Comparison of Weight Loss Data Collected by Research Technicians Versus Electronic Medical Records: The PROPEL Trial

Peter T. Katzmarzyk, PhD[1], **Emily F. Mire, MS**[1], **Corby K. Martin, PhD**[1], **Robert L. Newton Jr., PhD**[1], **John W. Apolzan, PhD**[1], **Eboni G. Price-Haywood, MD, MPH**[2], **Kara D. Denstel, MPH**[1], **Ronald Horswell, PhD**[1], **San T. Chu, MS**[1], **William D. Johnson, PhD**[1],

PROPEL Research Group

[1]Pennington Biomedical Research Center, Baton Rouge, LA

[2]Ochsner Xavier Institute for Health Equity and Research, New Orleans, LA

Abstract

Background/Objectives: Pragmatic trials are increasingly used to study the implementation of weight loss interventions in real-world settings. This study compared researcher-measured body weights versus electronic medical record (EMR)-derived body weights from a pragmatic trial conducted in an underserved patient population.

Subjects/Methods: The PROPEL trial randomly allocated 18 clinics to usual care (UC) or to an intensive lifestyle intervention (ILI) designed to promote weight loss. Weight was measured by trained technicians at baseline and at 6, 12, 18, and 24 months. A total of 11 clinics (6 UC/5 ILI) with 577 enrolled patients also provided EMR data (n = 561), which included available body weights over the period of the trial.

Results: The total number of assessments were 2 638 and 2 048 for the researcher-measured and EMR-derived body weight values, respectively. The correlation between researcher-measured and EMR-derived body weights was 0.988 (n = 1 939; p<0.0001). The mean difference between the EMR and researcher weights (EMR-researcher) was 0.63 (2.65 SD) kg, and a Bland-Altman graph showed good agreement between the two data collection methods; the upper and lower boundaries of the 95% limits of agreement are −4.65 kg and +5.91 kg, and 71 (3.7%) of the values were outside the limits of agreement. However, at 6 months, percent weight loss in the ILI compared to the UC group was 7.3% using researcher-measured data versus 5.5% using EMR-derived data. At 24 months, the weight loss maintenance was 4.6% using the technician-measured data versus 3.5% using EMR-derived data.

Conclusion: At the group level, body weight data derived from researcher assessments and an EMR showed good agreement; however, the weight loss difference between ILI and UC was blunted when using EMR data. This suggests that weight loss studies that rely on EMR data may require larger sample sizes to detect significant effects.

Clinical trial registration: ClinicalTrials.gov number NCT02561221

Introduction

Pragmatic trials are designed to study the implementation of health interventions with previously demonstrated efficacy and effectiveness, in "real-world" settings (1, 2). As such, pragmatic trials are often conducted in schools, communities, medical clinics, or other settings. For example, weight loss interventions conducted in primary care settings offer promise to the millions of Americans living with obesity (3–5). In support of this contention, two recent pragmatic trials demonstrated that weight loss approaches conducted in primary care settings are feasible and effective (6, 7). While these trials were conducted in the context of primary care, data collection was conducted using a "gold standard" approach by having trained research technicians measure the outcomes at fixed intervals (8). Recently, there has been increasing interest in collecting outcome data in pragmatic trials conducted in healthcare settings from existing electronic medical record (EMR) platforms (9, 10).

Few studies have compared EMR-derived body weight data to data collected for research purposes in the context of a weight loss trial (11–13). However, two studies demonstrated strong linear associations (r~0.99) between researcher-measured and EMR-derived body weights at different time points (11, 13), and weight-loss intervention effects have been shown to be similar when using researcher-measured and EMR-derived body weights (12, 13). These previous studies provide evidence of the utility of EMR-derived body weight, though these trials were conducted within single health systems that were typically large integrated health organizations with a single EMR system. These trials were also conducted in predominantly White, middle class populations. There is a need to better understand the utility of EMR-derived estimates of body weight for use in pragmatic trials from a variety of settings, including multi-site trials involving different types of health systems such as federally qualified health centers (FQHCs), and trials conducted in low income populations.

The purpose of this study was to compare researcher-measured body weights versus EMR-derived body weights from a pragmatic weight loss trial conducted in primary care clinics,

including several FQHCs and clinics associated with a large integrated health system, in a predominantly underserved, low-income patient population.

Materials/Subjects and Methods

The PROPEL study was a weight loss cluster-randomized trial conducted in 18 primary care clinics in Louisiana (ClinicalTrials.gov number, NCT02561221)(7, 14). The clinics were from five health systems; one large, nonprofit academic multispecialty healthcare delivery system, and four systems of FQHCs. The clinics were randomized by the study statistician to either an intensive lifestyle intervention (ILI) or a usual care (UC) group, after stratification by health system. All enrolled patients received the intervention to which their clinic was assigned. All patients provided written informed consent prior to participating in the study, which was conducted between April 2016 and September 2019. The study protocol was approved by the Pennington Biomedical Research Center Institutional Review Board.

Study Arms

A detailed description of the two study arms has been published previously (14). Briefly, the ILI was delivered by health coaches embedded in the primary care clinics and focused on achieving a personal weight loss goal of 10% for each patient. Patients in the ILI attended weekly sessions (16 in-person and 6 telephone) in the first 6 months, followed by sessions (alternating in-person and telephone) held at least monthly for the remaining 18 months. Furthermore, patients in the ILI group were provided with an electronic scale (BodyTrace) and were encouraged to weigh themselves daily. Patients worked with coaches to develop action plans for eating and physical activity to reach their weight loss goal. Patients in the UC group received routine care from their primary care team throughout the trial, and also received six newsletters covering topics related to sitting and health, goal setting, staying safe in the heat, memory health, self-care, sleep hygiene, and smoking cessation.

