EHR, hence mitigating the inclusion of prevalent users.

Additional details are given in the Supplementary Materials document. For the purposes of this study, we restricted attention to one monotherapy treatment episode for each patient. Furthermore, polytherapy treatment episodes (i.e. 1 antidepressant drug simultaneously or a combination of drug and psychotherapy) were excluded because we believed that it would be more challenging for subjects to recall specific information about treatment episodes that involved multiple treatments.

Only subjects who were continuously enrolled in the GH system since their depression diagnosis were included. We excluded subjects with conditions or treatments known to be associated with significant weight fluctuations including: diagnosis of cancer, psychotic disorders, cognitive impairment, cirrhosis, pregnancy, kidney disease requiring dialysis, and those who had been prescribed an obesity drug treatment or undergone bariatric surgery. Following application of these criteria, a total of N= 8,345 patients were identified and formed the parent study sample.

Electronic Health Record Data

Demographic and enrollment information, prescription medication use, health care encounters, and medical conditions for all eligible subjects were extracted from GH electronic databases and the EHR. All such information between the date of treatment initiation (referred to as “baseline” throughout) and the end of follow-up was extracted. Height, weight, and smoking status were determined from EHR-entered data fields.

Survey – Rationale, Design and Recruitment Process

During the planning phase of the parent study, as we considered the extent of missing data in the EHR, we defined a patient as having “complete” data if they had at least one baseline weight measurement (taken to be any measurement in the 180 days prior to treatment initiation) and at least one follow-up measurement (taken to be any measurement between

treatment initiation and the end of follow-up). Among the N=8,345 patients in the study sample, 5,630 (67.5%) patients had complete data; 2,715 patients (32.5%) had incomplete data and, in principle, could not be directly included in primary analyses for the parent study.

To investigate the potential for the EHR-based weight data to be MNAR, we conducted a one-time telephone survey. The goal was to collect additional detailed information, beyond that readily available in the EHR, on current weight and reasons for incomplete data. We began by stratifying the N=8,345 patients into one of five weight information groups depending on whether or not there was an opportunity to be weighed, at baseline and during follow-up, as well as on whether or not a weight measurement was observed (see Table 1). Each group was further stratified by treatment type, consisting of four strata: psychotherapy, drugs *a priori* hypothesized to be associated with weight gain (mirtazapine and paroxetine), drugs *a priori* hypothesized to be associated with weight loss (bupropion and fluoxetine), and other antidepressant drugs.

When designing the survey, we planned to recruit 200 subjects in each of the five main groups (1,000 total). Invitation letters were sent out in batches of 100 per week. To ensure that certain key sub-populations were well represented in the survey, we initially targeted patients who belonged to strata with fewer members (e.g. those prescribed either mirtazapine or paroxetine). Since this feature of the sampling scheme was by design we were able to adjust for it using appropriate inverse-probability weighting (see below). At the conclusion of the survey, a total of n*=2,109 patients had been invited; n=1,153 patients responded and completed the survey.

Survey – Data Collection

Individuals received an invitation letter with a \$5 bill inviting them to participate in a brief telephone survey and directing them how to opt out of additional contact. Those who did not opt out were contacted by study staff, screened for eligibility, and invited to participate in the survey. A copy of the complete instrument is provided in the Supplementary Materials. Briefly, the survey asked all respondents to assess their current weight as well as their recalled weight at the time of treatment initiation. The survey also asked a series of questions regarding factors hypothesized to be relevant to observance of complete weight data but were not measured in the EHR. These included reasons for not having had a weight assessed in the clinical setting, changes in weight prior to and during depression treatment, the perceived effect of treatment on weight, current depressive symptoms, weight control practices, attitudes toward weight, use of other medications or other health conditions that might have impacted weight, reasons for discontinuing depression treatment, and sociodemographic characteristics.

Statistical Analysis

For characteristics observable in the EHR, we calculated frequency distributions for all N=8,345 patients. We also calculated unadjusted and weighted distributions for all characteristics (i.e. those in the EHR and solely available from the survey) among the n=1,153 survey respondents. The weights, which account for the fact that the n=1,153

survey respondents are not a random sample of the N=8,345 patients in the study sample, consisted of two components. The first was obtained from a model for the probability of being invited to participate in the survey, which included an indicator of which of the five weight information groups the patient belonged to as well as treatment type, and their interaction. Parameter estimates were obtained by fitting this model to the N=8,345 patients in the study sample. The second component was obtained from a model for the probability of survey response, given that the patient was invited. This model included indicators of gender, age, smoking status over the last 9 months, depression diagnosis associated with the treatment episode and whether the patient had a concurrent diagnosis of anxiety or a sleep disorder. Parameter estimates were obtained by fitting this model to the n*=2,109 patients who were invited by the survey team to participate in the survey. Given estimates from these models, fitted values were obtained for the n=1,153 respondents; the two sets of fitted values were multiplied and inverted to form the final weights.

