Adapting Tools of Causality to Analyze the Obesity Paradox

Prasenjit Mitra (mitra@l3s.de)
L3S Research Center
Xinyuan (Cindy) Zhang
Harvard University
Connor Heaton
Pennsylvania State University

Article

Keywords:

Posted Date: February 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2515102/v1

Additional Declarations: No competing interests reported.

Adapting Tools of Causality to Analyze the Obesity Paradox

Connor Heaton[1], **Xinyuan Zhang**[2], **and Prasenjit Mitra**[1,3,*]

[1]College of IST, Pennsylvania State University
[2]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School
[3]L3S Research Center, Leibniz University Hannover
[*]mitra@l3s.de

ABSTRACT

The prevailing sentiment in the general population is that a Body Mass Index (BMI) above desired levels leads to less than optimal health outcomes and increased likelihood of death. Recent medical studies, however, have found evidence that the opposite is true for a certain subset of the population, particularly those with a history of cardiovascular disease (CVD). These studies found that individuals with a BMI *slightly above desired* experience more optimal health outcomes compared to those with a BMI in the *desired* range - a phenomenon known as the "obesity paradox." However, these studies have primarily been performed using classical tools of probability and statistics. In this study, we apply tools appropriated from causal inference, with some loosened restrictions, to the $age \geq 65$ population of the Kailuan dataset, a longitudinal dataset tracking almost 15,000 patients, 10,005 of which have a history of CVD , 4,641 have a BMI *slightly above desired*, 5,863 have a BMI in the *desired* range, and 317 a BMI *slightly below desired*. We ultimately find some evidence of the obesity paradox in this dataset.

Introduction

The recent increase in "health awareness" in the general population is motivated in part by the understanding that our weight, or body mass, can have a large influence on our health trajectory. Medical professionals have introduced the notion of Body Mass Index (BMI) to quantify the relation between an individual's weight and height, which can shed light on the overall health of the individual. The equation for BMI is given equation 1 below.

$$BMI = \frac{mass_{kg}}{height_m^2} = \frac{mass_{lb}}{height_{in}^2} \times 703 \tag{1}$$

BMI can be used to classify individuals as either *underweight*, *desired weight*, *overweight*, or *obese* as presented in table 1. In studies across the general population, it has been shown that individuals with BMI < 18.5 and in the range of $25 - 29.9$, *underweight* and *overweight* respectively, have shorter life expectancy from age 40 than individuals with BMI in the range $18.5 - 24.9$, the *desired weight*. The association between BMI and overall mortality is often called "J-shaped" in that a low BMI (< 18.5) puts you at moderate risk for mortality, but the risk decreases as BMI increases before reaching an inflection point where risk increases again as BMI approaches 25 and higher[1].

BMI Value	BMI Category
< 18.5	Underweight (OB 0)
18.5-24.9	Desired Weight (OB 1)
25-29.9	Overweight (OB 2)
30-34.9	Obese (OB 3)
≥ 35	Very Obese (OB 4)

Table 1. Typical BMI classification.

The correlation between BMI and mortality risk is not constant across all populations, however, and is in fact reversed in some populations, such as for *overweight* individuals with a history of cardiovascular disease (CVD)[2–4]. CVD refers to a number of conditions including heart disease, heart attack, stroke, heart failure, arrhythmia, or hearth valve problems. Specifically, the obesity paradox posits that for individuals with a history of CVD, having a BMI slightly *above desired* leads to better health outcomes compared to those with *desired* BMI.

While these findings have certainly been insightful, they have been performed using traditional statistical analysis methods, such as the adjusted Cox regression model used by Horwish et al[4]. Ideally, to understand the causal effect of some treatment a randomized controlled trial (RCT) would be performed, but this is not always suitable for a number of ethical and logistical reasons[5]. In the case of exploring the casual effect of BMI, not only would it be ethically questionable to induce patients to gain weight to achieve a certain BMI that may be unhealthy, but artificially inducing a person to gain weight will likely bias the results in some fashion. Furthermore, as noted by Franks et al[6], many weight loss trials, which touch upon BMI, often have very stringent participation criteria, perhaps limiting the degree to which results can be generalized to the population at large.

Causal analysis methods such as the potential outcomes framework[7,8], then, may become appealing for their ability to make use of existing observational data, performing *pseudo-RCT's*. Such methods, however, are typically used to estimate the effect of some *discrete intervention*, i.e. a treatment, while a BMI classification is more so a *state of being*. This then leads to the question of whether or not this difference prohibits the application, or perhaps adaptation, of causal methods for this assay. For causal analysis to yield *valid* results, in general, three conditions must hold[9]:

1. **Stable Unit Treatment Value**: The potential outcomes for any unit do not vary with the treatment assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.

