

5-15-2014

Validity of Medical Chart Weights and Heights for Obese Pregnant Women

Michael C. Leo PhD

Kaiser Permanente Center for Health Research, Michael.C.Leo@kpchr.org

Nangel M. Lindberg PhD

Kaiser Permanente Center for Health Research, Nangel.M.Lindberg@kpchr.org

Kimberly K. Vesco MD

Kaiser Permanente Center for Health Research, Kimberly.K.Vesco@kpchr.org

Victor J. Stevens PhD

Kaiser Permanente Center for Health Research, Victor.J.Stevens@kpchr.org

Follow this and additional works at: <http://repository.edm-forum.org/egems>



Part of the [Health Services Research Commons](#)

Recommended Citation

Leo, Michael C. PhD; Lindberg, Nangel M. PhD; Vesco, Kimberly K. MD; and Stevens, Victor J. PhD (2014) "Validity of Medical Chart Weights and Heights for Obese Pregnant Women," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 2: Iss. 1, Article 7.

DOI: <http://dx.doi.org/10.13063/2327-9214.1051>

Available at: <http://repository.edm-forum.org/egems/vol2/iss1/7>

This Methods Empirical Research is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Validity of Medical Chart Weights and Heights for Obese Pregnant Women

Abstract

Objective: To determine the validity of adult body weights and heights recorded in electronic medical records (EMRs) in the course of routine medical care.

Background: EMRs allow the potential use of data collected in the course of routine medical care for a variety of research applications in many fields including epidemiology and comparative effectiveness studies. However, researchers familiar with carefully controlled measurement protocols typically used in clinical trials may question the validity of data collected in the course of routine clinical care.

Methods: Weights and heights collected during a research project that focused on weight gain during pregnancy were compared to weight and height measurements coincidentally recorded in the research participant's medical records. For weight measures (N=102), data recorded within ± 14 days were compared, and for height measures (N=114), data recorded within ± 5 years were compared. We assessed agreement between medical and research measurements using the concordance and intraclass correlation coefficients, and Bland and Altman's limits of agreement.

Findings: The mean research and medical record weight measurements were 99.3 kg and 99.2 kg, respectively. The concordance and intraclass correlation coefficients for weight had similar estimates of .999 and 95 percent confidence intervals [.998, .999]. The 95 percent limits of agreement were -1.5 kg and +1.7 kg. The mean research and medical height measurements were 1.646 m and 1.654 m, respectively, and the concordance and intraclass correlation coefficients for height were .941 and .942, respectively. The 95 percent limits of agreement were -.031 m and +.047 m.

Conclusions: For pregnant women, body weights documented in the medical record are exchangeable with body weights recorded in a research setting. Height measurements recorded in the medical records were not in as close agreement as weights, but concordance between medical record and research height measurements are high enough to allow them to be used epidemiological and comparative effectiveness research.

Acknowledgements

This work was supported by a grant from the National Institute of Child Health and Human Development R01 HD0586-01.

Keywords

Weight, height, validity, pregnancy, medical record

Disciplines

Health Services Research

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

Validity of Medical Chart Weights and Heights for Obese Pregnant Women

Michael C. Leo, PhD; Nangel M. Lindberg, PhD; Kimberly K. Vesco, MD; Victor J. Stevens, PhD¹

Abstract

Objective: To determine the validity of adult body weights and heights recorded in electronic medical records (EMRs) in the course of routine medical care.

Background: EMRs allow the potential use of data collected in the course of routine medical care for a variety of research applications in many fields including epidemiology and comparative effectiveness studies. However, researchers familiar with carefully controlled measurement protocols typically used in clinical trials may question the validity of data collected in the course of routine clinical care.

Methods: Weights and heights collected during a research project that focused on weight gain during pregnancy were compared to weight and height measurements coincidentally recorded in the research participant's medical records. For weight measures (N=102), data recorded within ± 14 days were compared, and for height measures (N=114), data recorded within ± 5 years were compared. We assessed agreement between medical and research measurements using the concordance and intraclass correlation coefficients, and Bland and Altman's limits of agreement.

Findings: The mean research and medical record weight measurements were 99.3 kg and 99.2 kg, respectively. The concordance and intraclass correlation coefficients for weight had similar estimates of .999 and 95 percent confidence intervals [.998, .999]. The 95 percent limits of agreement were -1.5 kg and +1.7 kg. The mean research and medical height measurements were 1.646 m and 1.654 m, respectively, and the concordance and intraclass correlation coefficients for height were .941 and .942, respectively. The 95 percent limits of agreement were -.031 m and +.047 m.