To investigate completeness of weight information in the EHR, we considered four events: (i) a primary care visit must have occurred at baseline or within the prior 180 days, (ii) if a baseline visit occurred, a weight measurement must have been recorded, (iii) a primary care visit must have occurred during follow-up, and (iv) if a follow-up visit occurred, a weight measurement must have been recorded. For each of these we defined a corresponding binary outcome of whether or not the event occurred.

For each of the four outcomes, we compared the estimated distribution of recalled BMI among patients who experienced the event to the corresponding distribution among patients who did not. These estimates were based on the n=1,153 survey respondents with the counts weighted by the inverse probability of survey invitation/completion to ensure generalizability to the full parent study sample. In addition we fit a series of logistic regression models to investigate determinants of completeness, one for each of the four binary outcomes. Each was fit using the n=1,153 survey respondents with estimates obtained via inverse-probability weighting¹⁸, again using the survey invitation/response weights.

Finally, since the survey was retrospective there was the potential for recall bias. To investigate this we compared the observed EHR-based baseline weight with recalled baseline weight among the 478 survey respondents with a measured baseline weight (see Table 1). We calculated bias, defined as the absolute difference between the two measures, as well as percent bias. Comparisons were made graphically as well as via linear regression analyses with bias and percent bias as outcomes. For the latter, we considered all factors available in the EHR and the survey.

Throughout, all analyses were performed in R version 3.1.1, specifically the *survey* package for the weighted analyses^{19,20}. All p-values and confidence intervals were two-sided.

Results

Among the N=8,345 patients identified in the parent study, 6,860 (82%) had had an antidepressant drug treatment episode and 1,485 (18%) a psychotherapy episode (Table 1). Seventy percent were female, and while 55-65 year olds were the largest age group (34%), a

substantial portion (19%) was less than 35 years old at treatment initiation. Most diagnoses were for a major depressive episode (52%); 8% for dysthymia and 40% for non-specific depression. Response rates for the survey were slightly higher among those who had complete weight data in the EHR (58%) versus those who had missing data (range, 53-55%). In models investigating survey response, responders and non-responders differed with respect to age, sex, smoking status in the last 9 months and whether or not the patient had a concurrent diagnosis of an anxiety disorder. Factors not strongly related to survey response were depression diagnosis, treatment type, and availability of weight measures.

Table 2 summarizes the (weighted) distributions of factors *a priori* hypothesized to be related to missingness but not available in the EHR. We see that the majority of survey respondents were white (83%), married (62%), had at least some college education (51%), and were employed full or part time (72%). Fifteen percent reported engagement in a commercial weight loss program in the last 2 years; 30% reported use of meal replacement products, and 25% the use of prescription weight-loss medications. Seventeen percent of survey respondents reported having chosen not to be weighed during a prior clinic visit. At the time of treatment initiation, most patients (55%) were trying to lose weight. While 18% of respondents indicated that they believed their treatment would cause weight gain, most (65%) indicated they believed that their treatment would cause no change in weight.

The final column of Table 2 provides estimated distributions of for all N=8,345 patients in the study sample. That is, it provides a weighted estimate of the distribution that would have been observed had all N=8,345 patients been invited and responded to the survey. While many of the estimated distributions differ somewhat from those observed among the n=1,153 survey respondents, none of the differences are dramatic.

Table 3 provides estimated unadjusted distributions of recalled baseline BMI, by each of the four outcomes. From the first two columns, patients who had a primary care visit at baseline generally had higher BMI than those patients who did not. In contrast, patients who had a weight measurement recorded at baseline generally had lower BMI than patients who did not. That is, at baseline, the missing BMI measurements were generally lower than the observed BMI measurements. Similarly, recalled BMI values were generally lower for patients who had at least one follow-up visit and those who had at least one follow-up weight measurement. In each case, the difference in recalled BMI distributions between those patients with complete data and those with incomplete data is highly statistically significant ($p < 0.001$).

