

RESEARCH ARTICLE

Applying Data Envelopment Analysis to Preventive Medicine: A Novel Method for Constructing a Personalized Risk Model of Obesity

Hiroto Narimatsu^{1*}, Yoshinori Nakata², Sho Nakamura³, Hidenori Sato⁴, Ri Sho¹, Katsumi Otani¹, Ryo Kawasaki¹, Isao Kubota⁵, Yoshiyuki Ueno⁶, Takeo Kato⁷, Hidetoshi Yamashita⁸, Akira Fukao¹, Takamasa Kayama⁹

1 Department of Public Health, Yamagata University Graduate School of Medicine, Yamagata, Japan, **2** Department of Healthcare Management, Teikyo University Graduate School of Public Health, Tokyo, Japan, **3** Department of Clinical Oncology, Yamagata University Faculty of Medicine, Yamagata, Japan, **4** Genome Informatics Unit, Institute for Promotion of Medical Science Research, Yamagata University Faculty of Medicine, Yamagata, Japan, **5** First Department of Internal Medicine, Yamagata University Faculty of Medicine, Yamagata, Japan, **6** Second Department of Internal Medicine, Yamagata University Faculty of Medicine, Yamagata, Japan, **7** Third Department of Internal Medicine, Yamagata University Faculty of Medicine, Yamagata, Japan, **8** Department of Ophthalmology and Visual Sciences, Yamagata University Faculty of Medicine, Yamagata, Japan, **9** Department of Neurosurgery, Yamagata University Faculty of Medicine, Yamagata, Japan

* hiroto-narimatsu@umin.org



OPEN ACCESS

Citation: Narimatsu H, Nakata Y, Nakamura S, Sato H, Sho R, Otani K, et al. (2015) Applying Data Envelopment Analysis to Preventive Medicine: A Novel Method for Constructing a Personalized Risk Model of Obesity. PLoS ONE 10(5): e0126443. doi:10.1371/journal.pone.0126443

Academic Editor: Stefan Kiechl, Innsbruck Medical University, AUSTRIA

Received: December 22, 2014

Accepted: April 2, 2015

Published: May 14, 2015

Copyright: © 2015 Narimatsu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by JSPS KAKENHI (Grant-in-Aid for Challenging Exploratory Research) Grant Number 25560363 to Dr. Narimatsu. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Data envelopment analysis (DEA) is a method of operations research that has not yet been applied in the field of obesity research. However, DEA might be used to evaluate individuals' susceptibility to obesity, which could help establish effective risk models for the onset of obesity. Therefore, we conducted this study to evaluate the feasibility of applying DEA to predict obesity, by calculating efficiency scores and evaluating the usefulness of risk models. In this study, we evaluated data from the Takahata study, which was a population-based cohort study (with a follow-up study) of Japanese people who are >40 years old. For our analysis, we used the input-oriented Charnes-Cooper-Rhodes model of DEA, and defined the decision-making units (DMUs) as individual subjects. The inputs were defined as (1) exercise (measured as calories expended) and (2) the inverse of food intake (measured as calories ingested). The output was defined as the inverse of body mass index (BMI). Using the β coefficients for the participants' single nucleotide polymorphisms, we then calculated their genetic predisposition score (GPS). Both efficiency scores and GPS were available for 1,620 participants from the baseline survey, and for 708 participants from the follow-up survey. To compare the strengths of the associations, we used models of multiple linear regressions. To evaluate the effects of genetic factors and efficiency score on body mass index (BMI), we used multiple linear regression analysis, with BMI as the dependent variable, GPS and efficiency scores as the explanatory variables, and several demographic controls, including age and sex. Our results indicated that all factors were statistically significant ($p < 0.05$), with an adjusted R^2 value of 0.66. Therefore, it

is possible to use DEA to predict environmentally driven obesity, and thus to establish a well-fitted model for risk of obesity.