2. **Positivity**: For any value of variable X, treatment assignment is not deterministic; Each participant has a non-zero probability of being assigned to each treatment.

3. **Ignorability**: Given the background variable, X, treatment assignment W is independent to the potential outcomes; a hypothetical, unaccounted confounding variable will be evenly distributed among treatment and control groups.

Assumption one will hold in this application as one individual belonging to a certain BMI classification does not impact the BMI classification of some other individual in the study. Although two participants are unlikely to have the same exact *BMI value*, previous studies[9] have demonstrated the validity of grouping real-valued treatments in to discrete classes, such as BMI categories, satisfying assumption one.

Assumption two, positivity, is slightly more difficult to justify in this application, however, given the biological systems at play. For example, according to the CDC[10] , there may be some genetic factors at play, such as the *MC4R* gene, which has been shown to make individuals feel extremely hungry, inducing over-eating (*hyperfagia*). An individual with the *MC4R* gene, then, will be unlikely, although not explicitly prohibited, to be in the *underweight* BMI classification, for example, given their genetic composition. Although genes such as *MC4R* may perhaps impact other aspects of the genetic composition, the gene itself doesn't *directly* induce a high BMI per se. According to the same CDC source, genes associated with obesity influence the signals that are sent to the brain. In the case of *MC4R*, a signal of "extreme hunger" is sent to the brain, encouraging the individual to eat more. The source continues to say that although 50 genes have been associated with obesity, most have a very tiny effect. Furthermore, biological systems may also induce a person to be *underweight*, an effect in the opposite direction, such as repeated portions of chromosome 16[11]. The fact that genetic factors do not explicitly *prohibit* any individual from being in a certain BMI classification, in conjunction with the small effect size of genetic factors when they are present, thus we believe that the assumption of positivity is still valid for our purposes here.

The issues identified with assumption two also cause issues in satisfying assumption three, ignorability, which essentially asserts that "if a variable is not accounted for, it has no impact on the outcome or treatment and is safe to ignore." If some genetic factor influences an individual to feel more hungry more often, it is unlikely that individuals with this genetic factor will be evenly distributed across BMI classifications - we would expect such individuals to appear more often on the higher end of the BMI scale. As the community understands BMI to impact health outcomes, a gene that impacts BMI would be a confounding variable.

Yao et al[9] note that the assumption of ignorability is often difficult to satisfy in practice. As they mention, however, it is possible to use big data to find latent, proxy variables for this unaccounted for, confounding variable. One such big data approach to account for unobserved confounding variables is to use an *autoencoder*[12] to encode the records. An autoencoder essentially constructs a *low dimensional* representation from an original *high dimension* representation while retaining as much of the *important* signal in the original representation as possible. In doing so, the model is able to infer complex relationships between observed and latent confounding variables. While signal describing the unobserved confounding variable would be ideal and preferred, the community has found that latent variables discovered by autoencoders are a reasonable stand-in[9].

We propose adopting tools of causal inference, specifically by relaxing the assumption of positivity and definition of treatment, to study the obesity paradox, hypothesizing that it will not necessarily disprove the notion of the obesity paradox, but help refine the community's understanding of the role that obesity, as indicated by BMI, plays in patients' health prospects. Instead of comparing how average health outcomes differ across two groups, these causal analysis methods pair two *similar* individuals who underwent different *treatment* and compare their health outcomes, then report the average pair-wise difference

as the estimated effect. In doing so, the goal is to isolate the effect that the *treatment* has on the outcome of interest. In this study we relax the definition of *treatment* somewhat, such that it includes the five BMI classifications listed in table 1, and relax the assumptions of positivity slightly. Even if others disagree this technically qualifies as "causal inference" because of these changes, so be it, we believe such experiments can still provide valuable information to the community.

We apply our methodology to the subset of the Kailuan Study dataset, a dataset that follows over 100,000 individuals for up to three check-ins over a period of six years[13]. Specifically, we apply these methods to a subset of the Kailuan dataset including only participants 65 years of age or older with a history of CVD, a subset that includes around 15,000 individuals. Not all participants completed all three check-ins, however, meaning the larger time frame included in the analysis, the fewer number of individual participants are available. Our experiments explore how the tradeoff between temporal range of the data and raw magnitude of participants impacts the stability of our findings. We ultimately find evidence in support of the obesity paradox.

