A novel algorithm for describing population level trends in body weight

Azadeh Alimadad1,2*, Carrie Matteson3, Warren L. Hare2,4, Ozge Karanfil5, Diane T. Finegood3

1Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada; *Corresponding Author: aalimada@sfu.ca
2Modeling of Complex Social Systems Program, Simon Fraser University, Burnaby, Canada
3Department of Biomedical Physiology & Kinesiology, Simon Fraser University, Burnaby, Canada
4Mathematics, Irving K. Barber School of Arts and Sciences, University of British Columbia, Okanagan Campus, Kelowna, Canada
5Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA

Received 1 November 2012; revised 9 December 2012; accepted 24 December 2012

ABSTRACT

Modeling population trends and predicting the impact of interventions to address obesity requires algorithms for predicting body weight status in the future. Predictions can be based on statistical consideration of different risk factors, or be an extrapolation of past and current trends. Despite the well known correlation between previous and future weight, individual weight history has not been used to predict future trends. We developed a novel population-level model to examine trends of different classes of body weight considering individual body weight histories from the National Longitudinal Survey of Youth (NLSY79). A subset of data used to assess the predictive ability of our proposed model with actual data. Our results confirm the importance of weight history in determining future weight status. Over 80% of individuals in a specific weight category (normal, overweight, obese) will stay in the same weight category after two years (except overweight females). The length of body weight stability was also found to be important. The probability of remaining normal weight increased with longer prior periods of being at a normal weight over 18 years (0.834 to 0.893). We demonstrate that an individual's most probable weight class in the future is consistent with their maximal historical weight class.

Keywords: Body Weight; Modeling; Obesity

1. INTRODUCTION

Obesity and its associated chronic diseases are the leading causes of preventable death worldwide [1]. The obesity epidemic has grown rapidly over the last few decades and is associated with a growing morbidity and mortality due to diabetes, cardiovascular disease, and cancer [2,3]. To understand the implications for health care systems and public health programs, policy makers need a means of projecting these rates and their related costs into the future. Modeling provides a means of making these projections, as well as an approach to examining options for slowing or reversing current trends [4,5].

Regression analysis is a common modeling technique that has been used to extrapolate the prevalence of overweight and obesity into the future. Wang et al. used linear regression to project the prevalence of overweight and obesity among US adults [6]. Their projection suggests that by 2048 all Americans will be overweight or obese. One challenge of this approach, however, is in defining the time interval over which historical weight status can be extrapolated into the future [4]. The challenge is due in part to the dynamics of environmental factors like the availability and affordability of food, marketing, accessibility of public transport, and the structure and influence of social networks. In addition, interventions introduced to address overweight and obesity may influence trends into the future [7].

Markov modeling is another approach that has been suggested for predicting trends related to obesity and diabetes [5,8-10]. Markov models would consider individuals as belonging to groups based on a parameter like body mass index (e.g. normal weight, overweight or obese) and population trends are determined by the probability of transitioning from one group to another. Simple Markov models require the Markov assumption, that the transition from one state to another is only dependent on the current state and not any previous states. However, it is well known that weight status is correlated to prior weight status [11,12], making the Markov assumption questionable.

In the present paper we examined the validity of the
Markov assumption using data from the National Longitudinal Survey of Youth (NLSY’79). We calculated the transition probabilities between different weight categories over an 18 year period. Our analysis shows that transition probabilities between BMI categories are not solely dependent on an individual’s current weight status. In addition, we introduce a new approach, the Maxhist model, which extends the basic Markov model by including conditional probabilities, where the probability of being in a specific weight category depends on the highest historical body weight.

2. METHODS AND PROCEDURE

2.1. Description of Data

The National Longitudinal Survey of Youth (NLSY79) dataset is a nationally representative sample of 12,686 young men and women born in the late 1950s and early 1960s in the United States. These individuals were interviewed annually from 1979 to 1994 and then biennially from 1994 to 2006. Each individual reported his or her weight at each time point and their height in 1981, 1982 and again in 1985. Body Mass Index (BMI = Weight/\( \text{Height}^2 \)) was calculated using the reported weights and the highest reported height for each individual. We calculated the BMI biennially between the years 1986 and 2004, for all individuals who were 21 years or older in 1986. We excluded 491 subjects for a missing date of birth, 1 subject was excluded for missing height, and 200 subjects we excluded due to pregnancy. Since only 5 individuals were born in 1955, 1956 and 1965, we excluded them as well. The remaining number of subjects in this dataset was 11,989. Two subsets of the data were used for this study.