Introduction

Personalized medicine considers the patient's genetic characteristics and history of exposure to environmental factors, and is expected to comprise the next generation of therapy. In this context, personalized medicine has the potential to provide preventive and curative interventions that are personalized for each individual, and is anticipated to reduce the incidence of disease. Furthermore, the completion of the Human Genome Project has paved the way for significant developments in the field of personalized medicine [1]. However, genomic information alone is not sufficient for providing personalized medicine, because the development of non-communicable diseases is a multifactorial process that also involves environmental influences, lifestyle choices, and genetic influences on the individual's susceptibility. Therefore, population-based prospective cohort studies that incorporate genomic measurements have recently been recommended for determining the risk factors and etiologies of chronic diseases, while also accounting for gene-environment interactions [2–4]. However, these studies are expensive to conduct, and typically require 20–30 years to gather the relevant information [2–4]. Therefore, these challenges are major obstacles in establishing personalized preventative medicine, as the development of predictive models for disease susceptibility are essential to establishing personalized preventative treatments, and these models must incorporate both genetic and environmental factors. Historically, multivariate regression models have commonly been used to identify disease risk factors, although more flexible methods (such as machine learning) are currently used to predict the onset of disease [5].

In this context, data envelopment analysis (DEA) is a measure of efficiency that considers multiple inputs and outputs, and can be used to evaluate outputs while controlling for the inputs [6–8]. The DEA process calculates efficiency scores (a measure of relative efficiency) within a given sample of decision making units (DMUs) [9,10]. A completely efficient subgroup will provide an efficiency score of 1, while a DMU that achieves a score of < 1 is considered inefficient, and these scores can be used to calculate an individual's disease susceptibility. For example, if a person who has a high body mass index (BMI), consumes too much food, and never exercises, that person may be obese due to his/her environment. In contrast, if a person with a high BMI eats appropriately and exercises regularly, that person's obesity may be hereditary, which is independent of his/her environment. Therefore, by calculating and comparing individuals' efficiency scores via DEA, we can evaluate individuals' susceptibility to obesity. In that model, the inputs would be the individual's caloric intake and energy expenditure, and the efficiency score risk model would be highly fitted for the onset of obesity, due to the use of BMI as the objective variable (output).

In the field of healthcare, DEA has already been adopted in hospital management [10,11], and one recent study has used DEA to evaluate the efficiency of physical activity programs for elderly women [12]. In that study, the inputs were defined as the amount of time spent performing strength, flexibility, and aerobic exercises, and the outputs were defined as the levels of strength, flexibility, static balance, dynamic balance, and maximal oxygen consumption that were verified at the end of the physical activity program. The authors found that DEA facilitated the assessment of the exercise program using the time spent performing physical activity, and subsequently concluded that DEA has promising applications in

preventing fitness-related conditions. However, that study's findings have limited application in clinical or preventive practice.

Therefore, in the present study, we evaluated DEA in the field of preventive medicine, and investigated its feasibility using clinical and genetic data from a large prospective cohort of a genome-wide association study (GWAS). Using this information, we hoped to estimate personal susceptibility to obesity, using the efficiency scores for the DMUs (individual subjects), and to evaluate the feasibility of applying DEA in the field of preventative medicine.

Methods

DEA analysis

In this study, we employed the input-oriented Charnes-Cooper-Rhodes model of DEA, which is a constant returns-to-scale model that is particularly relevant, given its ability to include multiple inputs and outputs without requiring an a priori function specification [13]. In this context, a DMU is defined as the entity that is responsible for converting the inputs into outputs [14]; therefore, we defined the DMUs as individual subjects. The inputs were defined as (1) total physical expenditure (measured as calories expended) and (2) the inverse of food intake (measured as calories ingested). In DEA, efficiency is defined as high if the input is minimized when the outputs are held constant, or if the output is maximized when the inputs are held constant. Therefore, to fit this definition, we used the inverse of BMI and of caloric intake as the outputs for our models, as subjects can only directly control the inputs (e.g., exercise and food intake), and cannot directly control the output (BMI) [15].

Each subject's efficiency score was calculated using DEA-Solver-Pro Software (Saitech, Inc., Tokyo, Japan) [6]. In that model, the efficiency scores range between 0 and 1, and the individuals who are most efficient at burning calories are assigned an efficiency score of 1 [16].