Results

In this section we present results obtained from carrying out two pseudo-RCT experiments. These experiments - experiments A & B - were designed to asses the effect that being *overweight* compared to being of the *desired weight*, expressed as BMI, has on various health outcomes, specifically death, for individuals with and without a history of CVD. Experiment A will estimate the casual effect that being overweight (OB 2), the *treatment*, has on different health outcomes with respect to being the *desired* weight (OB 1), the *control*, for **only** individuals who have a history of CVD. Experiment B will be the same as experiment A, but performed on **all** individuals included in the dataset. Only about a third of the population aged 65 or older **do not** have a history of CVD, which we deem insufficient for performing an experiment on that subset alone given the "big data" nature of this methodology. For that reason, we will use experiment A as a reference point for the results of experiment B.

For each experiment, we explore how stable our findings are when more temporal data, but fewer individual participants, is included in the analysis. In addition to denoting how many check-ins were included in the analysis, we also note how many treatment and control records are in the population and how many pairs of "similar records" resulted, which are used to obtain the Estimated Casual Effect (ECE) on various outcomes. Alongside death, our primary outcome of interest, we explore the ECE of the treatment on Ischemic Stroke, Intracerebral Hemorrhage, and Incident Myocardial Infarction, as they are also included in the original Kailuan dataset. To gauge the impact of the matching process, we also present in parentheses the estimated effect of the treatment that would be obtained by *randomly* pairing records from the treatment and control groups.

Check-ins	0	0 & 1	0, 1 & 2	0, 1, 2 & 3
N Treatment	4,641	2,668	1,422	838
N Control	5,863	3,266	1,826	1,075
N Pairs	4,402	2,488	1,197	541
E.C.E. on Death	-0.038 (-0.026)	-0.031 (-0.027)	-0.051 (-0.0426)	-0.033 (-0.022)
E.C.E. on Ischemic Stroke	0.024 (0.028)	0.03 (0.025)	0.02 (0.028)	-0.009 (0.006)
E.C.E. on Intracerebral Hemorrhage	-0.001 (0.0)	0.001 (-0.002)	0.002 (0.0011)	-0.002 (-0.001)
E.C.E. on Incident Myocardial Infarction	0.012 (0.011)	0.001 (0.006)	-0.001 (-0.009)	-0.015 (-0.012)

Table 2. Results for experiment A, leveraging **only** individuals with a history of CVD. In general, we find *overweight* individuals are less likely to experience death compared to individuals of the *desired* weight. This is in line with the core assertion of the obesity paradox. The results for the other health outcomes are not as conclusive.

To gauge just how "similar" paired records are, we present a visualization of the attributes for the *treatment* and *control* records for each pair obtained in Experiment B in Figure 1. As we can see, both the *treatment* and *control* records from each pair do have similar attributes, both categorical and real-valued. For example, consider the plot for fasting blood glucose in Figure 1. Although there do appear to be clear regions of the plot where the treatment has a fasting blood glucose but the paired control record has a low fasting blood glucose and vice-versa, the majority of points in the scatter plot fall within the circumference of a unit-circle with radius 10, certainly a radius of 15.

Discussion

In analyzing the results from Experiment A presented in Table 2, we see that the ECE of death is consistently negative, and rather stable, across all check-ins. The same cannot be said for the other health outcomes explored, however, as they all see a change in sign across the different temporal scales. It is also worth noting that the estimated effect on these other outcomes found by randomly pairing records varies in sign across check-ins as well. This perhaps suggests that the causal effect of being *overweight* compared to the *desired weight* is marginal or negligible for for these outcomes. Furthermore, we see that the

Check-ins	0	0 & 1	0, 1 & 2	0, 1, 2 & 3
N Treatment	4,887	2,804	1,510	892
N Control	6,565	3,663	2,058	1,214
N Pairs	4,129	2,367	1,157	588
E.C.E. on Death	-0.011 (-0.02)	-0.026 (-0.027)	-0.03 (-0.031)	-0.022 (-0.011)
E.C.E. on Ischemic Stroke	0.03 (0.03)	0.021 (0.025)	0.027 (0.031)	0.0 (0.008)
E.C.E. on Intracerebral Hemorrhage	0.001 (-0.001)	0.001 (-0.002)	0.003 (0.001)	-0.002 (0.002)
E.C.E. on Incident Myocardial Infarction	0.008 (0.012)	0.003 (0.006)	-0.013 (-0.008)	-0.017 (-0.011)

Table 3. Results for experiment B, leveraging **all** individuals in the dataset. In general, we find *overweight* individuals are less likely to experience death compared to individuals of the *desired* weight. While these results are not in line with the traditional understanding of how BMI relates to health outcomes, we believe this can be attributed to the large portion of the dataset (2/3) that have a history of CVD.