Outcomes

The outcomes described here include percent (%) change in body weight from baseline, and absolute changes in body weight. Detailed descriptions of the outcomes have been presented previously (14). The assessment of outcomes occurred in two ways: 1) using the full PROPEL sample, all assessments were conducted by trained research technicians who did not deliver the intervention (gold standard), and 2) for a sub-sample of PROPEL clinics who had EMR data available, outcomes were extracted from the EMR databases of the participating health care systems. Data for patients who developed a major medical condition, underwent bariatric surgery, or died during the course of the trial were included up to the time the event occurred (7).

Researcher-Measured Outcomes—Outcomes were assessed at in-clinic visits at baseline and at 6, 12, 18 and 24 months of follow-up. We attempted to measure each patient as close to their follow-up visit time as possible; we established +/− 90-day window around each anticipated study visit. Of the 2 061 follow-up visits, 96% (1 968) occurred within 30 days, while 98% (2 010) of the visits occurred within 45 days of the expected visit date.

The mean time (absolute values) between expected and actual measurement visits was 10.4 (SD 12.5) days (Supplementary Table S1). Body weight was measured by one of 14 trained technicians in duplicate with the patient in light clothing without shoes to the nearest 0.1 kg using a digital scale (Seca Model 876). The patient was instructed to stand in the middle of the scale with head erect and eyes looking straight ahead. The accuracy of the scales was checked at scheduled monitoring visits using standard weights. If the two measurements differed by >0.5 kg, a third measurement was obtained, and the two closest measurements are averaged for analysis. Given that PROPEL was a cluster-randomized trial, it was not possible to blind the technicians to group assignment. However, study investigators were blinded to group assignment.

Electronic Medical Record Data—Clinical data for eligible patients were extracted from the EMR databases of participating health care provider systems. All extracted data were transformed and harmonized to a common format. The data extraction process inherently associated a particular date with each extraction data element; for example, a date was associated with each value for weight.

The data from each health system was organized by their respective encounter identifiers and/or dates. Any records with exactly duplicated dates and measurement values for an individual were deleted, and in cases where a participant had two measurements on a given day, the measurement associated with the first listed encounter was chosen.

In order to identify visits in the EMR data that aligned with the PROPEL assessment time points (baseline and months 6, 12, 18 and 24), the EMR dataset was merged with the PROPEL analysis set. Any EMR records that were more than 90 days before the PROPEL assessment baseline or more than 90 days after the expected or completed PROPEL Month 24 visit were deleted (an "expected" visit date was exactly 6, 12, 18 or 24 months after the baseline assessment visit). The final PROPEL EMR visits were identified by retaining the respective EMR visits that had a weight measured and were closest to the expected PROPEL assessment date (+/− 90 days). Of the 2 048 EMR visits (baseline and/or follow-up), 65% (1 333) occurred within 30 days, while 79% (1 624) of the visits occurred with 45 days of the expected study visit date (Supplementary Table S1). The mean time (absolute values) between expected study visit and actual EMR visits was 26.6 (SD 23.0) days.

Statistical Analysis

We used three approaches to examine the agreement between researcher-measured body weight and EMR-derived body weight. First, the linear association between the two sets of measurements was determined using the Concordance Correlation Coefficient (CCC), a statistic similar to the Pearson correlation coefficient, but incorporating the departure from a 45 degree line through the origin indicating perfect agreement (15). Second, we used the approach of Bland and Altman to assess the 95% limits of agreement between the two methods, and graphically plot the association between the difference between the two measurements against the mean of the two measurements for each patient (16). We used linear regression to determine if the differences between the two methods differed over levels of body weight. Third, the outcomes at 6, 12, 18, and 24 months were analyzed in

the context of repeated-measures linear mixed-effects multilevel models, which included random cluster (clinic) effects. Intervention group, assessment time, and their interaction terms were included in the models with and without additional covariates (age, sex and race). We performed intention-to-treat analyses, which included all patients (regardless of the number of assessments obtained) and used the restricted maximum likelihood method. The model assumed that missing values were missing at random, and all values presented in the tables and figure are model-based estimates. Power analyses conducted a priori indicated that the sample size provided at least 97% power to detect a mean weight loss in the ILI group of 3.5% at 24 months, relative to UC (7). We conducted a sensitivity analysis, by re-analyzing the general linear models after restricting the EMR sample to those that visits that occurred within 30 days of the expected study visit date. All analyses were conducted using data from patients from clinics that had EMR outcome data available, and the intent-to-treat analysis was conducted using patients with both researcher-measured and EMR-derived data available at baseline. Analyses were conducted with the use of SAS software, version 9.4 (SAS Institute).

Results

The PROPEL trial was conducted in 803 patients from 18 primary care clinics (9 ILI, 9 UC) from five health systems in Louisiana. EMR data were available from three of the health systems (seven clinics from two FQHC systems, and four clinics from one large integrated health system), including 11 clinics (5 ILI, 6 UC). In these 11 clinics, 577 patients had researcher-measured body weight at baseline (273 ILI; 304 UC). Of these patients, 561 patients (265 ILI, 296 UC) had EMR data available for at least 1 visit, while 530 patients had EMR data available at baseline. The total number of assessments in the trial were 2 638 and 2 048 for the researcher-measured and EMR-derived body weight values, respectively, while 1 939 patients had both assessments at the same time points.

Table 1 presents descriptive characteristics of the samples. There were no differences in body weight (p=0.32) or BMI (p=0.43) at baseline between patients who had EMR data available at baseline (n=530) versus those who did not (n=273). However, patients who had EMR data available at baseline were significantly older than those who did not (51.2 years versus 45.8 years; p<0.001). There was also a lower proportion of women (p=0.002) and a higher proportion of Black patients (p=0.009) in the sample with EMR data available.