Genotyping and imputation

Whole blood samples were used to obtain the subjects' DNA samples, and the single nucleotide polymorphisms (SNPs) were genotyped using an Infinium 660W BeadChip assay (Illumina, San Diego, CA, USA). For our analyses, SNPs with a minor allele frequency of $<0.5\%$ and a call rate of $<95\%$ were excluded. After confirming the quality of the genotyped SNP data, we performed genotype imputation using the MACH-Admix program [17], using a data set from the 1000 Genomes Project (194 ASN, 68 CHB, 25 CHS, 84 JPT, and 17 MXL, released in August 2010).

Genetic predisposition score

The genetic predisposition score (GPS) was calculated using the β coefficients of the SNPs, via a previously reported weighting method [18]. In this method, the GPS is calculated by multiplying the number of effect alleles (0, 1, or 2) at each locus by the β coefficient of that SNP (as obtained from the GWAS), dividing by the maximum allowable sum of the β coefficients, and then multiplying by twice the number of alleles. Higher scores indicate a greater genetic predisposition to obesity. The GPSs from the β coefficients of 29 SNPs from East Asian subjects were used to conduct these analyses [19,20].

Statistical analysis

The statistical analyses for this study were conducted during September 2014. The statistical analyses were performed using R software (version 3.1.1, R Foundation for Statistical Computing, Vienna, Austria), and differences with a p-value of <0.05 were considered statistically significant. Pearson's correlation coefficients were calculated to evaluate the relationships between the

efficiency scores, GPS, and BMI. Robust multiple linear regression analyses were also conducted to evaluate the effect of obesity risk factors, using the “robustbase” function in the R software. Five multiple linear regression models were used to evaluate the coefficients for BMI at baseline or yearly changes in BMI (the dependent variables). Model 1 evaluated the effect of environmental factors (energy expenditure and energy intake) as the explanatory variables, while Model 2 evaluated the effect of the efficiency score. Model 3 evaluated the effect of genetic factors using the GPS, Model 4 evaluated the effect of genetic and environmental factors (energy expenditure, energy intake, and GPS), and Model 5 used the GPS and efficiency scores to compare the effect of genetic factors to that of the efficiency score. All models were adjusted for age and sex.

Data from the cohort study

The Takahata study is a population-based cohort study (with a subsequent follow-up study) of Japanese people who were >40 years old, which sought to clarify the risk factors for certain life-style-related conditions, such as diabetes and obesity [21–24]. The details of the study design have been reported previously [21–24], and the participants’ physical activity status was calculated as metabolic equivalents, using The Japan Arteriosclerosis Longitudinal Study Physical Activity Questionnaire [25]. In addition, the daily intake of each nutrient was calculated using a brief self-administered diet history questionnaire [26]. The baseline survey of 3,522 participants was conducted from 2004 to 2006, and 1,620 unique DNA samples were extracted from the participants’ whole blood samples, and we were able to calculate the efficiency scores for those 1,620 participants. Among these participants, 1,079 completed the follow-up survey in 2011, at 5–7 years after the baseline survey, and we classified these participants as underweight (<18.5 kg/m²), normal (18.5–24.99 kg/m²), and overweight (≥25 kg/m²) [27]. Written informed consent was obtained from all participants for the original study. The protocols of both the Takahata study and the present study were approved by the Yamagata University Faculty of Medicine ethics committee.

Results

Efficiency score and baseline BMI

Among the participants in the Takahata study, both GPS and efficiency scores were available for 1,620 participants (726 men and 894 women). The characteristics of these subjects are shown in [Table 1](#). The median age at the baseline survey was 62 years (range, 40–84 years), and the mean baseline BMI was 23.4 kg/m² (standard deviation [SD], 3.1). The mean efficiency score of the participants was 0.51 (SD, 0.13), and the total physical expenditure and inverse of food intake according to the efficiency score quartiles are shown in [Fig 1](#).