Conclusions: For pregnant women, body weights documented in the medical record are exchangeable with body weights recorded in a research setting. Height measurements recorded in the medical records were not in as close agreement as weights, but concordance between medical record and research height measurements are high enough to allow them to be used epidemiological and comparative effectiveness research.

Introduction

Body weight and height measures are commonly recorded in patient medical records,¹ and thus could serve as a potential source of data for research purposes. Among possible uses of these data would be population-based, cross-sectional and longitudinal studies of obesity and osteoporosis, as well as large-scale comparative effectiveness studies for obesity prevention and treatment programs. Using data already recorded in electronic medical records (EMRs) allows for very low-cost, population-level research. However, there is a question of whether measurements collected in the course of routine clinical practice are sufficiently valid for research purposes.

Background and Context

A few previous studies have found close agreement between adult body weights measured in research studies and body weights coincidentally measured during routine clinical care. Stevens et al.² examined 123 pairs of weights, one recorded in a weight-loss study and one recorded in the patient's medical record as a part of routine clinical care. That study found that research and clinical care weights collected within 30 days of each other had a high level of concordance, with a mean difference of 0.03 kg, and a standard deviation of 1.08 kg. Similar results were found analyzing pairs of body weight data among 85 older patients undergoing elective surgery,³ and 224 obese depressed women.⁴

¹Kaiser Permanente Center for Health Research

Despite the evidence that weight data may be accurately collected in the course of routine clinical care, the validity of this data is still often questioned. The current study examines the accuracy of body weights and heights of obese pregnant women, comparing the data recorded in a research setting to data recorded in the research participants' EMRs in the course of routine medical care. Compared to previous research on this topic, this study uses a different population that experiences relatively rapid weight gain and loss, and examines the absolute agreement (i.e., exchangeability) between clinic and research measurements versus the consistency between these two sources of height and weight data.

Methods

The data for this analysis were collected as part of the Healthy Moms study, a randomized clinical trial testing a weight management intervention for obese pregnant women.⁵ The Kaiser Permanente Northwest's Institutional Review Board and an independent data safety and monitoring board approved the study protocol and consent procedures. All participants in the Healthy Moms had a body mass index (BMI) of 30 kg/m² or greater at the start of their pregnancy (mean baseline BMI = 36.72), and they entered the study at 12 to 20 weeks into their pregnancy (mean gestational age = 14.9 weeks).

Healthy Moms participants were members of Kaiser Permanente Northwest (KPNW), a nonprofit, group-model HMO that provides comprehensive medical care to 480,000 individuals in Southwest Washington and Northwest Oregon. All pregnant KPNW patients receive comprehensive prenatal care. Women with low risk pregnancies have prenatal visits at 8, 12, 16, 20, 24, 28, 32, 36, 38, 39, 40, and 41 weeks of pregnancy. Additional visits are added as needed for women with comorbidities (e.g., diabetes mellitus) or medical complications of pregnancy. All outpatient care in KPNW is recorded in the patient's EMR. Standard procedures require scale measurement of weight as a vital sign at all medical care visits, and, for those ages 18 years or older, height measurements every five years. Weights and heights are measured with the patient wearing indoor clothing, but without shoes. These guidelines for recording body weight and height measurements in clinical care were essentially the same as the procedures used for obtaining weight and height measurements for the Healthy Moms research project. However, there were differences in the way the weight and height measurements were collected. Specifically, the research weight measurements were taken using an electronic scale (Scale-Tronix model 5002), and the accuracy of this scale was assessed monthly. Weights recorded in patient medical records were made using a variety of scales, including balance beam mechanical scales and various digital scales. The medical record did not indicate the type of scale used for weights taken in the course of routine practice. Both research- and clinic weights were likely collected at different times of day. Research heights were measured using a calibrated, wall-mounted stadiometer with the participant standing shoeless on a firm, level surface, with her head in the horizontal (Frankfort) plane. Height was measured to the nearest 0.1 cm. In contrast, medical clinic heights were typically measured using the height assessment slide attached to a weight scale.

