scientific reports

Check for updates

OPEN

Derivation and validation of the Ottawa weight loss prediction model for patients on a low-calorie diet

Robert R. M. Dent[3] & Carl van Walraven[1,2]✉

Accurate weight predictions are essential for weight management program patients. The freely available National Institutes of Health Body Weight Planner (NIH-BWP) returns expected weights over time but overestimates weight when patients consume a low-calorie diet. This study sought to increase the accuracy of NIH-BWP predicted weights for people on low-calorie diets. People enrolled in a weight management program were included if they received meal replacements with defined caloric content for the 3 months of the weight loss phase of the program. The Ottawa Weight Loss Prediction Model (OWL-PM) modelled the relative difference between observed and NIH-BWP predicted weights using longitudinal analysis methods based on patient factors. OWL-PM was externally validated. 1761 people were included (mean age 46 years, 73.3% women) with a mean (SD) baseline weight in pounds and body mass index of 271.9 (55.6) and 43.9 (7.4), respectively. At the end of the program's weight loss phase, people lost a median (IQR) of 17.1% (14.8–19.5) of their baseline weight. Observed weight relative to NIH-BWP predicted weights had a median value of – 4.9% but ranged from – 32.1% to + 28.5%. After adjustment, weight *over*estimation by NIH-BWP was most pronounced in male patients, people without diabetes and with increased observation time. OWL-PM returned expected weights at 3 months that were significantly more accurate than those from NIH-BWP alone (mean difference observed vs. expected [95% CI] 6.7lbs [6.4–7.0] vs. 12.6lbs [12.1–13.0]). In the external validation cohort (n = 106), OWL-PM was significantly more accurate than NIH-BWP (mean squared error 24.3 vs. 40.0, p = 0.0018). OWL-PM incorporated patient-level covariates to significantly increase weight prediction accuracy of NIH-BWP in patients consuming a low-calorie diet.

Accurate weight predictions are very helpful for managing patients in a weight loss program. It helps with counselling patients about what they can expect from a particular weight management program. This may ensure realistic weight loss goals and possibly improve program adherence. In addition, notable deviations from accurate expected weights could help identify issues that must be addressed for successful weight management. Patients losing *less* weight than expected could have issues with program adherence, unidentified endocrine issues, or unaddressed edema from congestive heart failure or cirrhosis. Patients losing *more* weight than expected might be taking metabolic stimulants, have hyperthyroidism, or be presenting with an unidentified eating disorder. Accurate weight predictions over time are an essential tool for physicians monitoring patients in weight management programs.

One freely available weight prediction model is the National Institutes of Health Body Weight Planner (NIH-BWP) by Hall et al.[1] It is a non-linear model that uses a person's age and sex, baseline weight and height, activity level, and daily caloric intake to return individualized weights over time. In a previous study of 3703 adherent people taking a low-calorie meal replacement in a weight management program, we found that the NIH-BWP returned accurate body weight estimates but tended to *over*estimate weights.[2] After 7 weeks of dietary intervention, the mean (SD) relative difference between observed and NIH-BWP predicted weights was -1.8% (3.5). There was also notable variation between patients in the accuracy of NIH-BWP predicted weights; the relative

[1]Epidemiology and Community Medicine, Ottawa Hospital Research Institute, University of Ottawa, Carling Ave, Ottawa, ON ASB1-003 1053K1Y 4E9, Canada. [2]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada. [3]Division of Endocrinology, Department of Medicine, University of Ottawa, OTTAWA, Canada. ✉email: carlv@ohri.ca

difference between observed and expected weights after 7 weeks ranged from − 6.3% to + 2.8%, even after excluding the highest and lowest 10th percentiles.

Reasons for discrepancies between NIH-BWP predictions and observed weights are unknown. We wondered whether weight prediction accuracy might be increased by modelling differences from NIH-BWP predictions. This study attempted this using patient-level covariates of patients in a supervised weight loss program to determine whether we could increase weight prediction accuracy.

Methods
Study setting and program
This was a cohort study with prospective external validation approved by the Ottawa Health Science Network Research Ethics Board (OHSN-REB). All data collection and analyses were performed in accordance with OHSN-REB regulations. Data used in this analysis were solely from patients who provided written informed consent for its use. Data were collected prospectively at the Ottawa Hospital Bariatric Centre and recorded in a database.[3] This centre provides an intensive 26-week weight management skill-building program consisting of weekly 3-h group sessions facilitated by registered dietitians, social workers, behaviourists, and exercise specialists. Patients paid for the program up to 2010, after which it was funded by the Ontario Ministry of Health.