Subset 1: This set of the data was used to calculate transition probabilities between 2 time steps and to explore the effect of an individual’s historical BMI in driving their future BMI using the Markov and Maxhist models. We were interested in how individuals transitioned between different classes of BMI. As such we excluded subjects who did not have any consecutive reported weights (n = 1261). The remaining sample had 10,728 individuals (50% female). It was used to explore the effect of an individual’s historical BMI in driving their future BMI.

Subset 2: This subset was used to test the accuracy of our Maxhist hypothesis regarding probabilities of particular patterns of population weight transition over time.

To verify the accuracy of the Maxhist hypothesis, we used a dataset in which each individual’s BMI was known at every time point. Therefore we excluded individuals if their weight data was missing at any time step. This reduced our sample size to 5316 individuals (49% female). To ensure that this smaller sample is an accurate representation of the larger population, transition probabilities were also calculated on subset 1 using the Maxhist algorithm. Estimated values for transition probabilities were similar in both the larger and the reduced dataset.

2.2. Definition of Weight Classes and Age Groups

In order to calculate transition probabilities from one BMI category to another we defined three classes of body weight: NO: Normal weight (BMI < 25), OW: Overweight (25 ≤ BMI < 30), and OB: Obese (BMI ≥ 30) [5].

The dynamics of weight gain and weight loss in the NLSY79 dataset was investigated with longitudinal data that spanned the interval from 22 to 48 years of age for people born between 1957 and 1964. All subjects of the same age were grouped together regardless of their date of birth due to the small range of years in which subjects were enrolled in the study. Therefore, our group of 24 year olds would consist of individuals who were born in 1962 and were 24 in 1986, plus those individuals who were born in 1964 and were 24 in 1988, etc. We verified that there was no birth cohort effect by comparing the BMI trend lines for each birth cohort.

2.3. Design of the Markov Model

To explore the trends in obesity at the population level, we considered a simple Markov model with 3 states: Normal weight (NO), Overweight (OW), and Obese (OB). We calculated the probability of individuals’ movements between these three states (transition probabilities) as follows. At each time interval, individuals are assigned to a state and the probability of staying in that state or moving to one of the other 2 states is calculated for each class from the data in subset 1 (Figure 1, Simple Markov Model, left hand side of the figure). Calculations reflect only the current state of the system and not prior time points. In essence the calculation is “memory-less” and follows the Markov assumption. Since the NLSY79 dataset was collected every two years, we use this time interval as our natural time step (known as Markov cycle) between transitions. As time steps represent two year, we assume that any state can transition into any other state between each time step.

We also developed a higher order Markov model in order to test the basic Markov assumption and evaluate whether this higher order model behaves in the same fashion as the three state model. The three state model becomes a nine state model, where each state stores not only the individual’s current BMI state, but also their BMI state from the previous time step (Figure 1, Nine state Markov Model, full figure). This tests the basic
premise of the “memoryless” property of the Markov model.

2.4. Development of the Maxhist Model

To address the failure of the simple and the 9 state Markov models, we developed a new method for defining the state space by extending our basic model to include a conditional probability. The Maxhist model is based on the premise that an individual’s most probable weight class in the future is determined by their maximal historical weight class. The Maxhist function considers an individual’s BMI history for n time steps and returns the maximal historical BMI. We calculate the conditional probability of individual’s movements based on the result of Maxhist function (Figure 2).

2.5. Validation of the Maxhist Algorithm

We used two approaches to examine the validity of the
Maxhist algorithm. In the first approach, we determined the number of times the Maxhist hypothesis was broken. This equals the frequency of cases where OW individuals go on to become NO even though their maximum historical body weight was OW and the number of times OB people go on to become NO or OW even though their maximum historical weight was OB.

In order to assess the predictive validity of the proposed Maxhist model, we applied a split-sample validation method by separating the dataset into two parts: the first 10 years and the next 8 years of data on each subject. The first part of the dataset was used to calculate transition probabilities, which were then applied to the subsequent eight years to predict the outcomes. The predictions of the Maxhist model were calculated using MATLAB (version 7.5.0.338-R2007b). Values were compared to the results of a simple Markov model and a linear extrapolation based on the first 10 years of data. All three results were compared to the actual body weight classifications in the second half of the dataset. To measure the goodness of fit between the actual and predicted percentage of NO, OW and OB, the square root of the sum of the square differences are calculated between the age of 34 and 40 (see Equation (1) below).

3. RESULTS

3.1. Markov Transition Probabilities

The probabilities that an individual in each BMI category (NO, OW, OB) moves to any of these three states after two years (one time step) for males and females in the NLSY79 data subset #1 are shown in Table 1.