The CCC between researcher-measured and EMR-derived body weights was 0.988 (n = 1 939; p<0.0001) (Figure 1). The mean difference between the EMR-derived and researcher-measured weights (EMR-technician) was 0.63 (2.65 SD) kg (p<0.001), indicating that, on average, the EMR-derived weights were higher than researcher-measured weights. When restricting the sample to EMR visits that occurred within 30 days of expected, the mean difference between the EMR-derived and researcher-measured weights (EMR-technician) was 0.64 (2.32 SD) kg (p<0.001). Further, the correlation between the time between visits and the mean difference was −0.02 (p=0.39). There were also no differences between FQHC and non-FQHC clinics (p=0.27); the mean difference between the EMR-derived and researcher-measured weights (EMR-technician) was 0.56 (3.18 SD) kg in FQHC clinics versus 0.69 (2.19 SD) kg in non-FQHC clinics.

The Bland and Altman graph showing the agreement between the two methods is shown in Figure 2. The upper and lower boundaries of the 95% limits of agreement are −4.65 kg and +5.91 kg. A total of 71 (3.7%) of the values were outside the 95% limits of agreement. Furthermore, there were 40 (2.1%) cases where the EMR value was 5 kg heavier than the researcher measured weight and there were 38 (2%) cases where the researcher measure was 5 kg heavier than the weight found in the EMR. The regression analysis between the weight difference and the mean weight for the two methods indicated minimal variance explained ($R^2 = 0.003$) but a significant positive slope (0.00811; p=0.02), suggesting that the difference between the methods tended to increase across the range of body weight. Bland and Altman graphs are also presented in Supplementary Figure 1 and Supplementary Figure 2 by study arm and time point. In general, there is good agreement between the two data collections methods at the group level, though values for a small number of participants varied markedly.

Figure 3 presents the patient flow through the trial, based on patients who had data available at baseline. Figure 4 presents the mean percent weight loss in the UC and ILI groups at each time point using the EMR-derived data and researcher-measured data. The EMR and researcher-measured data are very similar across time points for the UC group. While the general trend in weight loss over time in the ILI is captured by both sets of data, the magnitude of the weight loss is less using the EMR dataset. For example, at 6 months, weight loss in the ILI compared to the UC group was 7.3% using researcher-measured data versus 5.5% using EMR-derived data (Table 2). Further, at 24 months, the weight loss maintenance was 4.6% using the researcher-measured data versus 3.5% using EMR-derived data (Table 2). The results of the sensitivity analyses where we restricted the sample to include only those visits that occurred within 30 days of the expected study visit date are presented in Supplementary Table S2. The results how modest improvement at some time points over the primary analyses which included all study visits that within 90 days of the expected study visit date.

Discussion

The results demonstrate good concordance between researcher-measured and EMR-derived body weights in the PROPEL trial. The correlation between measurement was ~0.99, and only 3.7% of the paired body weight measurements fell outside the 95% limits of agreement. However, the observed differences in response between the UC and ILI groups was attenuated when using the EMR-derived weights. This could be due to a variety of factors. For example, the difference in the time of measurement around each "expected" visit (i.e., 6, 12, 18 and 24 months) was a mean of 10.4 days for the researcher-measured weights versus a mean of 26.4 days for the EMR-derived weights. This is somewhat expected given that the timing of patient visits to their PCP were not specified in the PROPEL protocol, and were naturally occurring over the time period of trial. Further, the mean difference between EMR-derived and researcher-measured weights was 0.63 (SD 2.65) kg which could explain, in part, the difference observed (i.e., the difference of differences between the ILI and UC groups was approximately 1 kg at 24 months (Table 2), which is within the limits of agreement). The higher body weights observed in the EMR data could be explained, in part, by differences in measurement protocol between the research study and the primary

care clinics. For example, the research study followed a strict protocol where all patients removed all heavy outer clothing and shoes prior to measurement, whereas the protocols differed among the clinics, and in some cases, required patients to keep their shoes on while on the scale.

Inspection of Figure 4 suggests that the majority of the blunted treatment effect when using EMR data is primarily due to differences in measurements in the ILI group. We do not know definitively why this is, but one hypothesis is that the differences in timing between the researcher-measured and EMR-derived body weights might not have a big effect in the UC group, where weights are relatively stable over time. However, in the ILI group which is characterized by patients changing their weight over time, differences in timing of the measurements may play a bigger role; i.e., any lags between technician-measured and EMR-data collection may miss important changes in body weight during the active phase of weight loss. This highlights the need to carefully control for the timing of EMR data collection in future weight loss trials.

Results of the present study can be compared to a limited number of previous studies that examined similar issues. For example, Stevens and colleagues compared body weight measurements obtained during group meetings in a weight loss program in a large HMO and body weights abstracted from medical charts (not electronic)(17). Of 112 participants in the program, 64 (57%) had at least one body weight in their medical chart within 30 days of a weight recorded in the weight loss program. Based on data from a total of 123 paired body weight measurements, the Pearson correlation between measurements was 0.998, with a mean difference of 0.3 kg (SD 1.08 kg). DiMaria-Ghalili and colleagues compared body weights collected by a researcher versus those extracted from medical charts (not electronic) during a prospective study of older adults undergoing elective coronary artery bypass surgery (18). Data from 84 patients collected pre-operatively demonstrated a mean difference of 0.19 kg (SD 1.36 kg) between the medical chart and researcher-derived estimates of weight. Furthermore, a total of 2.4% of values fell outside of the 95% limits of agreement based on the method of Bland and Altman, similar to the current study (18).