The core program at the Ottawa Hospital Bariatric Center of excellence is an intensive skill building course in weight management that runs weekly for 6 months. The weight loss phase of one of its programs consists of 12 weeks on Optifast 900 meal replacement (3). Patients admitted to this program have a body mass index exceeding 30 kg/m², or a body mass index exceeding 27 kg/m² with comorbidities such as diabetes, lipid abnormalities, sleep apnea, or hypertension. The program's first intervention (week 0) was a 1200 kcal/day diet for which participants were given detailed instructions. From weeks 1–13, participants consumed exclusively a program-provided 900 kcal/day meal replacement program (Appendix A), paid for by the participants.

Creation of derivation cohort
The derivation cohort was created from 5057 people who enrolled in the program between 1992 (the year that the program started) and 2015 (final year of data collection prior to creation of research database) were eligible for study inclusion. Appendix B illustrates the reasons patients were excluded from the derivation cohort. Excluded patients did not differ from those kept in the study by any covariate listed in Table 1. The final derivation cohort included people who were adherent to the program meal replacement of known caloric content for 13 weeks. Participant adherence to the prescribed diet was elicited each week; people were deemed non-adherent if: meal replacement usage was less than 80% or exceeded 100% recommended; dietary non-adherence was documented in any weekly physician notes; or people attended less than two thirds of the weekly meetings.

Data collection
Baseline weights were calculated as the mean of weights taken at program intake (week − 1) and program initiation prior to any dietary intervention (week 0). In the derivation group, patient weights were measured at the end of each week on the same scale at each weekly program meeting.

NIH-BWP expected weights were generated using the online webpage calculator accessed in August 2022.[4] This model used patient age, sex, height, baseline weight, activity level and daily caloric intake to return daily expected weights over the specified time frame. Because all people included in the analysis were adherent to the prescribed diet, calculations used daily caloric intake that was exclusively from the recommended (week 0) or provided diets (weeks 1–13). Patient activity level was defaulted to the lowest strata available for the NIH-BWP (i.e. a physical activity level of "1.4", described on the website as 'sedentary') for the following seven reasons:

1. This was the approach used in a weight prediction validation study by one of the model's developers.[5]
2. In the 2017–18 Canadian Community Health Survey, 101 668 people self-reported data required to calculate body mass index (BMI) and average energy expenditure (Appendix C). These 2 variables were very strongly associated with each other; patients with a BMI exceeding 40 (class 3 obesity, the most common BMI group in our cohort) had a median daily energy expenditure (25.7 Mets*Minutes) that was less than 71% lower than people with a normal BMI. 25.7 METS*Minutes is equivalent to less than 15 min of dish washing[6].
3. 168 patients (9.5%) in our cohort reported a significant physical barrier, such as paraparesis or significant arthritis, that prohibited exercise.
4. Of the 929 patients in our cohort who reported work-based activity levels on a 5-point Likert scale, 648 (69.8%) described themselves as 'very inactive' or 'inactive'.
5. Through informal discussions with program participants, we found it very uncommon for people to participate in physical activity substantial enough to importantly modify weight.
6. Patients were recommended to minimize all physical activity during the first 6 weeks of the weight loss program to avoid pre-syncopal or syncopal episodes while adjusting to their low-calorie diet.
7. A meta-analysis of 343 studies[7] involving diet alone vs. diet plus aerobic exercise in obese patients found no significant difference in weight loss between those with diet alone (5.1 kg [standard error of the mean 0.5]) vs. diet plus aerobic exercise (5.5 kg [0.7]).

Patients treated for diabetes mellitus were classified as having weight neutral medications if their treatment consisted solely of biguanides, or acarbose (glucose-like peptide 1 agonists and sodium-glucose cotransporter-2 inhibitors were not available in Canada at the time of the study).