The baseline BMI and efficiency scores were significantly correlated ($r = -0.78$, $p < 0.01$) ([Fig 2a](#)), as were the baseline BMI and GPS ($r = 0.14$, $p < 0.01$) ([Fig 2b](#)) and the efficiency scores and GPS ($r = -0.12$, $p < 0.01$) ([Fig 2c](#)) ([Table 2](#)). The results of the subgroup analysis according to age are also shown in [Table 2](#). We then evaluated the coefficients using the five multiple linear regression models, and found that the models that included efficiency score had greater adjusted R² values than those that did not include efficiency score. In Model 5, which used the GPS and efficiency scores to compare the effect of genetic factors with that of efficiency score, the equation for yearly changes in BMI was:

$$\begin{aligned} \text{Baseline BMI (kg/m}^2\text{)} = & 28.84 + \{0.25 \times \text{Sex (women vs. men)}\} + \{0.05 \times \text{Age} \\ & \text{(years)}\} + (0.03 \times \text{GPS}) + (-19.47 \times \text{efficiency score}) \end{aligned} \quad (1)$$

In this formula, all factors were statistically significant ($p < 0.05$) ([Table 3](#)).

Table 1. Participant characteristics (n = 1,620).

Variable	Number
Age, years (median [range], mean [SD])	62 (40–84), 61.3 (10.1)
Sex (men/women)	726/894
Baseline BMI, kg/m ² (median [range], mean [SD])	23.2 (15.0–35.5), 23.4 (3.1)
Efficiency score (median [range], mean [SD])	0.49 (0.23–1.00), 0.51 (0.13)
Efficiency score according to age group	
40–49 years	0.49 (0.26–0.91), 0.50 (0.12)
50–59 years	0.47 (0.26–1.00), 0.49 (0.13)
60–69 years	0.49 (0.24–0.99), 0.50 (0.12)
≥70 years	0.54 (0.23–1), 0.55 (0.14)
GPS (median [range], mean [SD])	25.9 (14.5–42.3), 26.1 (3.1)
Total physical expenditure, METs-h/day (median [range], mean [SD])	35.3 (25.8–74.8), 36.2 (5.8)
Food intake, kcal/day (median [range], mean [SD])	2,175 (307–7090), 2,257 (673)
Change in BMI, kg/m ² /year (median [range], mean [SD])	-0.01 (-0.92–1.14), -0.02 (0.22)

SD: standard deviation; BMI, body mass index; GPS, genomic predisposition score; METs, metabolic equivalents.

doi:10.1371/journal.pone.0126443.t001

Efficiency score and yearly change in BMI

After excluding underweight and overweight participants, data regarding yearly changes in BMI were available for 708 of the 1,079 subjects who participated in the follow-up survey. The yearly change in BMI was correlated with the efficiency score ($r = 0.038$, $p = 0.31$) and the