ECE on death is consistently higher than the estimate obtained by randomly pairing individuals from the treatment and control groups.

The results of Experiment B presented in Table 3 tell a similar, but slightly different story. Of the possible outcomes analyzed in this assay, the ECE on Death is again the most stable, followed by the ECE on Ischemic Stroke. Additionally, the ECE on Death being negative implies that individuals who are *overweight* are less likely to experience death than those of the *desired weight*. The ECE on Death in Experiment B, which leverages the entire population of the dataset, is lower in magnitude than the ECE on Death identified in Experiment A, which only leverages the subset of the population with a history of CVD. Here, we see that the ECE on death is lower than estimated effect found by *randomly* pairing individuals from the treatment and control populations in all but one sub-experiment for Experiment B, albeit much closer in magnitude compared to Experiment A.

In conjunction, these findings reinforce the notion of the obesity paradox. The ECE on death for being *overweight* as opposed to the *desired weight* is higher (more protective) for the subset of the population in the Kailuan dataset aged 65+ *with* a history of CVD than it is for the whole population aged 65+. In Experiment A, the average ECE on Death across all check-ins is -0.038, while for Experiment B the average ECE on Death is -0.022. The negative estimated effect implies *overweight* individuals are less likely to experience death than individuals of the *desired* weight, and the estimate is higher for the population with a history of CVD.

Finally, comparing the ECE of death with that found by randomly pairing individuals from the treatment and control groups highlights the impact of the matching process. Treating each member of the treatment or control populations as identical results in different results compared to when the unique attributes of each user are taken into account.

Methods

Here we describe the methods we employed to carry out the experiments described above. As mentioned above, we leverage the potential outcomes framework to estimate a causal effect by finding *similar* individuals who underwent different treatment. Participants in the Kailuan study participated in a variable number of check-ins, so we train an auto-encoder to encode individual check-in records, and describe individuals in terms of their available encoded check-in records. All methods were performed in accordance with the relevant guidelines and regulations.

Data Collection

The dataset was generated from the Kailuan Study, an ongoing, prospective, community-based cohort study in North China. Detailed data collection methods in this cohort has been described previously[14,15]. This study was approved by the Ethics Committee of the Kailuan General Hospital. Participants gave their written informed consents on using the de-identified data collected at each check-in.

Encoding Check-In Records

As part of the Kailuan study, at each check-in, participants would complete a questionnaire describing their smoking status, physical activity level, and salt intake among other lifestyle characteristics. Height and weight were measure at each check-in to compute BMI, and blood-samples were collected to identify blood glucose and cholesterol level, among others. From this set of attributes, we identify 26 confounding variables which we control for in our experiments described in Table 4. Some values are real-valued, described by a single number, while others are categorical, described using a one-hot encoding. In total, these 26 features result in a check-in vector that is 61-dimensions wide.