Table 4 reports adjusted odds ratio (OR) estimates and 95% confidence intervals (CI) from the weighted logistic regression analyses for each of the four outcomes. From the first five rows, we see that the suggested associations between recalled BMI and each of the four sub-mechanisms in Table 3 are not corroborated in these adjusted analyses. While some of the estimated ORs suggest an association, there are no clear systematic patterns with recalled BMI and all but two of the component 95% CIs cover the null of no association. Beyond recalled BMI, patients with antidepressant treatment were more likely than psychotherapy to have a primary care visit (ORs ranging from 2.54-3.03 at baseline and 2.36-3.72 during follow-up). With the exception of patients treated with mirtazapine or paroxetine (OR 3.89,

95% CI 1.07-14.21), there was no evidence that treatment type was related to observation of a weight measurement when a primary care visit occurred. Conversely, while female gender did not appear to be related to whether a visit occurred, women were more likely than men to have a weight measurement taken at any given visit (OR 2.81, 95% CI 1.59, 4.97 at baseline; OR 4.03 95% CI 2.06, 7.91 during follow-up). Of the four outcomes, smoking status was only associated with whether the patient had a baseline visit (OR 2.97, 95% CI 1.37, 6.45). Patients' previous refusal to have their weight taken was negatively associated with a weight measurement being taken at both baseline and follow-up, and positively associated with occurrence of a baseline visit. Finally, each weight control activity was associated with having either at least one baseline primary care visit in the EHR or at least one follow-up visit.

Figure 1 compares recalled weight at baseline with observed EHR-based measurements, among patients with both. From the left-hand panel, we see that the two measures are highly correlated (estimated correlation 0.97). While some patients over- or under-report weight, the right-hand panel indicates no evidence of systematic recall bias as a function of actual weight. As discussed in the Supplementary Materials document, linear regression analyses of bias in recalled weight indicate little evidence of differential recall with differences either being not statistically significant and/or small in magnitude.

Discussion

When addressing missing data in EHR-based studies and the form of selection bias it may introduce, researchers must first make an assumption regarding whether or not the available data are MAR and second, if necessary, perform some form of statistical adjustment. In practice, since the MAR assumption can never be empirically verified, it is incumbent upon researchers to critically evaluate it to the extent possible. As researchers do so, however, it may become clear that the available EHR does not contain sufficient information. In this case, the only reliable strategy for assessing missing data assumptions in EHR-based settings is to conduct an internal validation study that supplements the EHR data with: (i) the missing values themselves, and/or (ii) information on factors not available in the EHR that are related to whether a patient has complete data. While the general strategy of collecting additional data to inform assumptions is frequently used in assessing confounding bias or bias due to measurement error/misclassification, we do not believe it to be common in the missing data context. As we have shown, however, adopting such a strategy can give important insights into the complexity of missingness in an EHR, as well as the roles of certain covariates in determining the completeness of a patient's data.

Moving beyond the specific context of this study, a key question is: how should researchers design and conduct their own surveys? Although this is likely to be a complex methodologic challenge, we believe that important lessons can be drawn from our experience with the current survey. Before discussing these, it is useful to consider some limitations specific to our survey. First, to investigate missingness during follow-up, we considered any post-baseline measurement(s) regardless of when they occurred. This was done because, although not reported here, the analyses for the parent study are inclusive in the sense of taking full advantage of as much of the available follow-up weight data as possible. Second, to simplify

the survey we focused on querying baseline factors. As such we did not query patients about different time points (e.g. how their perceptions changed over time), although such information would likely enhance our understanding of missingness. Third, the focus of the survey was missingness in the outcome for the parent study (i.e. weight data). While missingness in other covariates for the parent study was minimal, this will clearly not be the case for all EHR-based studies. Fourth, we did not collect information on a number of factors that could influence missingness including whether the patient had a diagnosis of an eating disorder. This likely limits the interpretation of the substantive results, specifically Tables 2 and 4, although not the overarching strategy. Finally, although we found little evidence of systematic differences between the weight measurements observed in the EHR and recalled values of these weights, we cannot completely rule out the possibility of recall bias.