Several studies have examined associations between body weight measurements obtained by researchers and those derived from EMRs (11–13, 19), Leo and colleagues compared EMR body weights collected during routine care versus body weights collected during a research project designed to study weight gain in pregnant women in a large HMO (86% White)(19). Among 102 women with both sets of measurements, the CCC between the two measurements was >0.99, with 95% limits of agreement of −0.15 and +1.7 kg. Arterburn and colleagues compared EMR data collected during routine medical care versus study-determined body weights among 291 women (81% White) with obesity from a prepaid health plan and delivery system (11). Among 224 women with data at 12-months of follow-up, the Pearson correlation between researcher-determined and EMR-derived body weights was 0.99. Further, the Pearson correlation between 12-month changes in study-determined and EMR-derived body weights was 0.96 (11). Gallis and colleagues compared clinic EMR body weights to researcher-determined body weights using data from two studies: one weight loss trial, and one weight gain prevention trial (12). The participants included 307 adults from the Track study (28% White)(20) and 139 adults from the Shape study (100%

Black)(21) who consented to EMR data extraction. For both trials, the study-determined body weights were lower and less variable than the EMR-derived weights; however, the intervention effects relative to usual care were similar when using study-determined versus EMR-extracted body weights (12). Similarly, Xiao and colleagues examined the agreement between EMR-derived and researcher-measured body weights in the BE WELL (n=330; 50% White) and E-LITE (N=241; 78% White) trials, conducted in two large integrated health systems in California (13). The CCCs between measurements at different time points in the BE WELL trial were >0.99, and Bland and Altman graphs demonstrated no systematic differences between EMR-derived and researcher-measured weights (13). The results of the present study extend those of previous work by exploring the associations between researcher-measured and EMR-derived body weight data in the context of a large cluster-randomized trial conducted in multiple health systems serving low-income patients, including several FQHCs.

The NIH Health Care Systems Research Collaboratory recently identified several challenges to using EMR data in research, including 1) inadequate collection of outcome data, 2) lack of structured data collection, 3) lack of data standardization, 4) inadequate resources to support EMR customization, 5) difficulties in aggregating data across sites, and 6) accessing EMR data (22). We encountered some of these challenges in the PROPEL trial, including lack of structured data collection, lack of data standardization, difficulties in aggregating data across sites, and accessing EMR data. Given that the primary data collection was conducted by researchers, PROPEL did not invest in structuring or standardizing data across EMR system a priori, which would have made it easier to aggregate data across sites. Alternatively, we relied on the "natural" collection of outcome data during clinical encounters, and we invested efforts on the back end to standardize the data across several different EMR platforms. This was a time-consuming process, and future studies that plan to rely on EMR data should address these issues at the study design phase. Furthermore, while the original PROPEL trial included 18 clinics, our analysis was limited to the 11 clinics that were able to provide EMR data. In particular, we were limited by one FQHC system which was unable to provide data on their clinics given their reliance on third-party EMR vendors.

While we corrected several errors that we identified in the EMR data (i.e. weight recording pounds rather than kilograms, clear data entry transposition errors, etc.), it should be noted that we did not exclude outliers from the analyses in the present study, either from the EMR-derived data or researcher-measured data. We felt that this was the most transparent approach given the purpose of the study was to compare the two data collection methods. Six of the paired body weight assessments (0.3%), however, demonstrated more than a 20 kg difference between the researcher-measured versus EMR-derived data. Nonetheless, based on trends in data among these patients over time and information from other sources, we estimate that at least five of these cases involved incorrect EMR data rather than errors in the researcher-measured data. While these outliers had little impact on the current analysis, it highlights: 1) the benefit of relying on researcher-measured vs. EMR-derived body weights, when possible, and 2) the need to carefully inspect EMR data from outliers when solely relying on that data source.

This study has several strengths and limitations. A major strength is the inclusion of patients from several FQHCs in addition to primary care clinics affiliated with a large integrated health system, which increases the generalizability of the results to other underserved, low income populations living in the U.S. However, given this, it is possible that the results may not be directly applicable to future studies conducted in health care settings with high-functioning EMR systems in place to support clinical research. A limitation is that while EMR data were collected in parallel with the research project, the EMR data collection occurred in relation to naturally occurring patient- or PCP-driven clinical visits (i.e., EMR data collection was not timed to occur in relation to the measurement of the trial outcomes). Future studies that plan to rely on EMR data to assess body weight outcomes should coordinate clinical care visits to reflect the needs of the study as much as possible. Furthermore, primary care clinics should strive to accurately measure and record the weights of their patients using consistent methods, including the removal of heavy items of clothing.

In conclusion, body weight data derived from researcher assessments and an EMR showed good agreement in this pragmatic trial. However, a small number of EMR-derived weights appeared to be erroneous and grossly inaccurate, and the weight loss difference between the two conditions (i.e., ILI versus UC) was blunted when using EMR data, which may be a concern for studies that rely solely on EMR data to assess outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability

The datasets analyzed and computer code utilized in the current study are available from the corresponding author on reasonable request.