Variable	Derivation			Validation
	End-program weight change relative to NIH-BWP			
	Less than median	More than median	Total	
	N = 894	N = 867	N = 1,761	N = 106
Mean age, in years (SD)	46.5 ± 11.7	45.8 ± 10.0	46.2 ± 10.9	48.8 ± 12.6
Female	728 (81.4%)	562 (64.8%)	1,290 (73.3%)	75 (70.8%)
Mean baseline weight, in pounds (SD)	268.0 ± 53.4	276.0 ± 57.6	271.9 ± 55.6	273.8 ± 55.8
Mean height, inches (SD)	65.4 ± 3.3	66.3 ± 3.8	65.9 ± 3.6	66.5 ± 3.5
Mean baseline body mass index (SD)	43.9 ± 7.7	43.9 ± 7.1	43.9 ± 7.4	43.3 ± 7.1
Mean baseline girth, inches (SD)	48.4 ± 10.9	49.3 ± 7.3	48.8 ± 9.3	-
Smoking status				
Never	458 (51.2%)	398 (45.9%)	856 (48.6%)	67 (63.2%)
Previous	361 (40.4%)	387 (44.6%)	748 (42.5%)	33 (31.3%)
Current	75 (8.4%)	82 (9.5%)	157 (8.9%)	6 (5.7%)
Obstructive Sleep Apnea status				
None	708 (79.2%)	629 (72.5%)	1,337 (75.9%)	62 (58.5%)
OSA, on CPAP	110 (12.3%)	144 (16.6%)	254 (14.4%)	23 (21.7%)
OSA, not on CPAP	76 (8.5%)	94 (10.8%)	170 (9.7%)	21 (19.8%)
Diabetes status				
None	593 (66.3%)	576 (66.4%)	1,169 (66.4%)	88 (83.0%)
Diet controlled	156 (17.4%)	182 (21.0%)	338 (19.2%)	4 (3.8%)
Weight neutral meds only	53 (5.9%)	33 (3.8%)	86 (4.9%)	13 (12.3%)
Non-Weight neutral meds	92 (10.3%)	76 (8.8%)	168 (9.5%)	1 (0.9%)
Oral glucocorticoids	10 (1.1%)	7 (0.8%)	17 (1.0%)	1 (0.9%)
Antipsychotic	15 (1.7%)	12 (1.4%)	27 (1.5%)	3 (2.8%)
Mean TSH, (SD)	2.2 ± 1.1	2.1 ± 1.0	2.2 ± 1.1	2.4 ± 1.4
Mean HDL, (SD)	1.2 ± 0.3	1.1 ± 0.3	1.2 ± 0.3	1.2 ± 0.3
ATPIII metabolic syndrome criteria met				
Blood pressure	621 (69.5%)	639 (73.7%)	1,260 (71.6%)	74 (69.8%)
Girth	892 (99.8%)	865 (99.8%)	1,757 (99.8%)	106 (100%)
Blood glucose	352 (39.4%)	380 (43.8%)	732 (41.6%)	52 (49.1%)
High density lipoprotein	553 (61.9%)	539 (62.2%)	1,092 (62.0%)	68 (64.2%)
Triglycerides	352 (39.4%)	377 (43.5%)	729 (41.4%)	47 (44.3%)
Number of criteria met				
0	1 (0.1%)	1 (0.1%)	2 (0.1%)	0 (0%)
1	64 (7.2%)	53 (6.1%)	117 (6.6%)	6 (5.7%)
2	221 (24.7%)	176 (20.3%)	397 (22.5%)	18 (17.0%)
3	275 (30.8%)	275 (31.7%)	550 (31.2%)	39 (36.8%)
4	226 (25.3%)	240 (27.7%)	466 (26.5%)	27 (25.5%)
5	107 (12.0%)	122 (14.1%)	229 (13.0%)	16 (15.1%)

Table 1. Description of study cohort.

Creating and evaluating the Ottawa weight loss prediction model (OWL-PM)

Differences between observed weights and those predicted by the NIH-BWP were expressed as percent relative weight differences (%RWD), calculated as:

$$\%RWD = \frac{100 * (O - \widehat{E_N})}{\widehat{E_N}} \tag{1}$$

where O is the observed weight and $\widehat{E_N}$ is the expected weight from the NIH-BWP. Negative %RWD values occurred when the observed weights were less NIH-BWP expected weights (i.e. NIH-BWP *under*estimated weight).

We then used longitudinal data analysis methods to model the %RWDs at each week as a function of patient-level covariates. This longitudinal model was created using PROC MIXED (SAS, Cary NC) to cluster data within patients. We used the maximum likelihood method to estimate covariance parameters and used an unstructured covariance structure since all subjects had their weights measured on a weekly basis and missing observations were minimal. Continuous variables were centered. Total degrees of freedom used for Type 3 significance tests were corrected for clustering of data within patients using the general Satterwaite approximation.

Construction of the Ottawa Weight Loss Prediction Model (OWL-PM) started by determining *a-priori* the order in which we added variables to the model based on clinical experience and previous studies. Each variable was sequentially added to the model in the specified order along with an interaction term of that variable with both week and week[2] to capture potential linear and quadratic effects, respectively, over time. We used the likelihood ratio test to determine if the addition of these terms importantly improved the model fit; criteria for significantly improved model fit was a decrease in the − 2 log likelihood of at least 3*(# degrees of freedom added to the model). If this threshold was not met, modeling was repeated (and the likelihood ratio test recalculated) after excluding the quadratic interaction term (i.e. variable*week[2]). If *this* model did not significantly improve fit, modeling was repeated (and the likelihood ratio test recalculated) after excluding the linear interaction term (i.e. variable*week). Modelling was complete when no other candidate variables met these inclusion criteria.