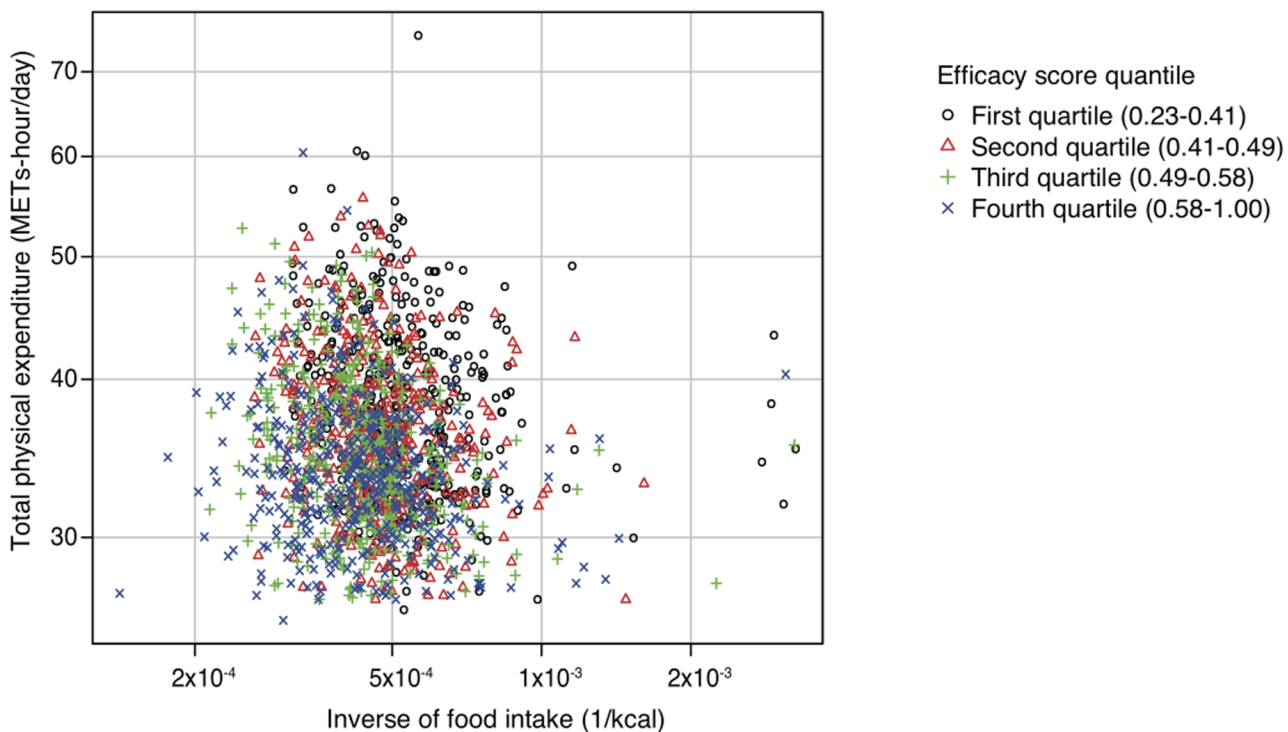


Fig 1. Total physical expenditure and the inverse of food intake according to the efficiency score quartiles. METs, metabolic equivalents.