Considering the aforementioned issues suggests the following recommendations for designing surveys meant to uncover missingness mechanisms. Critical at the outset is developing a framework that outlines the sub-mechanisms that could result in a patient having incomplete data in the EHR and their interplay. In addition to the mechanisms we consider, these could include whether clinical practice standards changed and whether a patient remained continuously enrolled in the health plan, received health care outside the EHR catchment, or died. Second, researchers may need to consider many of these mechanisms at multiple, specific time points. Decisions regarding which time points to consider will be closely related to the design of the parent study (e.g. cross-sectional vs. longitudinal) as well as the nature of the missing data (e.g. the EHR data may be complete at baseline but not during follow-up). Third, researchers need to accommodate the possibility that different sub-mechanisms are driven by different sets of factors, for which both their values at the time under consideration and their history may be of interest. Towards this, the directed acyclic graph framework of Hernan et al¹⁵ could be used to ensure that all variables thought to be relevant to missingness are included. Fourth, by their nature, surveys are subject to both participation bias (i.e. some patients may refuse to participate) and recall bias. Researchers should therefore consider strategies for mitigating recall bias and/or collecting information on factors that may result in differential recall^{21,22}. Fifth, researchers may consider adopting a stratified sampling scheme when identifying patients to be surveyed. Intuitively, stratifying will improve the ability to characterize the impact of certain factors on missingness; in our survey, for example, we chose to stratify on treatment type since we wanted to fully understand the role that the primary exposure of interest for the parent study played in whether or not a patient had complete data. Finally, tied to each of these considerations, is the overarching question of how many patients to survey. Typically, sample size considerations attempt to balance statistical power with budgetary/logistical constraints. Since the primary goal of the proposed survey framework is to learn about missingness mechanisms, statistical power for any single association is, arguably, not relevant. New frameworks for sample size in this context need to be developed and, in the meantime, we recommend collecting as much information as possible within practical constraints.

Finally, as frameworks for the design of surveys in the context of missing data are developed, methods for integrating survey results into the analyses for the parent study will

also need to be developed. Current work in the statistical literature that could be used and/or extended include methods for double-sampling schemes²³ and methods for non-response in case-control studies^{22,24,25}. From a substantive perspective, a recent series of papers by Geng and colleagues describe a survey-based strategy to account for loss to follow-up²⁶⁻²⁸. In those papers outcome information recovered on a sub-sample of patients ostensibly lost to follow-up was incorporated into the main analyses via inverse probability weighting. A distinction between their survey and ours, however, is that they did not collect information specifically to understand why some patients had incomplete data and others not. Nevertheless, the methods they used could be expanded to incorporate rich information collected in the survey^{29,30}. Whichever approach is used, and as methods are developed, a key practical question will be how to select variables for use in the adjustment of selection bias. In principle, one could view this as a variable selection problem and use statistical significance to make decisions. From Table 4, however, we see that there are a number of factors that are suggestive of an association with one or more of the four sub-mechanisms and yet do not achieve statistical significance at the 0.05 level (e.g. employment status, smoking status and whether the patient had an anxiety or sleeping disorder). Practical guidelines are needed to help balance the magnitudes of estimated associations and statistical significance, especially given the inherent bias-variance trade-off that accompanies this decision; a strategy of inclusiveness, for example, will minimize potential bias at the expense of decreasing statistical power if unnecessary variables are included.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was funded under grant R01 MH083671 from the National Institute of Mental Health (NIMH) – Principle investigator: D. Arterburn

References

1. Hernan MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. *Epidemiology*. May; 2011 22(3):290–291. [PubMed: 21464646]
2. Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, D.C: The National Academies Press; 2009.
3. Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. Washington, DC: US Dept. of Health and Human Services; 2009.
4. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology*. Apr; 2005 58(4):323–337. [PubMed: 15862718]
5. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Annals of Internal Medicine*. Sep 1; 2009 151(5):359–360. [PubMed: 19638404]
6. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. Jan 1; 2013 20(1):144–151. [PubMed: 22733976]
7. Gallego B, Dunn AG, Coiera E. Role of electronic health records in comparative effectiveness research. *J Comp Eff Res*. Nov; 2013 2(6):529–532. [PubMed: 24236790]
8. Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clinical Pharmacology and Therapeutics*. Aug; 2007 82(2):143–156. [PubMed: 17554243]

9. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. *Journal of clinical epidemiology*. Jul; 2007 60(7):734–741. [PubMed: 17573990]
10. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev*. Oct; 2010 67(5):503–527. [PubMed: 20150441]
11. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care*. Aug; 2013 51(8 Suppl 3):S30–37. [PubMed: 23774517]
12. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Medical Care*. Aug; 2013 51(8 Suppl 3):S80–86. [PubMed: 23774512]
13. Boudreau DM, Arterburn D, Bogart A, et al. Influence of body mass index on the choice of therapy for depression and follow-up care. *Obesity*. 2013; 21(3):E303–E313. [PubMed: 23404891]
14. Arterburn D, Bogart A, Coleman KJ, et al. Comparative effectiveness of bariatric surgery vs. nonsurgical treatment of type 2 diabetes among severely obese adults. *Obesity Research and Clinical Practice*. 2012; 282:1–11.
15. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. Sep; 2004 15(5):615–625. [PubMed: 15308962]
16. Haneuse S. Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical Care*. Dec 3.2013
17. Rubin DB. Multiple imputation after 18+ years. *Journal of the American Statistical Association*. Jun; 1996 91(434):473–489.
18. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. Sep; 1994 89(427):846–866.
19. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
20. Lumley, T. *Complex Surveys: A Guide to Analysis Using R*. Hoboken, New Jersey: John Wiley & Sons; 2010.
21. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *International journal of epidemiology*. 2014 dyu149.
22. Haneuse S, Chen J. A multiphase design strategy for dealing with participation bias. *Biometrics*. Mar; 2011 67(1):309–318. [PubMed: 20377576]
23. Frangakis CE, Rubin DB. Addressing an idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics*. 2001; 57(2):333–342. [PubMed: 11414553]
24. Scott AJ, Wild CJ. Fitting regression models with response-biased samples. *Canadian Journal of Statistics*. 2011; 39(3):519–536.
25. Jiang Y, Scott AJ, Wild CJ. Adjusting for Non-Response in Population-Based Case-Control Studies. *International Statistical Review*. 2011; 79(2):145–159.
26. Geng EH, Emenyonu N, Bwana MB, Glidden DV, Martin JN. Sampling-based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *Journal of the American Medical Association*. Aug 6; 2008 300(5):506–507. [PubMed: 18677022]
27. Geng EH, Bangsberg DR, Musinguzi N, et al. Understanding reasons for and outcomes of patients lost to follow-up in antiretroviral therapy programs in Africa through a sampling-based approach. *J Acquir Immune Defic Syndr*. 2010; 53(3):405–411. [PubMed: 19745753]
28. Geng EH, Glidden DV, Bangsberg DR, et al. A causal framework for understanding the effect of losses to follow-up on epidemiologic analyses in clinic-based cohorts: the case of HIV-infected patients on antiretroviral therapy in Africa. *American Journal of Epidemiology*. May 15; 2012 175(10):1080–1087. [PubMed: 22306557]
29. Rotnitzky A, Scharfstein D, Su TL, Robins J. Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*. 2001; 57(1):103–113. [PubMed: 11252584]

30. Molenberghs G. Missing Data: Sensitivity Analysis. Encyclopedia of Statistical Sciences. 2007

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

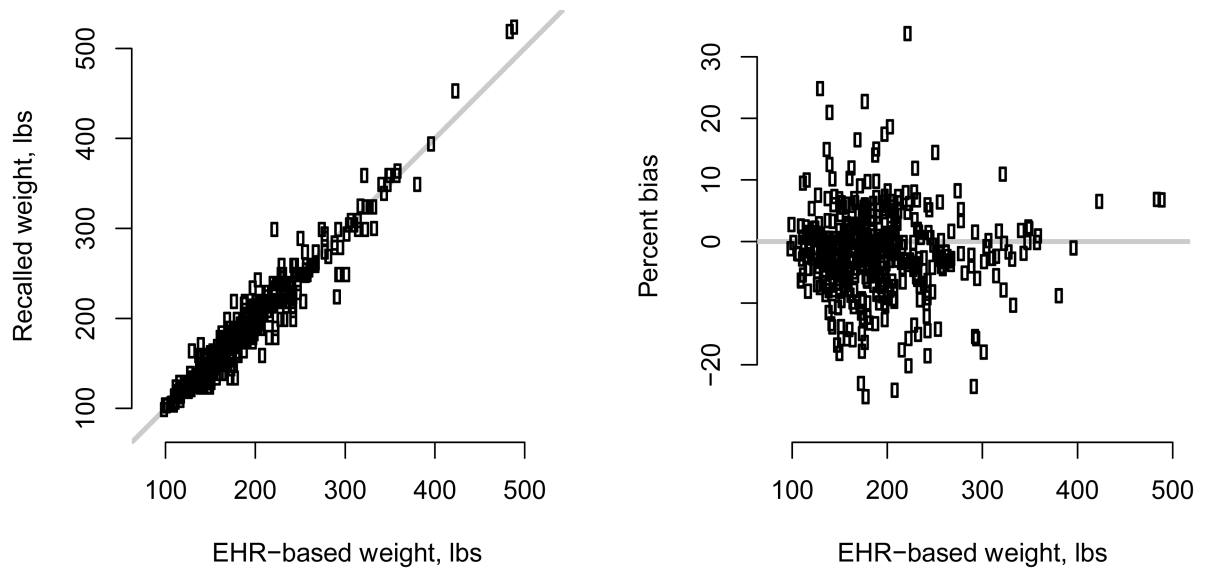


Figure 1. Comparison of self-reported weight at treatment initiation (baseline) and EHR-based weight. Shown are results based on 478 of 1,153 survey respondents (41%; Table 1) who had complete baseline weight data in the EHR. In the left-hand panel, the grey line is the 45° line.