Weight and Height Data Comparisons

For this analysis, we used all available medical record weight measurements collected within ± 14 days of weight measurements made in the research project. This window was recommended by the study's obstetrician as the maximum length of time during which weight would not change dramatically, thus maximizing the number of available pairs of weights for analysis. For height, we used all available medical record height measurements that were collected within ± 5 years, except for heights collected before age 18. In a case where more than one medical record weight or height was available, we chose the medical record measure that was closest in time to the research measure. Applying these criteria, pairs of research- and medical record weights were obtained for 102 research participants, which constituted 90 percent of the full sample of 114 Healthy Moms participants. For height, measurements were available for all 114 research participants.

Statistical Analyses

We examined the agreement between research- and medical record weights and heights using three indices designed for continuous measures: the intraclass correlation coefficient (ICC), concordance correlation coefficient (CCC), and limits of agreement. We calculated the ICC using a two-way random effects model for the reliability of a single observation for absolute agreement, which treats participants and weight/height measurements as random effects. The CCC, which is similar to the Pearson correlation coefficient in that it measures the linear relationship between two variables, also takes into account any departure from a line of perfect agreement. Values for the ICC can vary from 0 to 1, where 0 indicates no agreement and 1 perfect agreement. The CCC can vary from -1 to +1, where -1 indicates perfect discordance, 0 no concordance, and +1 perfect concordance. The limits of agreement consists of a plot based on the difference between each pair of weights/heights (bias), the average of the research- and medical record weights/heights, and a reference interval in which 95 percent of the differences are expected to lie (this is determined by the product of the standard deviation of the differences and the critical two-tailed z value of ± 1.96). A limits of agreement plot that demonstrates a bias near zero, narrow reference interval, and no discernible pattern between the average weight/height and difference in weight/height is evidence for strong agreement and the exchangeability of the measures.

Findings

Research participants included in the analysis ($n = 114$) had a mean age of 31.8 years ($SD = 4.8$). They were primarily non-Hispanic (91 percent) and white (86 percent). The mean gestational age at measurement of the research weight and height was 14.8 weeks (range = 6.9 to 21.1). The mean gestational age for the medical record weight was 14.6 weeks (range = 7.3 to 20.6) and the absolute difference in days between the research- and medical record weight measurements was 5.6 ($SD = 4.1$). The absolute difference in days between the research- and medical record height measurements was 668.9 ($SD = 380.5$).

The mean research weight was 99.3 kg (218.9 lb), ($SD = 15.6$ kg or 34.4 lb), and the mean medical record weight was 99.2 kg (218.7 lb, ($SD = 15.5$ kg or 34.1 lb). A paired t-test for the difference between these weights was not significant, $M = .2$, $SD=1.8$; $t(101) = 1.1$, $p = .29$. The correlation between the absolute difference in days between the weight measurements and the difference in weights was extremely small ($r = .01$, $p = .94$), so controlling for the elapsed time between measures would have not changed the paired t-test results.

The mean research height was 1.646 m (64.8 inches), ($SD = 0.062$ m or 2.4 inches), and the mean medical record height was 1.654 m (65.1 inches, ($SD = 0.063$ m or 2.5 inches). A paired t-test for the difference between these heights was significant, $M = .008$ m, $SD = .020$ m; $t(113) = 4.4$, $p < .001$. The correlation between the absolute difference in days between the height measurements and the difference in heights was negligible ($r = -.12$, $p = .21$).

For weight, the ICC and CCC had similar estimates of .999 and 95 percent confidence intervals [.998, .999]. A scatterplot of the research- and medical record weights graphed along with the line representing perfect agreement is presented in Figure 1.

The limits of agreement plot for weight is presented in Figure 2. The 95 percent reference interval (indicated by the top and bottom dashed horizontal lines) is between -1.5 kg and 1.7 kg. There were no large outliers, and the directionality of the differences was balanced. Finally, there does not appear to be any pattern in the variability of the differences throughout the range of participants' weights.

The ICC for height was .942 (95 percent CI [.900, .964]) and the CCC was .941 (95 percent CI [.921, .962]). A scatterplot of the research- and medical record heights graphed along with the line representing perfect agreement is presented in Figure 3.

The 95 percent limits of agreement for height was -.031 m to .047 m, and is presented in Figure 4. There was one major outlier, and a sensitivity analysis of its removal resulted in an increase in the ICC and CCC to .959 and a reduction in the -.025 m to .040 m. As with weight, the directionality of the differences between research- and medical record heights was balanced, and there does not appear to be any pattern in the variability of the differences.