doi:10.1371/journal.pone.0126443.g001

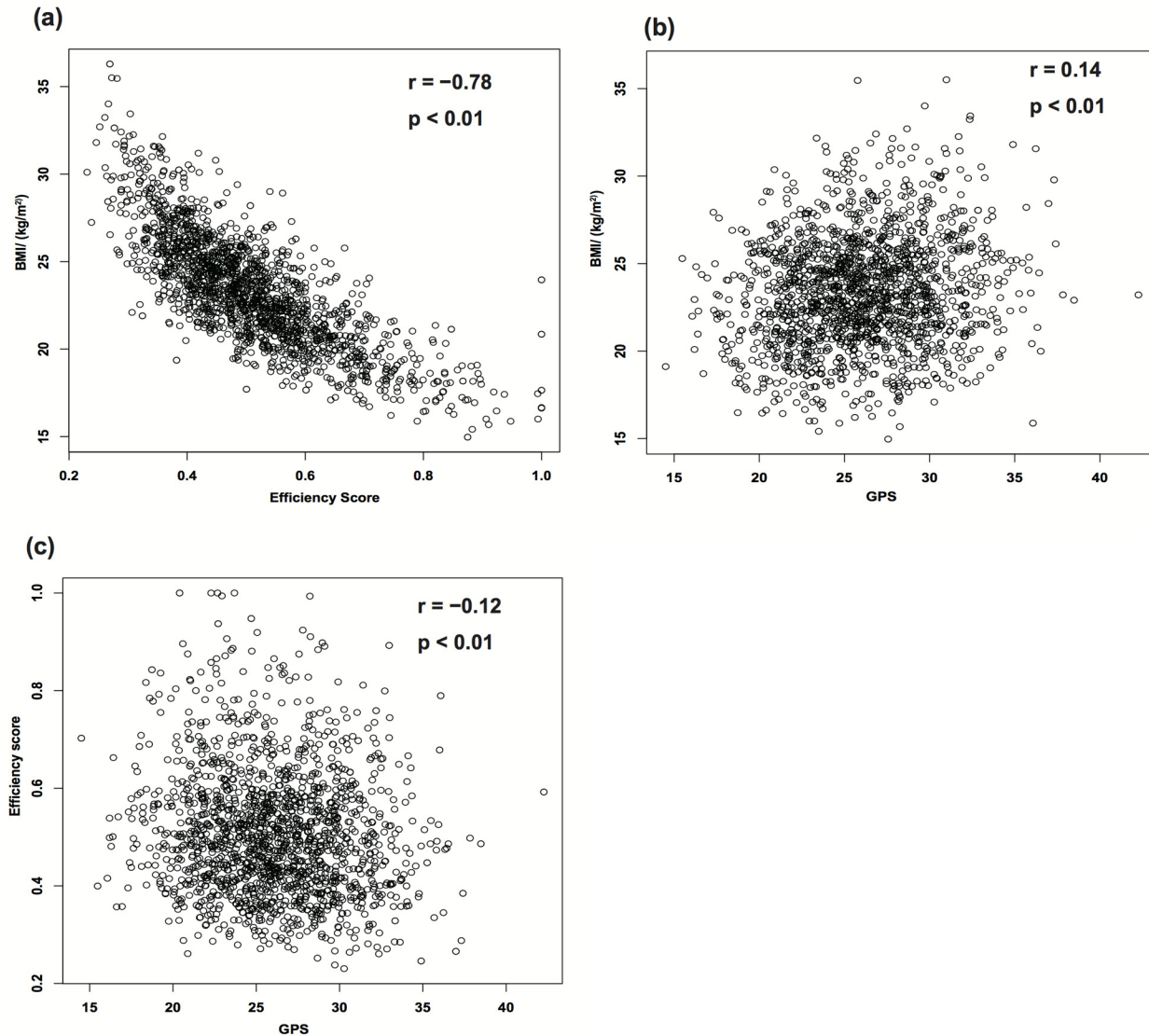


Fig 2. Correlations between clinical variables, genomic predisposition score (GPS), and efficiency score. Correlations are shown for (a) baseline body mass index (BMI) and efficiency score ($r = -0.78$, $p < 0.01$), (b) baseline BMI and GPS ($r = 0.14$, $p < 0.01$), and (c) efficiency score and GPS ($r = -0.12$, $p < 0.01$).

doi:10.1371/journal.pone.0126443.g002

yearly change in BMI and the GPS were also correlated ($r = 0.048$, $p = 0.12$) (Fig 3). Furthermore, we evaluated the coefficients for the change in BMI using the five multiple linear regression models, although the R^2 values for these models were very low (< 0.04). In Model 5 (Table 3), we used the GPS and efficiency scores to compare the effect of genetic factors with that of efficiency score, which provided the following equation:

$$\begin{aligned} \text{Changes in BMI (kg/m}^2\text{/year)} = & 0.19 + \{-0.02 \times \text{Sex(women vs. men)}\} + \\ & \{-0.004 \times \text{Age(years)}\} + (0.0001 \times \text{GPS}) + (0.14 \times \text{efficientscore}) \end{aligned} \quad (2)$$

Table 2. Factors associated with baseline body mass index (n = 1,620).