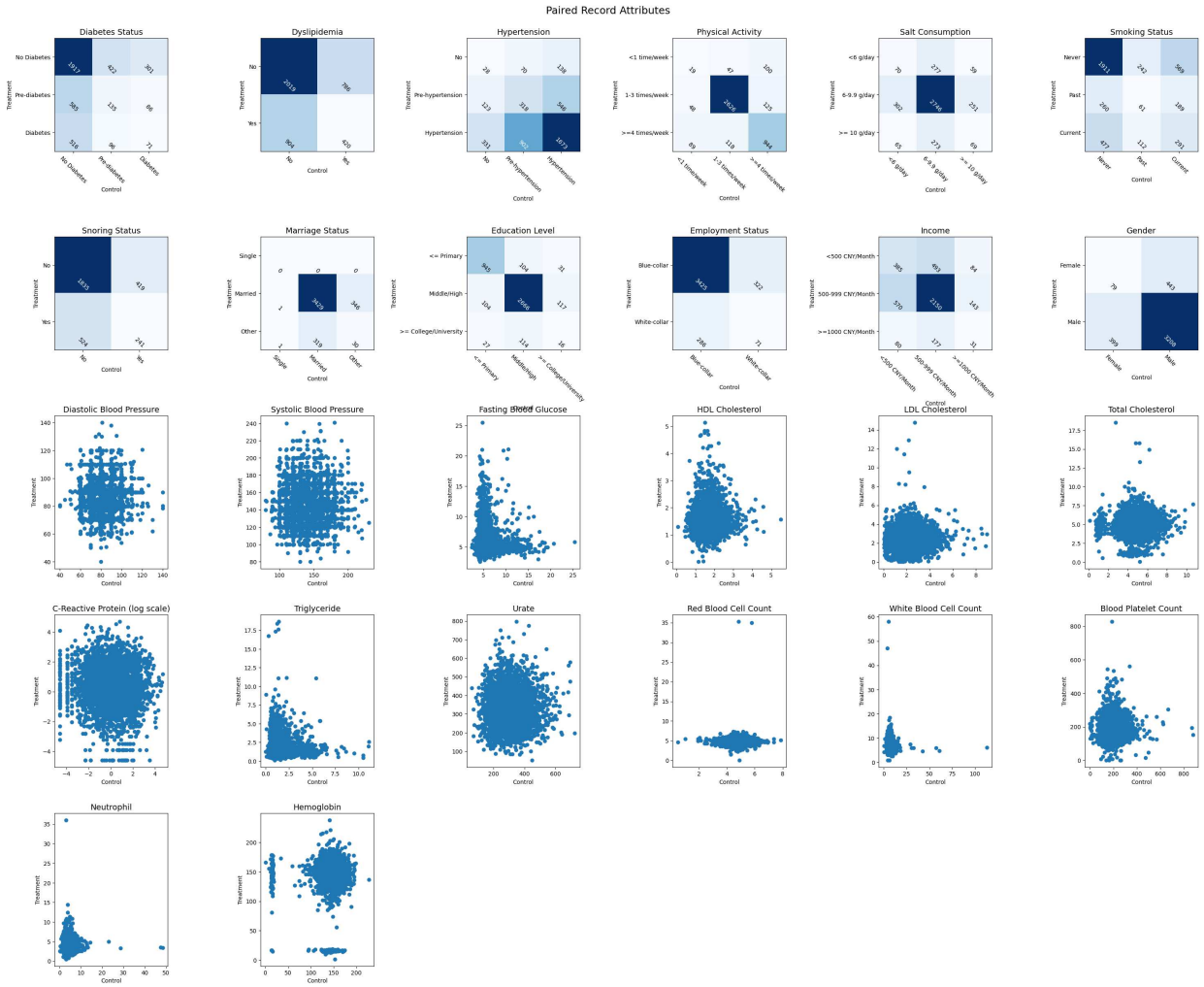

Figure 1. Visualization of attributes between each record in a pair.

#	Confounding Variable	#	Confounding Variable	#	Confounding Variable
1	Diabetes status	11	Physical activity	21	Neutrophil
2	Diastolic blood pressure	12	Salt consumption	22	Marriage status
3	Systolic blood pressure	13	Smoking status	23	Education status
4	Dyslipidemia	14	Triglyceride	24	Employment status
5	Fasting blood glucose	15	Urate	25	Income level
6	Hypertension	16	Snoring status	26	Gender
7	HDL cholesterol	17	Red blood cell count		
8	LDL cholesterol	18	White blood cell count		
9	Total cholesterol	19	Blood platelet count		
10	C-Reactive protein levels	20	Hemoglobin		

Table 4. Identified confounding variables that were controlled for in our experiments.

To encode these features, we train an auto-encoder, taking advantage of the benefits described above in the introduction. Specifically, using the `Pytorch` framework, we train an auto-encoder comprised of an encoder and decoder, each consisting of 7 fully-connected dense layers, that constructs an 8-dimension representation given the 61-dimension check-in record. When describing a participant who completed multiple check-ins we simply concatenate their corresponding 8-dimensional check-in representations. Not all individuals in the dataset participated in each check-in, so 28,869 unique check-in records result from the 15,000 or so individuals aged 65+. Given these 30,000 or so records, we train our autoencoder to minimize the mean squared error (MSE) between the original check-in record and the record recovered by the model. Specifically, we use an Adam optimizer, learning rate of $1e-5$, dropout rate of 0.25, and batch size of 72 when training for 95 epochs.