For example, men who are OB have a 14.1% chance of becoming OW in the subsequent two years. The probability that individuals could have an extreme weight change over two years (e.g. NO to OB) is very rare in both male and female groups (0.5% and 0.8% respectively). Likewise the transition from OB to NO is rare but slightly higher for women as compared to men (M: 0.7% and F: 1.7% respectively).

The results demonstrate that there is an over 80% chance that a male in the NO weight category will stay in the same weight category, and an almost 80% chance that an OW male will stay in the OW category in the next two-years. This probability is even higher for the OB, as we find that 85% of individuals are likely to remain OB after two years. Women in the NO and OB categories behave similarly to men, but the transition probability for an overweight woman to remain OW is only 65%, whereas it is almost 80% for men. Thus, there is an almost 15% difference between males and females in this category, which suggests that women are more susceptible to weight fluctuations than men, and in both directions.

The above results are based on two years weight transitions, which shows the current weight of individuals can play an important role in driving their future weight status. If an individual’s future body weight depends only on their current body weight and is independent of their previous body weight, we could use a simple Markov model to project individual’s body weight over time.

3.2. Evaluating the Markov Assumption

For the basic Markovian principles to hold true, we would have to see that an individual’s previous body weight would have no impact on their future weight. Consider for example the following 3 transitions: 1) NO → NO → NO; 2) OW → NO → NO; 3) OB → NO → NO. If prior weight has no effect on future weight then the second transition from NO → NO should have the same transition probability in all three cases.

Figure 1 illustrates that these probabilities and their 95% confidence intervals are 0.868 (0.863, 0.874), 0.455 (0.428, 0.481) and 0.238 (0.138, 0.337) respectively. Hence, the Markov assumption does not hold. Likewise the Markov assumption does not hold for the transition from OB → OB where an OB prior state increases the probably of remaining OB (0.913 (0.905, 0.921)) as compared to an

\[
\sqrt{\sum_{\text{age=34}} \left(\text{actual}_{\text{age=34}} - \text{predicted}_{\text{age=34}}\right)^2 + \cdots + \left(\text{actual}_{\text{age=40}} - \text{predicted}_{\text{age=40}}\right)^2}
\]
Table 1. Transition Probabilities for each BMI state independent of age (row to column).

<table>
<thead>
<tr>
<th>Current Weight</th>
<th>Transition Probabilities and 99% CI after two years (Male)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>NO</td>
<td>0.804 (0.795, 0.812)</td>
</tr>
<tr>
<td>OW</td>
<td>0.091 (0.085, 0.097)</td>
</tr>
<tr>
<td>OB</td>
<td>0.007 (0.005, 0.010)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current Weight</th>
<th>Transition Probabilities and 99% CI after two years (Female)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>NO</td>
<td>0.857 (0.851, 0.864)</td>
</tr>
<tr>
<td>OW</td>
<td>0.168 (0.158, 0.179)</td>
</tr>
<tr>
<td>OB</td>
<td>0.017 (0.013, 0.021)</td>
</tr>
</tbody>
</table>

Data for transition probabilities are presented with 99% confidence intervals in the parenthesis.

3.3. Maxhist Model

Given the importance of prior body weight, we examined a new model (Maxhist) which calculates transition probabilities for a future transition based on the maximum historical body weight. Transition probabilities were calculated for each possible subset of historical data, i.e. for prior knowledge of 2, 4, 6, etc. or up to 16 prior years of information (at 2 year intervals). Table 2 provides the results for the maximum historical period available (16 years) in the NLSY79 dataset (other data not shown). This table demonstrates that the highest probabilities are for transitions that are determined by the highest historical body weight, e.g. Maxhist (NO) → NO is 0.892, Maxhist (OW) → OW is 0.712 and Maxhist (OB) → OB is 0.758.

Not surprisingly, the length of historical body weight maintenance in a single category increases the probability of remaining in that category as illustrated in Table 3.

The probability of remaining in NO increased with longer prior periods of being NO (from 0.834 to 0.893), and the probability of transitioning to OW decreases slightly with longer periods of NO weight maintenance (from 0.159 to 0.104). The probability of transitioning to OB from NO is low regardless of the length of time being NO prior to the transition (less than 1%).

To check the robustness of these results, we repeated this analysis with the smaller dataset #2 (individuals with weight recorded at every time step) who had 18 years of data. For each individual, we counted the number of times when BMI was lower than the maximum historical weight. Well over half of the subjects in this dataset never broke the Maxhist hypothesis and more than 25% break it only once or twice, while a small percentage of all people break the Maxhist hypothesis more than 3 times within an 18-year interval (Figure 3).