Figure 1. Scatterplot of research and clinic weights in pounds

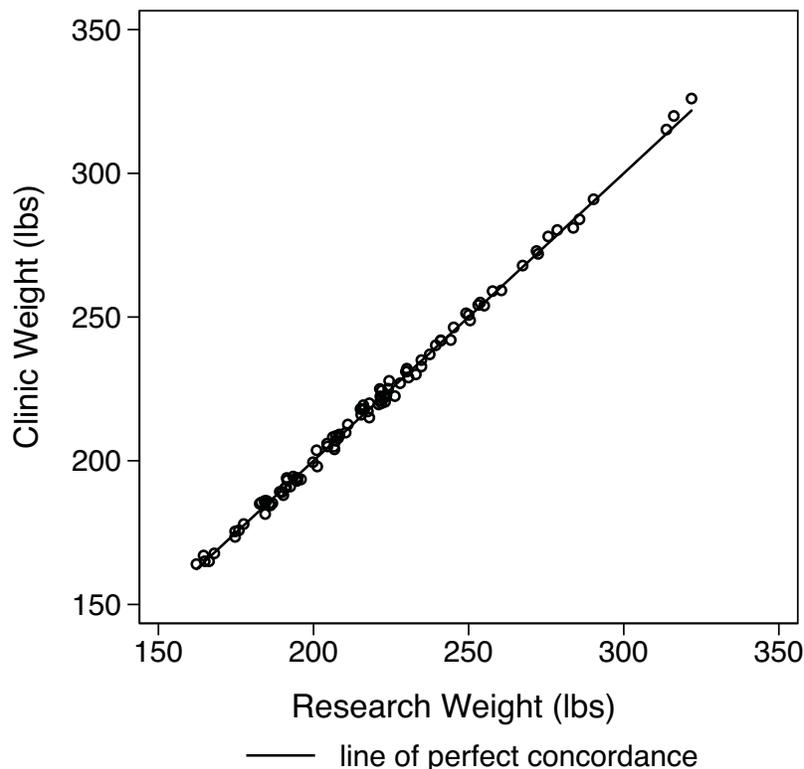


Figure 2. Bland and Altman plot of the mean of clinic and research weight (x-axis) and difference between research and clinic weights (y-axis) with overall mean difference (long dashed line) and lower and upper bounds for the 95% limits of agreement (short dashed lines) expressed in pounds