The OWL-PM outputted the *expected* %RWD (equation [1]) as a function of the model's covariates. This value was back-transformed to the predicted body weight with the following equation:

$$\widehat{WT} = \frac{(XB * \widehat{E}_N)}{100} + \widehat{E}_N \tag{2}$$

where $\widehat{WT}$ is the predicted body weight, XB is the expected %RWD from OWL-PM and $\widehat{E}_N$ is the expected weight from the NIH-BWP.

We quantified NIH-BWP and OWL-PM error using mean squared error (MSE):

$$\sum \left(O - \widehat{WT} \right)^2 / n \tag{3}$$

where O is the observed body weight, $\widehat{WT}$ is the predicted body weight from NIH-BWP or the OWL-PM, and n is the total number of people in the analysis.

Creation external validation cohort

The external validation cohort was also prospectively created from patients entering the program between 2021 and 2023. Appendix B illustrates the external validation cohort creation. The weight loss program in the external validation group differed from that used in the derivation group by three factors: (i) Patients proceeded directly to the Optifast® 900 meal replacement; (ii) Weight loss program lasted 6 weeks instead of 13; and (iii) Participant weights were measured at home.

We used methods from Archer et al.[8] to calculate the sample size required for the validation cohort. Regressing week-13 OWL-PM predicted weights on observed weights returned an adjusted R^2 value of 96.1% (i.e. 96.1% of the variation in observed weights was explained by the model). We expected that this would drop to 90% in the validation patients. Assuming good calibration, we used a slope in the observed vs. expected regression of 1 with a standard error of 0.035. These values returned a sample size estimate of 92 which we rounded up to 100.

Results

The derivation cohort included 1761 people (Table 1). They were middle-aged (mean age 46.2 years) with almost three quarters (73.3%) being female. Mean baseline weight (standard deviation, SD) was 271.9 pounds (55.6) with a mean (SD) body mass index of 43.9 (7.4). Almost half of patients (48.6%) never smoked while approximately one tenth (8.9%) were current smokers. Almost a quarter of patients (24.1%) had a diagnosis of obstructive sleep apnea, with almost two-thirds of these people receiving continuous positive airway pressure treatment. One third of the cohort had a diagnosis of diabetes mellitus with most of these people on diet treatment only. Of the ATPIII criteria for metabolic syndrome, all except four people met girth criteria while blood glucose and triglyceride criteria were the least commonly met (at 41.6% and 41.4%, respectively).

Weight loss in the derivation cohort was considerable (Fig. 1A). At the start of the program, the median weight (25th–75th percentile or interquartile range [IQR]) was 261 lbs (232–303); at week 13, the median weight had dropped to 216 lbs (192–249). The relative weight change [calculated as 100*(weight-baseline weight)/baseline weight] decreased progressively during the program, attaining a median value of − 17.1% (− 14.8%, − 19.5%) by week 13 (Fig. 1B). Variance between patients in relative weight change increased over time (Fig. 1B).

Overall, weight loss in the derivation cohort exceeded that predicted by the NIH-BWP (Fig. 1C). By program's end, the median percent relative weight difference (%RWD, Eq. 1) was − 4.9% (IQR − 2.6 to − 7.1%). When we stratified the derivation cohort by this amount, patients in the two strata were similar except that patients having *less* than the median end-program weight change were more likely to be women (81.4% vs. 64.8%, Table 1).