3.4. Evaluating the Maxhist Assumption

According to the Maxhist hypothesis, individuals have a strong tendency to return to their highest historical BMI status at the population level, however, there are individuals who don’t necessarily return to their highest historical weight. This is reflected for example in the fact that even individuals with a Maxhist of OW have a 21.2% chance of becoming NO (Table 2). This would reflect a “break” from the Maxhist hypothesis. An individual who was OB or OW once and then maintained a reduced body weight (e.g. they were OB → NO → NO → NO → NO → NO) would repeatedly break the Maxhist hypothesis because they never returned to their highest historical body weight.

To explore the transitions of the BMI status at the individual level, we used subset #2 (individuals with weight recorded at every time step) who had 18 years of data. For each individual, we counted the number of times when BMI was lower than the maximum historical weight. Well over half of the subjects in this dataset never broke the Maxhist hypothesis and more than 25% break it only once or twice, while a small percentage of all people break the Maxhist hypothesis more than 3 times within an 18-year interval (Figure 3).

3.5. Model Simulation, Validation and Predictions

In order to validate the various model predictions, we used the full 18-year weight history data available on a subset of the subjects. The dataset was split into the first 10 years which was used to predict weight for the last 8 years. Since weight status was known for these last 8 years, we were able to calculate the sum-of-squares difference between the percentage of people predicted in each weight category and the actual number for the Maxhist, regression and Markov models. Figure 4 illustrates the deviation of the models from actual for normal weight individuals. Between the dashed lines, one can
Table 2. Probability of being in NO, OW, or OB given maximum historical BMI in the last 16 years is NO, OW, or OB.

<table>
<thead>
<tr>
<th>Maximum historical BMI within the last 16 years</th>
<th>Probability of being in specific BMI 2 years after</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>NO</td>
<td>0.893 (0.882, 0.905)</td>
</tr>
<tr>
<td>OW</td>
<td>0.212 (0.200, 0.225)</td>
</tr>
<tr>
<td>OB</td>
<td>0.027 (0.022, 0.033)</td>
</tr>
</tbody>
</table>

*Data for transition probabilities are presented with 99% confidence intervals in the parenthesis.

Table 3. Probability of being in state NO, OW, or OB given that the maximal historical BMI state is NO.

<table>
<thead>
<tr>
<th>Time step</th>
<th>P (NO) given Maxhist = NO</th>
<th>P (OW) given Maxhist = NO</th>
<th>P (OB) given Maxhist = NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.834 (0.829, 0.839)</td>
<td>0.159 (0.154, 0.164)</td>
<td>0.007 (0.006, 0.008)</td>
</tr>
<tr>
<td>4</td>
<td>0.860 (0.855, 0.865)</td>
<td>0.135 (0.130, 0.140)</td>
<td>0.004 (0.003, 0.005)</td>
</tr>
<tr>
<td>6</td>
<td>0.870 (0.864, 0.875)</td>
<td>0.127 (0.121, 0.132)</td>
<td>0.004 (0.003, 0.005)</td>
</tr>
<tr>
<td>8</td>
<td>0.876 (0.870, 0.822)</td>
<td>0.121 (0.115, 0.126)</td>
<td>0.003 (0.002, 0.004)</td>
</tr>
<tr>
<td>10</td>
<td>0.882 (0.876, 0.888)</td>
<td>0.115 (0.109, 0.121)</td>
<td>0.003 (0.002, 0.004)</td>
</tr>
<tr>
<td>12</td>
<td>0.886 (0.878, 0.893)</td>
<td>0.111 (0.104, 0.118)</td>
<td>0.003 (0.002, 0.005)</td>
</tr>
<tr>
<td>14</td>
<td>0.889 (0.881, 0.897)</td>
<td>0.108 (0.100, 0.116)</td>
<td>0.003 (0.002, 0.005)</td>
</tr>
<tr>
<td>16</td>
<td>0.892 (0.882, 0.901)</td>
<td>0.106 (0.096, 0.115)</td>
<td>0.003 (0.001, 0.004)</td>
</tr>
<tr>
<td>18</td>
<td>0.893 (0.882, 0.905)</td>
<td>0.104 (0.093, 0.115)</td>
<td>0.003 (0.001, 0.004)</td>
</tr>
</tbody>
</table>

*Data for transition probabilities are presented with 99% confidence intervals in the parenthesis.

Figure 3. Illustration of the strength of Maxhist hypothesis.

See that the Maxhist model provides a significantly better fit to the actual data than the linear regression, and also better than the three state Markov model. Table 4 presents the sums-of-squares results for each of the models tested for NO, OW and OB.