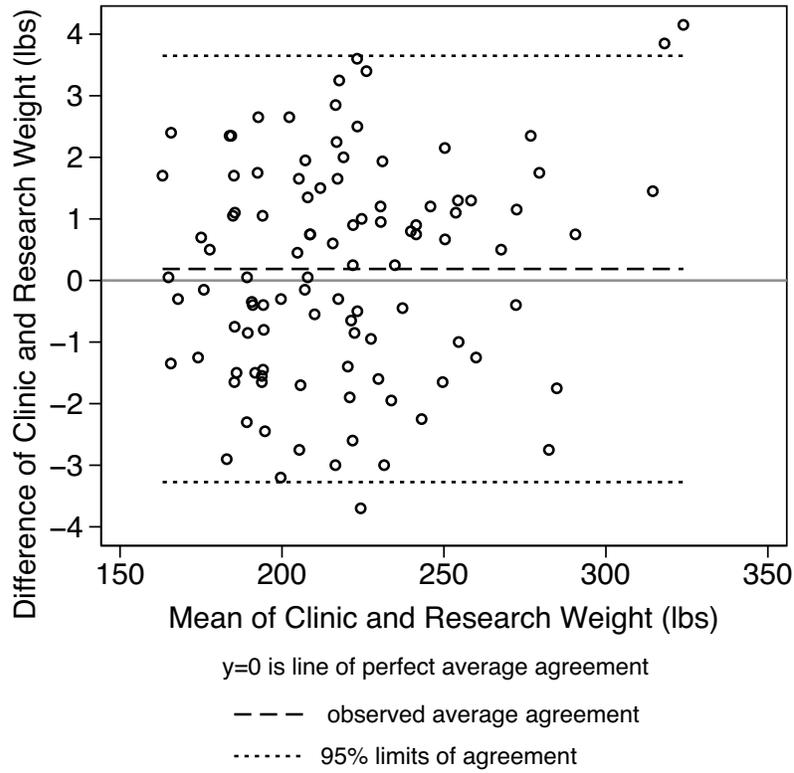


Figure 3. Scatterplot of research and clinic heights in inches

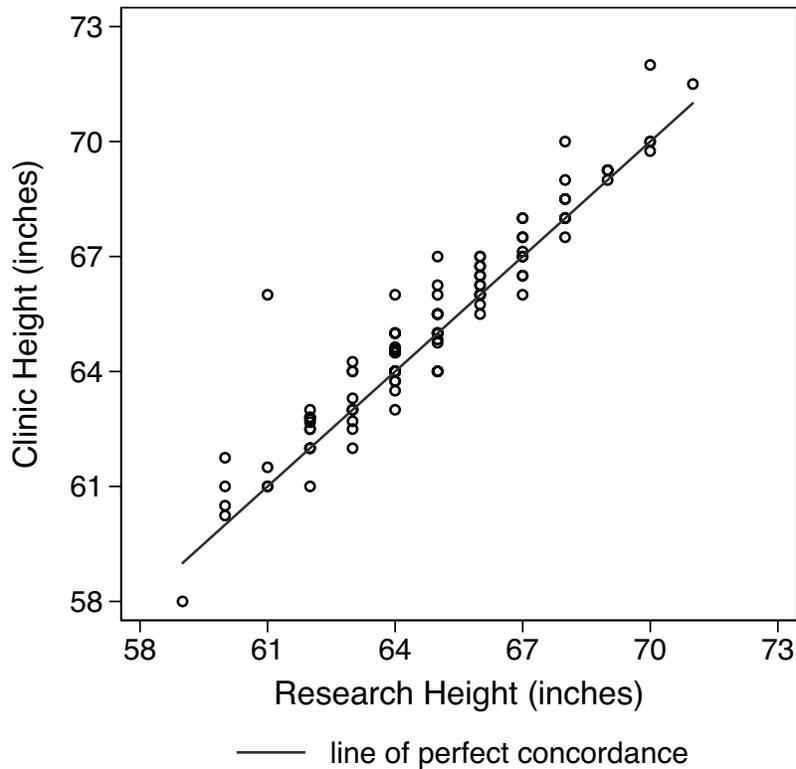
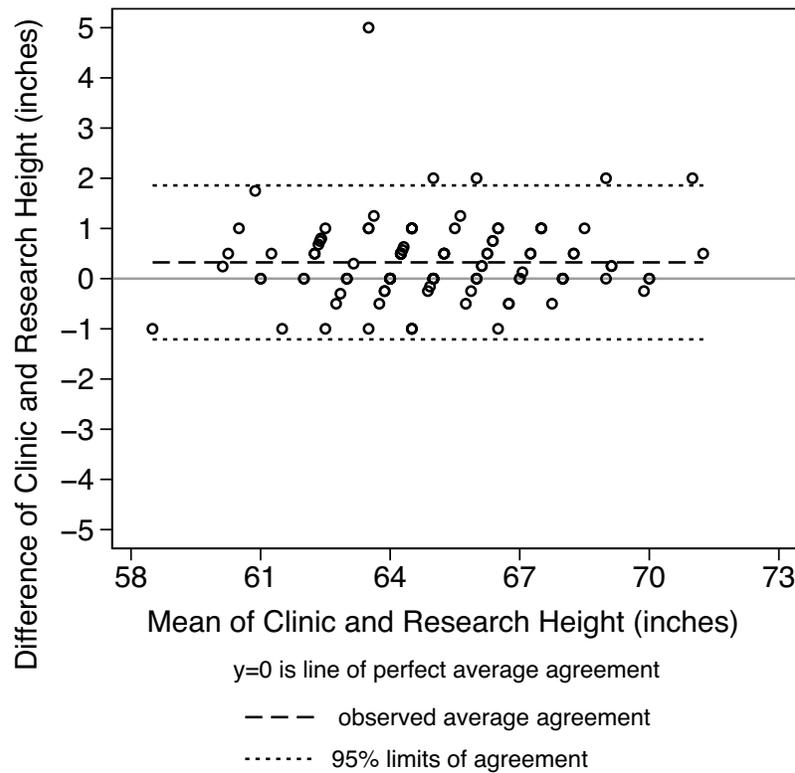


Figure 4. Bland and Altman plot of the mean of clinic and research height (x-axis) and difference between research and clinic heights (y-axis) with overall mean difference (long dashed line) and lower and upper bounds for the 95% limits of agreement (short dashed lines) expressed in inches