Table 1

Distributions (counts and percentages) of characteristics observed in the Group Health EHR for N=8,345 patients satisfying inclusion/exclusion criteria, for n*=2,109 patients invited to participate in the survey and for n=1,153 patients who complete the survey. Also shown are the estimated distributions (percentages) for the N=8,345 patients in the full sample, based on re-weighting the observed distributions for the n=1,153 survey respondents to account for survey invitation and response.

	Observed information in the EHR				Estimated distribution for all N=8,345 patients			
	Full sample	Invited to participate	Survey respondents	%	Full sample	Invited to participate	Survey respondents	%
Total	8,345	2,109	1,153		8,345	2,109	1,153	
Female	5,823	1,398	790	66	5,823	1,398	790	69
Age, years								
<= 35	1,570	430	175	20	1,570	430	175	15
(35, 40]	760	190	98	9	760	190	98	9
(40, 45]	833	213	100	10	833	213	100	9
(45, 50]	1,065	275	142	13	1,065	275	142	12
(50, 55]	1,261	305	187	14	1,261	305	187	16
> 55	2,856	696	451	33	2,856	696	451	39
Diagnosis								
Major depressive episode - single	1,817	416	215	20	1,817	416	215	19
Major depressive episode - recurrent	2,534	641	343	30	2,534	641	343	30
Dysthymia	672	204	114	10	672	204	114	10
Depression NOS	3,322	848	481	40	3,322	848	481	41
Anxiety disorder	2,880	714	353	34	2,880	714	353	31
Sleep disorder	698	131	81	6	698	131	81	7
Smoker in the prior 9 months	1,491	322	134	15	1,491	322	134	12
Treatment type								
Psychotherapy	1,485	566	317	27	1,485	566	317	28
Mirtazapine/Paroxetine	363	202	116	10	363	202	116	10
Bupropion/Fluoxetine	2,837	688	367	33	2,837	688	367	32
Other	3,660	653	353	31	3,660	653	353	30

	Observed information in the EHR						Estimated distribution for all N=8,345 patients		
	Full sample		Invited to participate		Survey respondents		n	%	%
	N	%	n*	%	n	%			
Weight information in the EHR									
Complete BL and FU	5,630	67	337	16	197	17	70		
BL missing w/visit	623	8	518	25	273	24	7		
BL missing w/o visit	1,271	15	741	35	402	35	14		
BL complete & FU missing w visit	77	1	76	4	41	4	1		
BL complete & FU missing w/o visit	744	9	437	21	240	21	8		

Table 2

Distributions (counts and percentages) of characteristics that are unavailable in the Group Health EHR but observed in the survey, based on n=1,153 survey respondents. Also shown are the estimated distributions (percentages) for all N=8,345 patients in the full data based on re-weighting the observed distributions for the n=1,153 survey respondents to account for survey invitation and response.

	Observed information for n=1,153 survey respondents		Estimated distribution for all N=8,345 patients	
	n	%	n	%
Total	1,153			
Race				
White	958	83		82
Black/AA	23	2		2
Hispanic	45	4		4
Asian	14	1		2
Multiple races endorsed	90	8		7
Other	23	2		2
Married ^a	718	62		61
Education ^d				
Less than high school	565	49		53
At least some college	304	26		27
College graduate	283	25		20
Household income				
Less than \$50K	401	35		37
\$50-100K	485	42		44
More than \$100K	221	19		17
Don't know/refused	46	4		2
Employed full or part time	830	72		72
Weight control activities in the last 2 years				
Commercial weight loss program ^d	176	15		12
Meal replacement products	342	30		31
Prescription medication	282	24		27
Missing	32	3		2

	Observed information for n=1,153 survey respondents		Estimated distribution for all N=8,345 patients	
	n	%	n	%
Ever had weight loss surgery	41	4	4	4
Ever chosen not to be weighed during clinic visit ^a	193	17	17	15
Recalled pre-treatment weight trajectory				
Gaining weight	232	20	20	17
Losing weight	117	10	10	10
Stayed about the same	771	67	67	69
Missing	33	3	3	4
General attitude about weight in the last 2 years				
Trying to maintain	221	19	19	20
Trying to lose	635	55	55	51
Trying to gain	26	2	2	2
Not thinking about it	263	23	23	26
Missing	8	1	1	1
Recalled BMI at treatment initiation ^b				
< 25.0	383	33	33	40
25.0 - 29.9	387	34	34	31
30.0 - 34.9	181	16	16	14
35.0 - 39.9	85	7	7	7
>= 40.0	110	10	10	8
Missing	7	1	1	0
Perceived effect of treatment on weight				
Caused gain	205	18	18	22
Caused loss	97	8	8	7
Caused no change	752	65	65	62
Missing	99	9	9	10

^aUnless otherwise indicated, 2 or fewer observations were missing

^bBased on recalled weight and observed height information

Table 4

Estimated odds ratios (OR) and 95% confidence intervals (CI) based on four logistic regression analyses for whether or not a patient have a primary care visit or a weight measurement, at treatment initiation and during follow-up. Inverse-probability weighting was used to account for survey non-participation and non-response.