4. DISCUSSION

Our results confirmed previous studies that demon-
Using longitudinal cohorts, Miller et al. [11-13] demonstrated that prior body weight of individuals can play an important role in defining an individual’s future weight. Using longitudinal cohorts, Miller et al. [11] and Stark et al. [12] demonstrated the risk of being overweight in adulthood is greater for overweight children compared to those who were normal weight as children. Abraham and Nordsieck also showed overweight children (age 10 - 13) are about twice as likely to become overweight adults 20 years later as compared to children with average weight [13]. Consistent with these observations, our study shows that overweight and obese young adults are likely to remain overweight or obese over the next 18 years. In addition, we show that the probability of remaining NO increased with longer prior periods of being NO and individuals who become OW and OB are not likely to return to a lower body weight over time.

This modeling approach helps visualize the need for stronger emphasis on preventive measures and making concerted efforts to curb childhood obesity as early as possible to avoid lengthy time intervals where increased weight status becomes a powerful predictor of continued weight management challenges.

Although the relationship between early body weight and body weight in adulthood was well known, current modeling approaches for projecting future trends in obesity have not taken this relationship into account. Current methods for projecting trends in obesity have used linear extrapolations based on the prevalence of obesity over time. As such these models do not take into account individual weight status. Some authors have suggested that Markov models could be used to consider individual weight status, but we demonstrate that a simple Markov model is not adequate and there is insufficient data available on which to base higher order Markov models.

We developed a novel model (Maxhist) by considering the importance of an individual’s weight history. Our algorithm uses conditional probabilities, where the likelihood of being in a specific weight category depends on the highest historical body weight. The main advantage of the Maxhist model is that it takes into account the weight history of individuals with a minimum amount of data. In theory we could use a higher order Markov model (e.g. 9 state model) but this requires a very large dataset for the calculation of all of the relevant transition probabilities.

We compared the capability of the Maxhist model to predict future weight status to a simple 3 state Markov model and a regression model. Our results demonstrate that Maxhist is more accurate than both the Markov and regression models for predictions over a period of 6 years. In addition, the Maxhist model produced more plausible predictions further into the future. It should be noted however, that there was no significant difference in the projections specifically for OB subjects between Maxhist and Markov or Maxhist and regression model.

Not all individuals followed the Maxhist algorithm perfectly during the 18 years of follow up. When individuals move from a higher weight category to a lower weight category they are considered to break the Maxhist hypothesis. However, few people break the Maxhist hypothesis (by maintaining a lower weight than the previous time step) more than twice during the 18 year follow-up period. Most of the breaks in the Maxhist hypothesis are due to OB individuals becoming OW or OW individuals becoming NO. Women are more likely to break the Maxhist hypothesis than men, indicating a greater propensity for weight fluctuations in females.

Our results are based on the analysis of individuals between 22 and 46 years of age who lived in United States. Future studies need to examine generalizability of our algorithm considering different age categories, social, cultural, environmental and demographic factors. There was no birth cohort effect within our study, although this may be due to the small range of birth dates (1957-1964) of individuals in the dataset.

In summary, our analysis showed that transition probabilities between BMI categories depend on the weight history of individuals. To our knowledge this is the first study that identified a specific pattern for the progression of NO, OW and OB over time. Our results showed that an individual’s most probable weight class in the future is determined by the maximal historical weight class. This new technique provides an avenue for more accurate weight transition predictions which are valuable for personalizing therapeutic approaches and generating improved disease risk and cost estimates adjusted for popu-

<table>
<thead>
<tr>
<th>Weight status</th>
<th>Sum-of-squares difference between the actual weight and predicted values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maxhist</td>
</tr>
<tr>
<td>NO</td>
<td>33</td>
</tr>
<tr>
<td>OW</td>
<td>40</td>
</tr>
<tr>
<td>OB</td>
<td>25</td>
</tr>
</tbody>
</table>
lation prevalence data.

5. ACKNOWLEDGEMENTS

This work is the result of an interdisciplinary collaboration. We thank Dr. Krisztina Vasarhelyi and Dr. Alexander Rutherford for having enabled this connection across departments. We are indebted to Dr. Nate Osgood for his feedback on the analysis. This work was sponsored by grants from the Canadian Institutes of Health Research to Dr. Diane T. Finegood and was supported in part by the SFU CTEF MoCSSy Program. We also are grateful for technical support from the IRMACS Centre, Simon Fraser University. Parts of this study were presented in preliminary form as an abstract at the First Annual Workshop on Dynamic Modelling for Health Policy.

REFERENCES