Matching Records to Estimate Causal Effect

A key aspect of the potential outcome framework is the finding of records that are "similar enough" such that they can serve as counterfactuals for each other. To do so, we *match* records based on their compressed check-in representations obtained from our autoencoder. We determine that two records who underwent different treatments can serve as a counterfactual for one another if the *distance* between their compressed check-in representation(s) is less than some threshold d. When quantifying the distance between records, we use a metric known as Mahalanobis distance. Mahalanobis distance is preferred to other distance metrics, such as the Euclidean distance, in that it accounts for correlations between variables, and gives a distance with respect to some statistical distribution. The equation to compute Mahalanobis distance is given in equation 2.

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T \mathbf{S}^{-1} (\vec{x}-\vec{y})} \tag{2}$$

To find an appropriate threshold d, for each experiment, we first compute the pairwise distance between all treatment records and all control records. We then determine the first percentile value as d. A different threshold value is computed for each experiment for each check-in configuration. Once a record has been *matched*, it is then ineligible to be included in other pairs of records. In general, it is not expected that all records from either the treatment or control group will be paired in the process.

Once all pairs of records have been found, the estimated causal effect of the treatment with respect to the control on various health outcomes can be obtained. First, the pair-wise difference in outcomes between the individual from the treatment population and the individual from the control population is computed. In our application, the various health outcomes explored in this study are encoded as binary values - a 1 denoting the outcome occurred, a 0 denoting it did not. Then, the estimated causal effect can be found by computing the average pair-wise distance between all pairs of matched records. In our case, the estimated causal effect can be interpreted as the difference in the *units* of each outcome that would be expected for the same individual if the only thing that changed about them was their BMI classification.

Data Availability

The data analyzed during the current study are available from the corresponding author on reasonable request and with permission of the Kailuan Study.

References

1. Bhaskaran, K., dos Santos-Silva, I., Leon, D. A., Douglas, I. J. & Smeeth, L. Association of bmi with overall and cause-specific mortality: a population-based cohort study of 3· 6 million adults in the uk. *The lancet Diabetes & endocrinology* **6**, 944–953 (2018).

2. Elagizi, A. *et al.* An overview and update on obesity and the obesity paradox in cardiovascular diseases. *Prog. cardiovascular diseases* **61**, 142–150 (2018).
3. Hainer, V. & Aldhoon-Hainerová, I. Obesity paradox does exist. *Diabetes care* **36**, S276–S281 (2013).
4. Horwich, T. B. *et al.* The relationship between obesity and mortality in patients with heart failure. *J. Am. Coll. Cardiol.* **38**, 789–795 (2001).
5. Sourial, N., Longo, C., Vedel, I. & Schuster, T. Daring to draw causal claims from non-randomized studies of primary care interventions. *Fam. Pract.* **35**, 639–643 (2018).
6. Franks, P. W. & Atabaki-Pasdar, N. Causal inference in obesity research. *J. internal medicine* **281**, 222–232 (2017).
7. Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. educational Psychol.* **66**, 688 (1974).
8. Splawa-Neyman, J., Dabrowska, D. M. & Speed, T. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Stat. Sci.* 465–472 (1990).
9. Yao, L. *et al.* A survey on causal inference. *ACM Transactions on Knowl. Discov. from Data (TKDD)* **15**, 1–46 (2021).
10. Genes and obesity (2013).
11. Jacquemont, S. *et al.* Mirror extreme bmi phenotypes associated with gene dosage at the chromosome 16p11. 2 locus. *Nature* **478**, 97–102 (2011).
12. Bank, D., Koenigstein, N. & Giryes, R. Autoencoders. *arXiv preprint arXiv:2003.05991* (2020).
13. Liu, X. *et al.* Cumulative exposure to ideal cardiovascular health and incident diabetes in a chinese population: The kailuan study. *J. Am. Hear. Assoc.* **5**, e004132 (2016).
14. Wang, L. *et al.* A prospective study of waist circumference trajectories and incident cardiovascular disease in china: the kailuan cohort study. *The Am. J. Clin. Nutr.* **113**, 338–347 (2021).
15. Zhang, X. *et al.* Alcohol consumption and risk of cardiovascular disease, cancer and mortality: a prospective cohort study. *Nutr. journal* **20**, 1–10 (2021).

Acknowledgements

Funding was provided in part by a Pennsylvania State University (PSU) Institute for Computational and Data Sciences (ICDS) seed grant.

Author contributions statement

X.Z. cleaned and formatted the data. P.M. conceived of experiment and C.H. conducted the experiments. P.M. and C.H. both analysed the results. All authors reviewed manuscript.