The Ottawa Weight Loss Prediction Model (OWL-PM) is presented in Table 2. It contained program week along with 8 other covariates including: age, sex, baseline weight, thyroid stimulating hormone level, total number of ATPIII metabolic syndrome criteria met, and status of smoking, diabetes, and obstructive sleep apnea (Table 2). All covariates significantly interacted with program week or the squared term for program week, indicating that their influence on the percent relative weight difference (%RWD—Eq. 1) changed over time. Four factors decreased the %RWD (i.e. were associated with *greater* weight loss than that predicted by NIH-BWP): program week (− 9.65); an additional ATPIII metabolic syndrome criterion being met (− 0.28); previous or current smoking status (former: − 0.68; current: − 0.04); and having a diagnosis of obstructive sleep apnea (on CPAP: − 0.80; not on CPAP: − 0.58). Five factors increased the %RWD (i.e. were associated with *less* weight loss than that predicted by NIH-BWP): being female (+1.71); greater baseline body weight (+0.41); increased age (+0.15); greater baseline thyroid stimulating hormone levels (+0.08); and the presence of diabetes (diet controlled: +0.16; treated with weight neutral medications: +1.86; treated with non-weight neutral medications: +1.43). The most influential factors on body weight included program week, sex, and diabetes status.
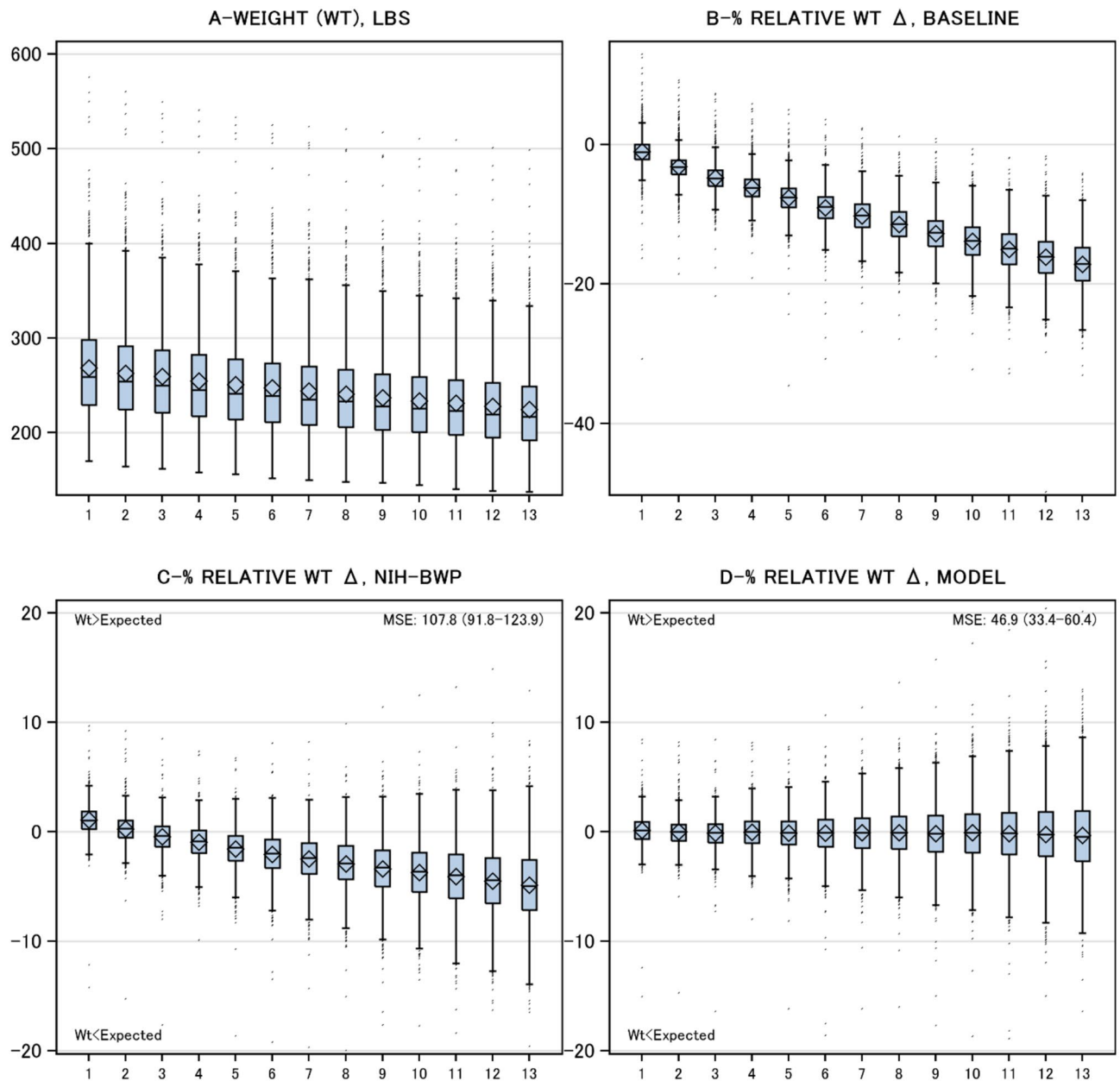
Figure 1. Weight changes during program. Plots summarize distributions in study cohort of 4 different weight statistics by program week (horizontal axis): **A** Cohort weight in pounds (LBS); **B** Percent relative weight difference compared to baseline (calculated as 100*[weight-baseline]/baseline); **C** Percent relative weight difference compared to NIH-BWP estimate (Eq. 1); **D** Percent relative weight difference compared to model estimate (Table 2). All plots summarize statistics using Tukey plots in which the box's middle, lower end, and upper end indicate the median, 25th percentile, and 75th percentile, respectively, and the whiskers extend 1.5 times the interquartile range (i.e. length of box) from each end of the box. The mean and standard deviation is indicated by the diamond in each box plot.