References

1. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290(12):1624–32. [PubMed: 14506122]

2. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011;13(2):217–24. [PubMed: 21842619]

3. Tsai AG, Remmert JE, Butryn ML, Wadden TA. Treatment of obesity in primary care. Med Clin North Am. 2018;102(1):35–47. [PubMed: 29156186]

4. Carvajal R, Wadden TA, Tsai AG, Peck K, Moran CH. Managing obesity in primary care practice: A narrative review Ann NY Acad Sci. 2013;1281:191–206. [PubMed: 23323827]

5. Tronieri JS, Wadden TA, Chao AM, Tsai AG. Primary care interventions for obesity: Review of the evidence. Curr Obes Rep. 2019;8(2):128–36. [PubMed: 30888632]

6. Befort CA, VanWormer JJ, Desouza C, Ellerbeck EF, Gajewski B, Kimminau KS, et al. Effect of behavioral therapy with in-clinic or telephone group visits vs in-clinic individual visits on weight loss among patients with obesity in rural clinical practice: A randomized clinical trial. JAMA. 2021;325(4):363–72. [PubMed: 33496775]

7. Katzmarzyk PT, Martin CK, Newton RL Jr., Apolzan JW, Arnold CL, Davis TC, et al. Weight loss in underserved patients - A cluster-randomized trial. N Engl J Med. 2020;383(10):909–18. [PubMed: 32877581]

8. Katzmarzyk PT, Apolzan JW, Gajewski B, Johnson WD, Martin CK, Newton RL Jr., et al. Weight loss in primary care: A pooled analysis of two pragmatic cluster-randomized trials. Obesity. 2021;29(12):2044–54. [PubMed: 34714976]

9. Ellenberg SS, Ellenberg JH. Proceedings of the University of Pennsylvania 12th annual conference on statistical issues in clinical trials: Electronic health records (EHR) in randomized clinical trials-Challenges and opportunities. Clin Trials. 2020;17(4):343–5. [PubMed: 32522023]

10. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - What is it and what can it tell us? N Engl J Med. 2016;375(23):2293–7. [PubMed: 27959688]

11. Arterburn D, Ichikawa L, Ludman EJ, Operskalski B, Linde JA, Anderson E, et al. Validity of clinical body weight measures as substitutes for missing data in a randomized trial. Obes Res Clin Pract. 2008;2(4):277–81. [PubMed: 19956347]

12. Gallis JA, Kusibab K, Egger JR, Olsen MK, Askew S, Steinberg DM, et al. Can electronic health records validly estimate the effects of health system interventions aimed at controlling body weight? Obesity (Silver Spring). 2020;28(11):2107–15. [PubMed: 32985131]

13. Xiao L, Lv N, Rosas LG, Au D, Ma J. Validation of clinic weights from electronic health records against standardized weight measurements in weight loss trials. Obesity (Silver Spring). 2017;25(2):363–9. [PubMed: 28059466]

14. Katzmarzyk PT, Martin CK, Newton RL Jr., Apolzan JW, Arnold CL, Davis TC, et al. Promoting Successful Weight Loss in Primary Care in Louisiana (PROPEL): Rationale, design and baseline characteristics. Contemp Clin Trials. 2018;67:1–10. [PubMed: 29408562]

15. Lawrence IKL. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45:255–68. [PubMed: 2720055]

16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10. [PubMed: 2868172]

17. Stevens VJ, Wagner EL, Rossner J, Craddick S, Greenlick MR. Validity and usefulness of medical chart weights in the long-term evaluation of weight loss programs. Addict Behav. 1988;13(2):171–5. [PubMed: 3369325]

18. DiMaria-Ghalili RA. Medical record versus researcher measures of height and weight. Biol Res Nurs. 2006;8(1):15–23. [PubMed: 16766625]

19. Leo MC, Lindberg NM, Vesco KK, Stevens VJ. Validity of medical chart weights and heights for obese pregnant women. EGEMS (Wash DC). 2014;2(1):1051. [PubMed: 25848585]

20. Foley P, Steinberg D, Levine E, Askew S, Batch BC, Puleo EM, et al. Track: A randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials. 2016;48:12–20. [PubMed: 26995281]

21. Bennett GG, Foley P, Levine E, Whiteley J, Askew S, Steinberg DM, et al. Behavioral treatment for weight gain prevention among black women in primary care practice: a randomized clinical trial. JAMA Intern Med. 2013;173(19):1770–7. [PubMed: 23979005]