To demonstrate the expected impact the error in these measurements could have on BMI, we provide an example using the average medical record height ($M = 1.654$ m) and weight ($M = 99.2$ kg). Using these point estimates, the BMI is 36.3. Using the 95 percent limits of agreement for height and holding weight constant, 95 percent of BMI estimates can be expected to range from 34.3 to 37.7. Using the 95 percent limits of agreement for weight and holding height constant, 95 percent of BMI estimates can be expected to range from 35.7 to 36.9. Given the functional form of the BMI, the expected measurement error in BMI as a function of measurement error in weight is constant throughout the weight and height range, whereas the error in BMI as a function of measurement error in height will increase as weight increases or height decreases. Thus, height not only contributes to more error in BMI estimates compared to weight, but the error is greater at larger BMIs. In a simulation of all possible pairs of height between 1.4 and 2.0 m and weights between 40 and 200 kg, we calculated BMI and then classified value based on WHO cutoffs for underweight, normal, overweight, obesity I, II, and III. We also calculated BMIs for each pair using the lower and upper bounds of the limits of agreement for weight and height, while holding the other constant, and then made WHO classifications based on these values to represent a plausible worst case scenario. We then computed the raw agreement between the classifications of the “true” BMI value with its potential lower and upper bounds. When we varied weight and held height constant, lower bound estimates

were misclassified 4.7 percent and upper bound estimates were misclassified 5.1 percent. When we varied height and held weight constant, lower bound estimates were misclassified 14.6 percent and upper bound estimates were misclassified 9.5 percent.

Discussion

This study found only minimal measurement error for both heights and weights recorded in patient medical records. Weights documented in the medical record are as accurate as research weights obtained within 14 days, and medical record weights can be used interchangeably with weights collected by researchers. Studies conducted among nonpregnant adults have also shown a high degree of association between research- and medical record weights for adults.²⁻⁴ These previous studies used a wider time window between the research- and medical record weights (ranging from 30 to 90 days), but we selected a narrower window of + 14 days given the rapid change in weight typically experienced during pregnancy.

Compared to the weight measures, there was a bit more discrepancy for height data. Whether the detected error would preclude the use of these data for research purposes would depend on the specific research question. An error of 2.5 cm (one inch) roughly translates into one BMI point, and therefore could result in misclassification of patients when categorizing them into defined

BMI categories (e.g., WHO). On the other hand, in a large population study, measurement error of that magnitude would probably be acceptable. In such a case, it would be advisable to conduct a validity study on a sample of cases to assess measurement error.

There are some limitations to this analysis. We only included weights measured in early pregnancy (late first and early second trimesters), prior to the third trimester when the majority of fetal weight gain occurs and when women may experience more dramatic fluctuations in weight due to gestational edema. While our analysis was limited to obese pregnant women receiving care within a single health system, it seems unlikely that measurement error would be larger for women of lower BMIs who would have lower weight measures. In addition, given that different health care settings may use different EMR systems, as well as different equipment and calibration procedures for measuring weight and height, it is important that validation work be conducted in other settings. However, the results of our analysis support findings from previous studies that suggest that adult body weights recorded in the medical record are among the most accurate assessments compared to other measures and acceptable for research use. Heights documented in the course of medical care may be acceptably accurate for large population studies.

There are many possible applications for using heights and weights recorded in EMRs. A few examples include real-time monitoring of obesity rates in whole patient populations, evaluating the comparative effectiveness of obesity prevention and treatment interventions, longitudinal studies of obesity development, and longitudinal studies of osteoporosis and change in adult height.

Acknowledgements

This work was supported by a grant from the National Institute of Child Health and Human Development R01 HD058061.

References

1. Arterburn DE, Alexander GL, Calvi J, Coleman LA, Gillman MW, Novotny R, Quinn VP, Rukstalis M, Stevens VJ, Taveras EM, Sherwood NE. Body mass index measurement and obesity prevalence in ten U.S. health plans. *Clin Med Res* 2010 December;8(3-4):126-30.
2. 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.
3. DiMaria-Ghalili RA. Medical record versus researcher measures of height and weight. *Biol Res Nurs* 2006 July;8(1):15-23.
4. Arterburn D, Ichikawa L, Ludman EJ, Operskalski B, Linde JA, Anderson E, Rohde P, Jeffery RW, Simon GE. Validity of Clinical Body Weight Measures as Substitutes for Missing Data in a Randomized Trial. *Obes Res Clin Pract* 2008 December;2(4):277-81.
5. Vesco KK, Karanja N, King JC, Gillman MW, Perrin N, McEvoy C, Eckhardt C, Smith KS, Stevens VJ. Healthy Moms, a randomized trial to promote and evaluate weight maintenance among obese pregnant women: Study design and rationale. *Contemp Clin Trials* 2012 July;33(4):777-85.