Predicting weights using the OWL-PM rather than the NIH-BWP had two important effects. First, relative differences between observed and weights predicted by OWL-PM had median and mean values centering on zero throughout the study (Fig. 1D). In contrast, mean or median relative weight differences with the NIH-BWP deviated progressively below zero over time (Fig. 1C). Second, accuracy of weight prediction was significantly greater with OWL-PM compared to the NIH-BWP with a significantly lower patient-specific mean squared error (46.9 [95%CI 33.4–60.4] vs. 107.8 [91.8–123.9]). At program's end, the mean absolute difference (95%CI) between observed and predicted weight was 6.7lbs (6.4–7.0) for OWL-PM vs. 12.6lbs (12.1–13.0) for NIH-BWP.

The 106 people in the validation cohort were similar to the derivation group except they had a higher prevalence of obstructive sleep apnea and a lower prevalence of both smoking and diabetes (Table 1). Weight prediction by the OWL-PM was significantly more accurate than that by NIH-BWP (Fig. 2), with significantly lower

Variable	Parameter estimate (standard error)	P-value	Independent mean change in % relative weight difference† at program's end due to factor
Intercept	1.848 (0.17)	<.0001	–
Program week increased by 1	− 0.762 (0.04)	<.0001	− 9.65
(Program week increased by 1)^2	0.02 (0.003)	<.0001	
Female	− 0.224 (0.116)	0.0531	+1.71
Female*Week	0.302 (0.027)	<.0001	
Female*Week 2	− 0.012 (0.002)	<.0001	
Baseline Weight^A increased by 25lbs	− 0.007 (0.001)	<.0001	+0.41
(Baseline Weight^A increased by 25lbs)*Week	0.001 (0)	<.0001	
(Baseline Weight^A increased by 25lbs)*Week^2	− 0.00003 (0)	0.0594	
Age^B increased by 10 years	− 0.011 (0.004)	0.0089	+0.15
(Age^B increased by 10 years)*Week	0.002 (0.001)	0.0021	
TSH^C increased by 1 IU	0.058 (0.042)	0.1645	+0.08
(TSH^C increased by 1 IU)*Week	0.024 (0.01)	0.0156	
(TSH^C increased by 1 IU)*Week^2	− 0.002 (0.001)	0.0074	
# of ATPIII MS Criteria increased by 1	− 0.027 (0.045)	0.5566	− 0.28
(# of ATPIII MS Criteria increased by 1)*Week	− 0.039 (0.011)	0.0003	
(# of ATPIII MS Criteria increased by 1)*Week^2	0.001 (0.001)	0.0292	
Smoking status			
Never	REF	–	–
Never*Week	REF	–	
Never*Week^2	REF	–	
Previous	0.006 (0.093)	0.9871	− 0.68
Previous*Week	− 0.042 (0.022)	0.0718	
Previous*Week^2	− 0.001 (0.001)	0.0525	
Current	− 0.02 (0.159)	0.9871	− 0.04
Current*Week	− 0.066 (0.038)	0.0718	
vCurrent*Week^2	0.005 (0.002)	0.0525	
Diabetes status			
None	REF	–	–
None*Week	REF	–	
None*Week^2	REF	–	
Diet controlled	− 0.038 (0.122)	0.6608	+0.16
Diet controlled*Week	0.044 (0.029)	0.1545	
Diet controlled*Week^2	− 0.002 (0.002)	0.0161	
Weight neutral meds only	0.199 (0.212)	0.6608	+1.86
Weight neutral meds only*Week	0.101 (0.05)	0.1545	
Weight neutral meds only*Week^2	0.002 (0.003)	0.0161	
Non-Weight neutral meds	0.111 (0.167)	0.6608	+1.43
Non-Weight neutral meds*Week	0.022 (0.039)	0.1545	
Non-Weight neutral meds*Week^2	0.006 (0.002)	0.0161	
Obstructive sleep apnea status			
None	REF	–	–
None*Week	REF	–	
OSA, on CPAP	− 0.045 (0.125)	0.9353	− 0.80
OSA, on CPAP*Week	− 0.058 (0.019)	0.0052	
OSA, not on CPAP	− 0.019 (0.143)	0.9353	− 0.58
OSA, not on CPAP*Week	− 0.043 (0.022)	0.0052	