	Pre-episode primary care visit			Pre-episode weight, given a pre-episode primary care visit			Follow-up primary care visit			Follow-up weight, given a follow-up primary care visit		
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Recalled BMI at treatment initiation ^d												
< 25.0	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF
25.0 - 29.9	1.89	(1.14, 3.12)	0.76	(0.42, 1.38)	1.04	(0.59, 1.85)	0.73	(0.36, 1.49)	0.73	(0.36, 1.49)	0.73	(0.36, 1.49)
30.0 - 34.9	1.29	(0.67, 2.51)	1.32	(0.56, 3.09)	0.57	(0.28, 1.19)	0.57	(0.22, 1.45)	0.57	(0.22, 1.45)	0.57	(0.22, 1.45)
35.0 - 39.9	0.79	(0.33, 1.89)	0.75	(0.26, 2.12)	0.56	(0.18, 1.74)	0.75	(0.22, 2.63)	0.75	(0.22, 2.63)	0.75	(0.22, 2.63)
>= 40.0	2.36	(1.10, 5.04)	1.26	(0.51, 3.13)	0.70	(0.27, 1.79)	0.57	(0.18, 1.77)	0.57	(0.18, 1.77)	0.57	(0.18, 1.77)
Age, years												
<= 35	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF
(35, 40]	0.87	(0.38, 1.98)	0.64	(0.25, 1.66)	1.07	(0.49, 2.35)	0.41	(0.14, 1.19)	0.41	(0.14, 1.19)	0.41	(0.14, 1.19)
(40, 45]	0.52	(0.21, 1.30)	0.85	(0.31, 2.30)	1.07	(0.33, 3.50)	1.00	(0.33, 3.08)	1.00	(0.33, 3.08)	1.00	(0.33, 3.08)
(45, 50]	0.61	(0.28, 1.32)	0.63	(0.25, 1.55)	1.12	(0.45, 2.82)	0.58	(0.20, 1.69)	0.58	(0.20, 1.69)	0.58	(0.20, 1.69)
(50, 55]	0.54	(0.24, 1.21)	1.72	(0.68, 4.38)	0.60	(0.26, 1.36)	2.73	(0.89, 8.35)	2.73	(0.89, 8.35)	2.73	(0.89, 8.35)
> 55	0.68	(0.38, 1.22)	1.30	(0.61, 2.76)	0.71	(0.36, 1.39)	0.98	(0.40, 2.44)	0.98	(0.40, 2.44)	0.98	(0.40, 2.44)
Female	0.99	(0.61, 1.61)	2.81	(1.59, 4.97)	1.65	(0.99, 2.75)	4.03	(2.06, 7.91)	4.03	(2.06, 7.91)	4.03	(2.06, 7.91)
Non-white	1.12	(0.64, 1.96)	1.14	(0.56, 2.32)	0.44	(0.24, 0.81)	0.90	(0.42, 1.97)	0.90	(0.42, 1.97)	0.90	(0.42, 1.97)
College graduate	1.35	(0.82, 2.22)	0.69	(0.38, 1.25)	0.75	(0.44, 1.27)	0.57	(0.29, 1.12)	0.57	(0.29, 1.12)	0.57	(0.29, 1.12)
Household income												
Less than \$50K	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF
\$50-100K	0.81	(0.50, 1.33)	1.19	(0.67, 2.13)	0.88	(0.51, 1.52)	1.87	(0.96, 3.62)	1.87	(0.96, 3.62)	1.87	(0.96, 3.62)
More than \$100K	0.72	(0.36, 1.45)	1.51	(0.67, 3.38)	1.00	(0.47, 2.12)	1.24	(0.46, 3.31)	1.24	(0.46, 3.31)	1.24	(0.46, 3.31)
Employed full or part time	1.21	(0.68, 2.16)	1.43	(0.81, 2.54)	0.96	(0.49, 1.88)	1.70	(0.88, 3.29)	1.70	(0.88, 3.29)	1.70	(0.88, 3.29)
Married	1.19	(0.75, 1.90)	0.89	(0.52, 1.52)	0.93	(0.57, 1.54)	1.13	(0.62, 2.05)	1.13	(0.62, 2.05)	1.13	(0.62, 2.05)
Smoker in the prior 9 months	2.97	(1.37, 6.45)	1.11	(0.55, 2.24)	0.68	(0.35, 1.35)	0.81	(0.35, 1.89)	0.81	(0.35, 1.89)	0.81	(0.35, 1.89)
Diagnosis												
Major depressive episode - single	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF	1.00	REF