22. Richesson RL, Marsolo KS, Douthit BJ, Staman K, Ho PM, Dailey D, et al. Enhancing the use of EHR systems for pragmatic embedded research: lessons from the NIH Health Care Systems Research Collaboratory. J Am Med Inform Assoc. 2021.
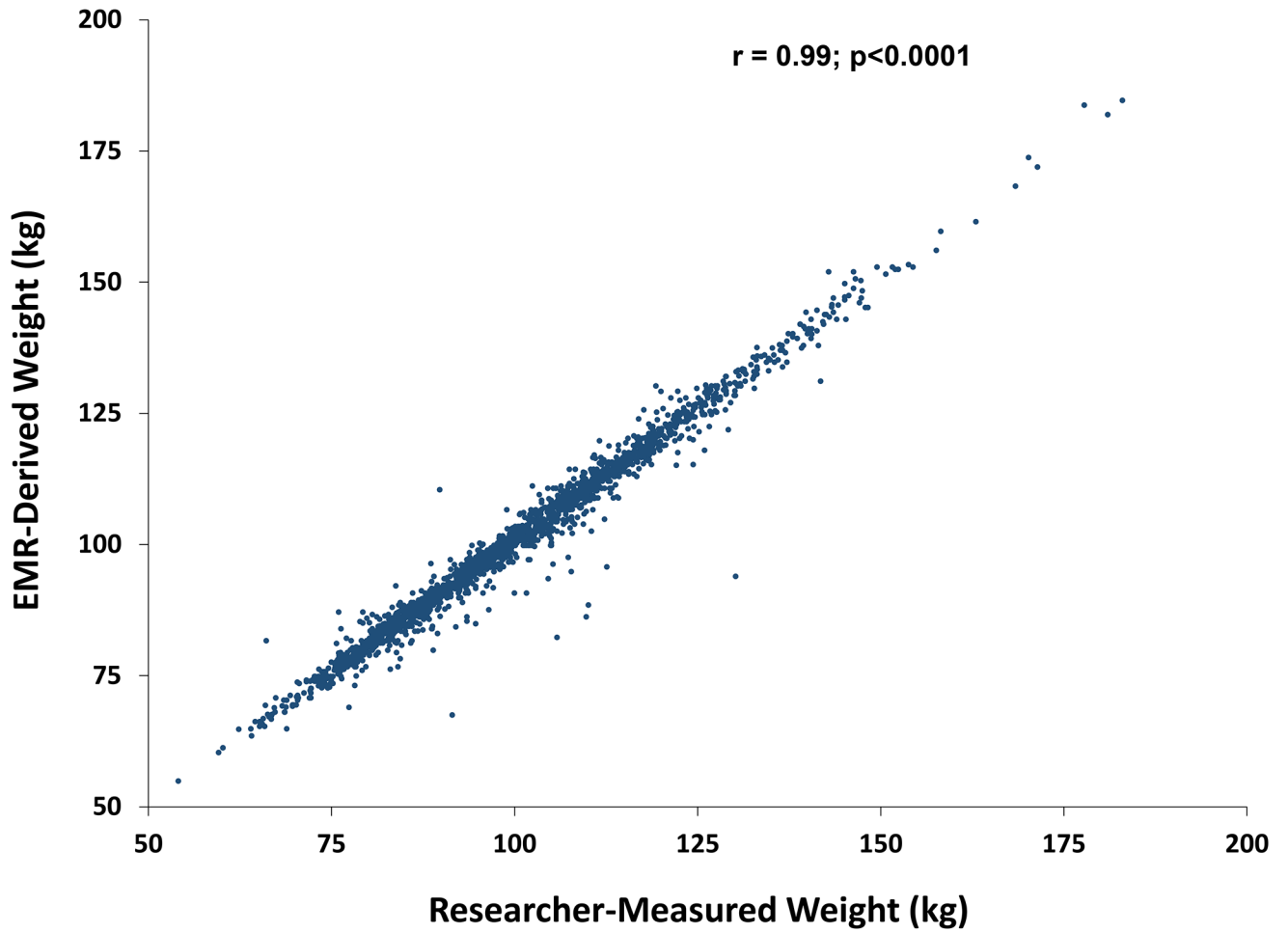
Figure 1.
Scatterplot of association between EMR-derived and researcher-measured body weights (n = 1 939).
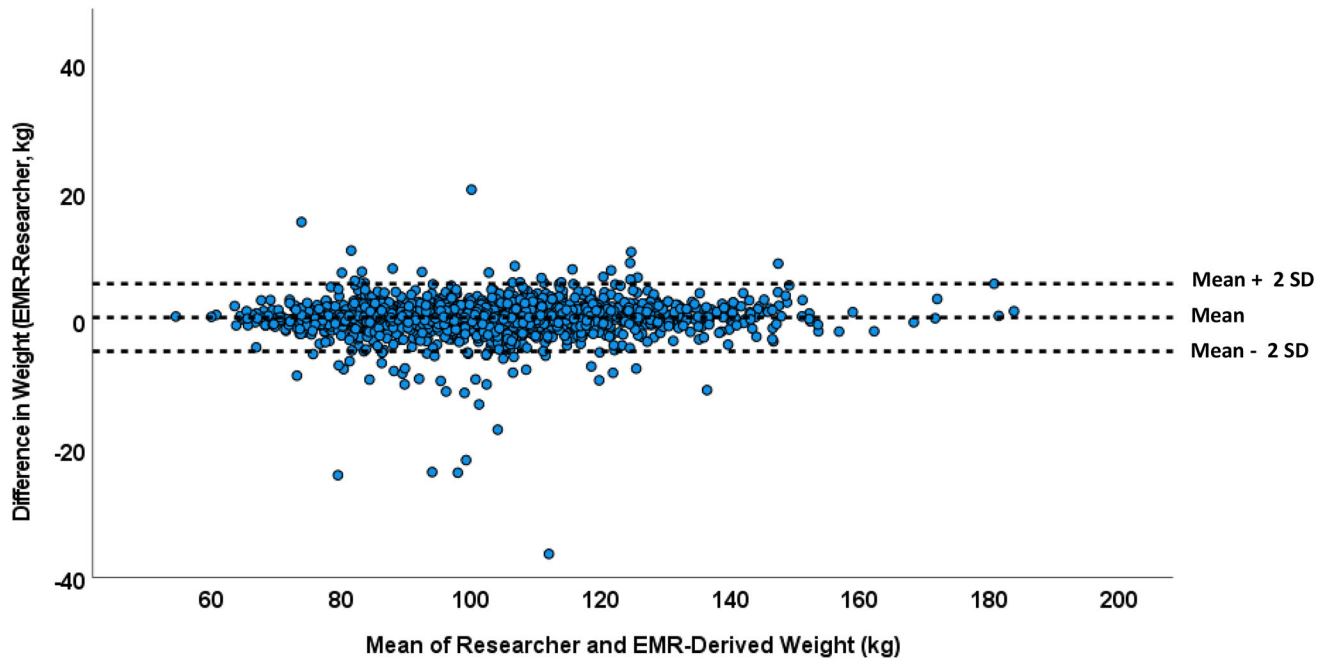
Figure 2.
Bland and Altman plot of mean differences and limits of agreement for EMR-derived and researcher-measured weights in the PROPEL trial (n=1 936).
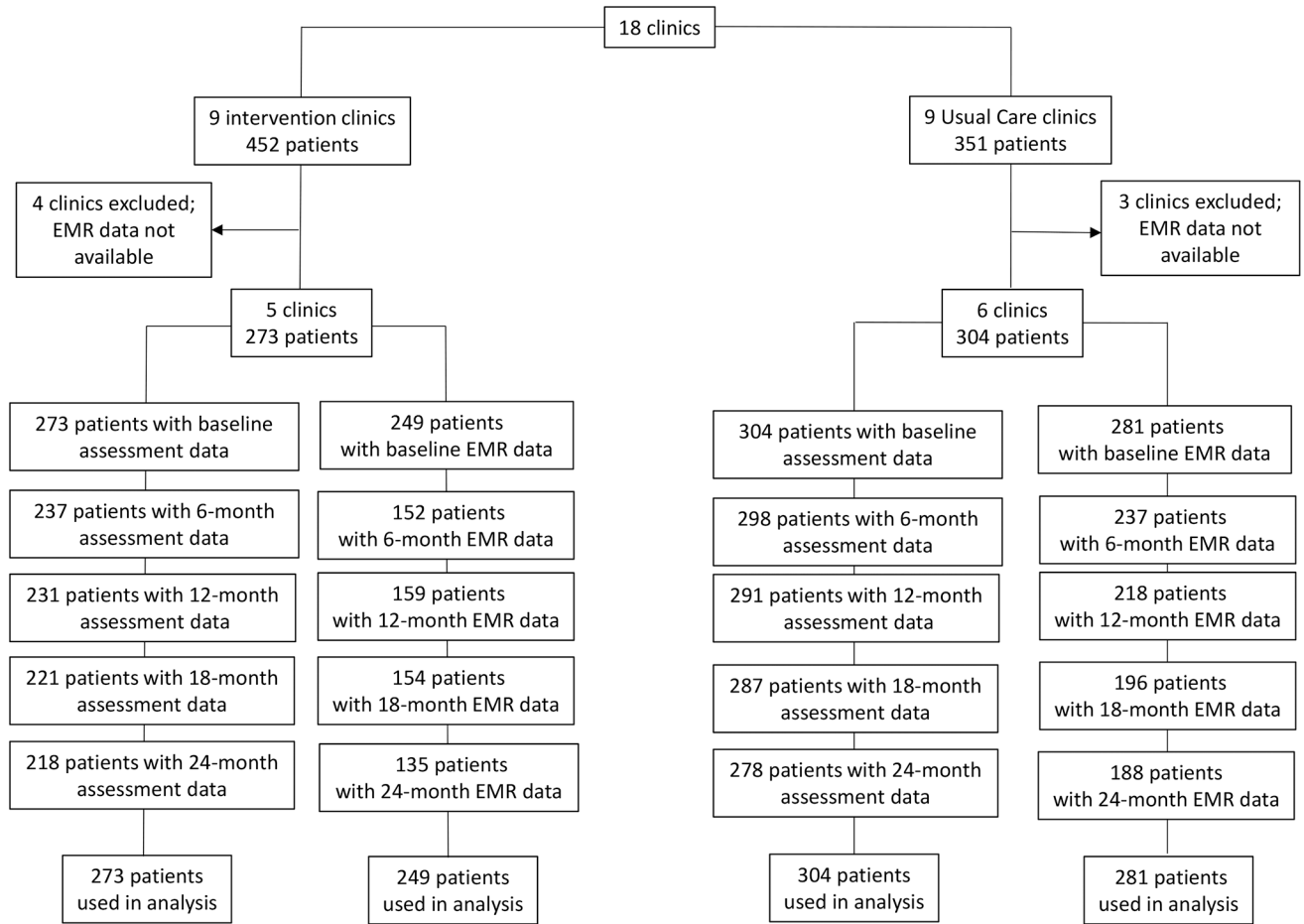
```
                                    ┌─────────────┐
                                    │  18 clinics │
                                    └─────────────┘

        ┌──────────────────────┐              ┌──────────────────────┐
        │ 9 intervention clinics│              │  9 Usual Care clinics │
        │    452 patients       │              │     351 patients      │
        └──────────────────────┘              └──────────────────────┘

 ┌──────────────────┐                                          ┌──────────────────┐
 │ 4 clinics excluded;│                                        │ 3 clinics excluded;│
 │   EMR data not    │                                         │   EMR data not    │
 │    available      │                                         │    available      │
 └──────────────────┘                                          └──────────────────┘

        ┌──────────────┐                              ┌──────────────┐
        │   5 clinics   │                              │   6 clinics   │
        │  273 patients │                              │  304 patients │
        └──────────────┘                              └──────────────┘
```