Table 2. The Ottawa weight loss prediction model (OWL-PM). †Compared to relative weight difference predicted by NIH-BWP. The OWL-PM is a longitudinal analysis model that predicts the relative difference between observed weights and those predicted by the NIH-BWP. The expected percent relative weight difference (Eq. 1) from the OWL-PM is calculated by multiplying patient-specific variable values by the corresponding parameter estimate. This value can be back-transformed to a predicted weight using Eq. 2 ^{A}Centered on 270lbs ^{B}Centered on 45 years ^{C}Centered on 2.2 IU. (lbs = pounds; TSH = thyroid stimulating hormone; IU = international units; ATPIII = 3rd adult treatment panel; MS = metabolic syndrome; OSA = obstructive sleep apnea).
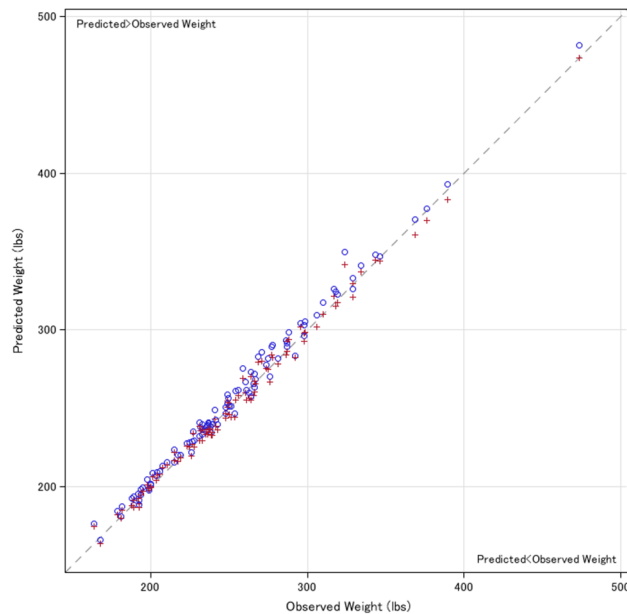
Figure 2. Observed vs. predicted body weights in validation cohort. Observed (horizontal axis) and predicted (vertical axis) weights in pounds for NIH-BWP (blue circles) and OWL-PM (red crosses) are plotted. Perfect agreement is indicated by dashed diagonal line. Final observations for validation cohort are presented. Mean squared error was significantly lower for OWL-PM (17.7 [95% CI 12.3–23.2] vs. 21.8 [14.3–29.3]; Student's t-statistic 3.20, p-value 0.0018).

end-program patient-specific mean squared errors (24.3 [95%CI 16.5–32.2] vs. 40.0 [24.7–54.6]; *t*-statistic 3.20, p-value 0.0018).

Discussion

Predicting patient weight is essential for managing patients weight management programs. This study closely followed a large cohort of patients taking a defined low-calorie intake for 3 months. We found that the NIH-BWP weight prediction model returned accurate weight predictions but systematically *under*estimated weight loss. We used longitudinal data analysis to measure the association of patient-level covariates with differences between observed and NIH-BWP predicted weights. The resulting Ottawa Weight Loss Prediction Model (OWL-PM) significantly improved weight prediction accuracy. Improved weight prediction using OWL-PM was confirmed in an external validation cohort.

We believe that our study makes several important points. First, we believe our modelling approach of quantifying the influence of covariates on deviations between actual and NIH-BWP predicted weight is innovative. With this analytical strategy, we took expected weights from a previously validated, widely available model—the NIH-BWP—and modelled its residuals using important patient-level factors to significantly increase weight prediction accuracy. We believe that this approach could be used by researchers to modify and improve previously created prediction models. Second, the two freely available weight prediction models include the NIH-BWP and the Pennington Biomedical Research Center Weight Loss Predictor (PBRC-WLP). In our previous analysis[2], we found that the NIH-BWP was significantly more accurate at predicting weight over time than the PBRC-WLP, with the latter more extensively over-estimating body weight. Since the OWL-PM is significantly more accurate than NIH-BWP, we can conclude that it is likely more accurate than PBRC-WLP. Third, even though our model significantly increased the accuracy of weight prediction, we still saw important variation between patients in the difference between observed and predicted weights. The end-program relative difference between the observed weight and OWL-PM expected weight (Table 2) ranged from − 26.7 to 38.2%. Further work is needed to determine if other factors, such as genetic traits or proteomic measures, could be added to the model to shrink this variation.

Several issues should be kept in mind when interpreting our results. First, although our study included a large cohort of patients accumulated over more than 22 years and was validated in a subsequent group of people, we consider it is essential that our model's performance be assessed in another patient population. This would ensure that the model is not biased by factors unique to our cohort.