	Pre-episode primary care visit			Pre-episode weight, given a pre-episode primary care visit			Follow-up primary care visit			Follow-up weight, given a follow-up primary care visit		
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Major depressive episode - recurrent	0.72	(0.39, 1.32)	1.03	(0.51, 2.12)	0.57	(0.29, 1.12)	0.82	(0.35, 1.92)				
Dysthymia	1.27	(0.60, 2.68)	0.77	(0.34, 1.77)	1.27	(0.56, 2.87)	1.12	(0.40, 3.10)				
Non-specific depression	0.64	(0.35, 1.17)	0.61	(0.32, 1.15)	1.43	(0.76, 2.67)	0.71	(0.34, 1.49)				
Anxiety disorder	1.06	(0.68, 1.66)	1.62	(0.88, 2.98)	0.56	(0.33, 0.94)	0.62	(0.31, 1.23)				
Sleep disorder	1.30	(0.64, 2.65)	1.74	(0.69, 4.42)	2.69	(1.20, 6.03)	1.66	(0.62, 4.43)				
Recalled pre-treatment weight trajectory												
Gaining weight	1.00	REF	1.00	REF	1.00	REF	1.00	REF				
Losing weight	1.23	(0.59, 2.57)	1.76	(0.71, 4.41)	1.36	(0.61, 3.06)	1.52	(0.53, 4.36)				
Stayed about the same	1.19	(0.71, 2.01)	1.49	(0.83, 2.68)	0.88	(0.49, 1.56)	1.32	(0.65, 2.68)				
Ever chosen not to be weighed during clinic visit	2.75	(1.44, 5.26)	0.32	(0.17, 0.61)	1.31	(0.63, 2.73)	0.29	(0.13, 0.63)				
Weight control activities in the last 2 years												
Commercial weight loss program	1.07	(0.57, 1.99)	0.59	(0.29, 1.21)	0.35	(0.16, 0.73)	0.44	(0.19, 1.02)				
Meal replacement products	1.16	(0.73, 1.84)	0.91	(0.51, 1.64)	2.78	(1.57, 4.91)	1.76	(0.85, 3.65)				
Prescription medication	1.81	(1.14, 2.89)	1.08	(0.61, 1.90)	1.19	(0.69, 2.04)	0.87	(0.47, 1.64)				
General attitude about weight in the last 2 years												
Trying to maintain	1.00	REF	1.00	REF	1.00	REF	1.00	REF				
Trying to lose	1.08	(0.61, 1.91)	1.29	(0.65, 2.56)	1.31	(0.70, 2.44)	1.30	(0.57, 2.98)				
Trying to gain	0.56	(0.15, 2.08)	0.19	(0.04, 0.85)	1.35	(0.29, 6.32)	0.19	(0.05, 0.77)				
Not thinking about it	1.33	(0.70, 2.52)	1.06	(0.50, 2.23)	2.29	(1.15, 4.55)	1.31	(0.53, 3.27)				
Treatment choice												
Psychotherapy	1.00	REF	1.00	REF	1.00	REF	1.00	REF				
Mirtazapine/Paroxetine	2.54	(1.36, 4.74)	1.08	(0.47, 2.45)	3.72	(1.73, 7.97)	3.89	(1.07, 14.21)				
Bupropion/Fluoxetine	2.60	(1.53, 4.41)	0.77	(0.40, 1.50)	3.26	(1.73, 6.12)	0.95	(0.42, 2.19)				
Other antidepressant	3.03	(1.84, 5.00)	0.82	(0.44, 1.55)	2.36	(1.33, 4.18)	0.95	(0.46, 1.97)				
Perceived effect of treatment on weight												
Caused gain			1.00	REF	1.00	REF	1.00	REF				
Caused loss			1.38	(0.60, 3.18)	1.07	(0.60, 3.18)	1.07	(0.32, 3.58)				
Caused no change			1.03	(0.54, 1.94)	1.26	(0.54, 1.94)	1.26	(0.59, 2.66)				

^aBased on recalled weight and observed height information