| 273 patients with baseline assessment data | 249 patients with baseline EMR data | 304 patients with baseline assessment data | 281 patients with baseline EMR data |

| 237 patients with 6-month assessment data | 152 patients with 6-month EMR data | 298 patients with 6-month assessment data | 237 patients with 6-month EMR data |

| 231 patients with 12-month assessment data | 159 patients with 12-month EMR data | 291 patients with 12-month assessment data | 218 patients with 12-month EMR data |

| 221 patients with 18-month assessment data | 154 patients with 18-month EMR data | 287 patients with 18-month assessment data | 196 patients with 18-month EMR data |

| 218 patients with 24-month assessment data | 135 patients with 24-month EMR data | 278 patients with 24-month assessment data | 188 patients with 24-month EMR data |

| 273 patients used in analysis | 249 patients used in analysis | 304 patients used in analysis | 281 patients used in analysis |

Figure 3.
Patient flow through the PROPEL trial among clinics with electronic medical record (EMR) data available.
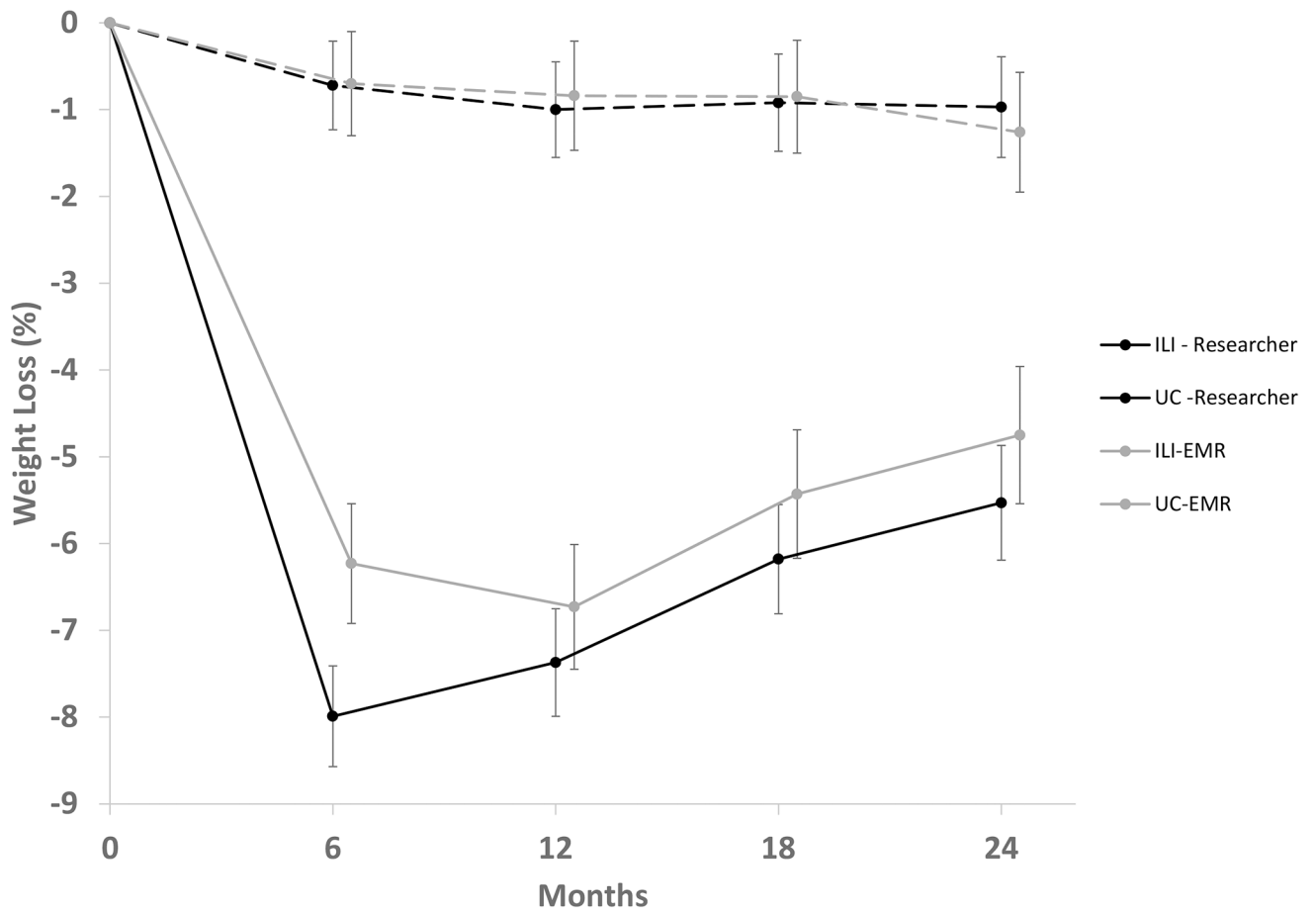
Figure 4.
Mean percent change in weight from baseline in the usual care (UC) group and the intensive lifestyle intervention (ILI) group) using data collected by trained research technicians versus data obtained from electronic medical records (EMR). The error bars represent standard errors.

Table 1.

Descriptive baseline measurements from the Promoting Successful Weight Loss in Primary Care in Louisiana (PROPEL) trial in the EMR-derived sub-sample and among patients who were excluded due to not having EMR data available.

	PROPEL EMR Sample*	PROPEL Non-EMR Sample
N	530	273
Age (y)	51.2 (12.8)	45.8 (13.0)
Weight (kg)	102.5 (17.1)	101.3 (15.9)
Body Mass Index (kg/m^2)	37.2 (4.6)	37.4 (4.8)
Sex (% Female)	81.5	90.1
Race (% Black)	70.4	61.2

*Results are restricted to the sub-sample that has EMR-derived data at baseline.

Data are presented as mean (SD) or percentage.

Table 2.

Differences between usual care and the intensive lifestyle intervention groups for changes in weight loss variables over two years using researcher-measured and EMR-derived data.[1]

Variable	Researcher-Measured	EMR–Derived
Change in Body Weight (%)		
At 6 months	−7.28 (−8.77, −5.79)	−5.53 (−7.30, −3.75)
At 12 months	−6.36 (−7.97, −4.76)	−5.89 (−7.75, −4.03)
At 18 months	−5.26 (−6.90, −3.62)	−4.57 (−6.49, −2.65)
At 24 months	−4.56 (−6.29, −2.83)	−3.49 (−5.56, −1.42)
Change in Body Weight (kg)		
At 6 months	−7.51 (−9.14, −5.88)	−5.80 (−7.72, −3.89)
At 12 months	−6.48 (−8.22, −4.73)	−6.12 (−8.12, −4.11)
At 18 months	−5.38 (−7.15, −3.60)	−4.86 (−6.92, −2.80)
At 24 months	−4.66 (−6.51, −2.80)	−3.70 (−5.90, −1.49)

Values are mean difference (95% C.I.);

[1] Adjusted for age, race and sex;