Second, while our model significantly increases the accuracy of weight estimates, its application is more cumbersome than most predictive models because it requires three steps. The first step involves the multiplication of patient-specific covariate values with the respective parameter estimates in Table 2 to return the expected percent relative weight difference (Eq. 1). In the second step, patient-specific NIH-BWP weights are calculated (based on their age, sex, baseline weight, height, daily caloric intake, and physical activity) using the NIH-BWP webpage (https://www.niddk.nih.gov/bwp). In the third and final step, these two values are inserted into Eq. 2 to return OWL-PM predicted patient weights. This difficulty of application could impede the model's update;

readers can request from the authors a spreadsheet program which eases implementation of the model. Third, the validation population used to test the OWL-PM were slightly different than that used to derive the model having a higher prevalence of obstructive sleep apnea and a lower prevalence of both smoking and diabetes (Table 1). It is unclear how these differences influenced model performance; however, since each of these covariates are in the OWL-PM, it is likely to be small. Finally, patients in our cohort were very heavy with a mean (standard deviation) baseline body weight and BMI of 272 (56) pounds and 43.9 (7.4), respectively. For the 7 reasons specified in the methods, we defaulted the physical activity variable in all patients to sedentary when applying the NIH-BWP model. Appendix D presents the percent relative weight difference (%RWD, equation [1]) at study's end with varying activity levels entered into the NIW-BWP model. For these calculations, we selected a patient with a %RWD value of − 4.94%, which is the median %RWD value for the entire cohort (Fig. 1C). Appendix D indicates that the default activity level of 1.4 returned a %RWD of − 4.94 at week 13. An NIH-BWP activity level of 1.6 resulted in a %RWD of − 2.68. The systemic weight *over*estimation by NIH-BWP (as seen in Fig. 1C) is eliminated when each person was defaulted to an activity level of 1.8. This is described on the website as someone whose occupation involves cleaning or delivering mail on foot and whose leisure activity involves walking or bicycling to work at least weekly. Therefore, future studies will return more accurate predicted patient weights over time by incorporating accurate patient activity levels into the NIH-BWP model. However, difficulties associated with using self-report to quantify energy expenditure should be recognized. Dowd et al.[9] identified 35 studies comparing self-reported energy expenditure to double labelled water measure and found study-specific mean relative differences ranged from approximately − 75% (self-report *under*estimated energy expenditure) to approximately +115% (self-report *over*estimated energy expenditure). Note should be made that these summary statistics capture only the *average* direction of bias in self-reported energy expenditure in each study and do not calculate the *total* individual differences between self-reported and objectively measured energy expenditure amongst all participants.

In summary, weight prediction in people consuming low-calorie diets can be made significantly more accurate by incorporating additional patient-level covariates. Future validation studies of the OWL-PM should be conducted to confirm its accuracy.

Data availability
The data may be available pending review and approval by the Ottawa Health Science Network Research Ethics Board. Dr. van Walraven can be contacted regarding all data requests.

References
1. Hall, K. D. *et al.* Quantification of the effect of energy imbalance on bodyweight. *Lancet* **378**, 826–837 (2011).
2. Dent, R. M., Harris, N. & van Walraven, C. Validity of two weight prediction models for community-living patients participating in a weight loss program. *Sci. Rep.* **13**, 11629 (2023).
3. Dent, R. M., Penwarden, R. M., Harris, N. & Hotz, S. B. Development and evaluation of patient-centered software for a weight-management clinic. *Obes. Res.* **10**, 651–656 (2002).
4. National Institutes of Diabetes and Digestive and Kidney Diseases. Body Weight Planner. 2021. 1–16–2021. (Online Source)
5. Brady, I. & Hall, K. Dispatch from the field: is mathematical modeling applicable to obesity treatment in the real world?. *Obesity* **22**, 1939–1941 (2014).
6. Jette, M., Sidney, K. & Blumchen, G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin. Cardiol.* **13**, 555–565 (1990).
7. Miller, W. C., Koceja, D. M. & Hamilton, E. J. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. *Int. J. Obes. (Lond)* **21**, 941–947 (1997).
8. Archer, L. *et al.* Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Stat. Med.* **40**, 133–146 (2021).
9. Dowd, K. P. *et al.* A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. *Int. J. Behav. Nutr. Phys. Act.* **15**, 15 (2018).

Author contributions
B.D. and C.W. conceived, analyzed, and wrote the study. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68454-z.

Correspondence and requests for materials should be addressed to C.W.

Reprints and permissions information is available at www.nature.com/reprints.

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