	Partial coefficient	Standard error†	p-value	
Model 1 (adjusted R ² = 0.0083)				
Intercept	22.19	0.92	N/A	
Sex (women vs. men)	-0.42	0.17	0.01	
Age (years)	0.02	0.01	0.03	
Energy expenditure (METs-h/day)	0.00	0.01	0.88	
Energy intake (kcal/day)	0.0001	0.0001	0.33	
Model 2 (adjusted R ² = 0.65)				
Intercept	29.76	0.35	N/A	
Sex (women vs. men)	0.24	0.10	0.01	
Age (years)	0.05	0.0046	<0.01	
Efficiency score	-19.57	0.41	<0.01	
Model 3 (adjusted R ² = 0.025)				
Intercept	19.69	0.75	N/A	
Sex (women vs. men)	-0.42	0.16	0.01	
Age (years)	0.02	0.01	0.02	
GPS	0.10	0.02	<0.01	
Model 4 (adjusted R ² = 0.024)				
Intercept	19.39	1.07	N/A	
Sex (women vs. men)	-0.38	0.16	0.02	
Age (years)	0.02	0.01	0.03	
Energy expenditure (METs-h/day)	0.0001	0.01	0.99	
Energy intake (kcal/day)	0.0001	0.0001	0.31	
GPS	0.10	0.02	<0.01	
Model 5 (adjusted R ² = 0.66)				
Intercept	28.84	0.48	N/A	
Sex (women vs. men)	0.25	0.10	0.01	
Age (years)	0.05	0.0046	<0.01	
GPS	0.03	0.01	0.01	
Efficiency score	-19.47	0.41	<0.01	
Age <60 years (n = 672) (adjusted R ² = 0.65)	Intercept	30.80	0.94	N/A
	Sex (women vs. men)	-0.02	0.16	0.92
	Age (years)	0.02	0.01	0.09
	GPS	0.03	0.02	0.08
	Efficiency score	-19.78	0.67	<0.01
Age ≥60 years (n = 948) (adjusted R ² = 0.66)	Intercept	27.84	0.83	N/A
	Sex (women vs. men)	0.44	0.12	<0.01
	Age (years)	0.07	0.01	<0.01
	GPS	0.04	0.02	0.02
	Efficiency score	-19.39	0.53	<0.01

N/A, not applicable; GPS, genomic predisposition score; METs, metabolic equivalents.

†Robust standard errors are reported.

doi:10.1371/journal.pone.0126443.t002

Discussion

In this study, we used DEA to estimate individuals' susceptibility to obesity. First, we compared a regression model that included the GPS with another model that included the efficiency score. Second, we compared the resulting coefficients with the coefficient provided by the

Table 3. Factors associated with yearly change in body mass index (n = 708).

	Partial coefficient	Standard error†	p-value
Model 1 (adjusted R ² = 0.032)			
Intercept	0.22	0.09	N/A
Sex (women vs. men)	-0.02	0.02	0.29
Age (years)	-0.004	0.001	<0.01
Energy expenditure (METs-h/day)	0.0004	0.001	0.79
Energy intake (kcal/day)	0.000001	0.00001	0.94
Model 2 (adjusted R ² = 0.038)			
Intercept	0.19	0.07	N/A
Sex (women vs. men)	-0.02	0.01	0.16
Age (years)	-0.0043	0.0009	<0.01
Efficiency score	0.14	0.07	0.06
Model 3 (adjusted R ² = 0.033)			
Intercept	0.25	0.07	N/A
Sex (women vs. men)	-0.02	0.01	0.25
Age (years)	-0.0041	0.0009	<0.01
GPS	-0.000045	0.0019	0.98
Model 4 (adjusted R ² = 0.031)			
Intercept	0.23	0.11	N/A
Sex (women vs. men)	-0.02	0.02	0.29
Age (years)	0.00	0.00	<0.01
Energy expenditure (METs-h/day)	0.0004	0.0015	0.79
Energy intake (kcal/day)	0.000001	0.000013	0.94
GPS	-0.00002	0.0019	0.99
Model 5 (adjusted R ² = 0.036)			
Intercept	0.19	0.08	N/A
Sex (women vs. men)	-0.02	0.02	0.16
Age (years)	-0.0043	0.0009	<0.01
GPS	0.0001	0.0019	0.97
Efficiency score	0.14	0.07	0.06

N/A, not applicable; GPS, genomic predisposition score; METs, metabolic equivalents.

†Robust standard errors are reported.

doi:10.1371/journal.pone.0126443.t003

model that included both variables. Based on our results, the efficiency score had a better predictive power when using baseline BMI, compared to GPS, as evidenced by the greater adjusted R² value and greater partial efficiency value. These findings suggest that DEA is feasible for establishing a model of an individual's risk of obesity. Similar to the previous study that used DEA to evaluate a physical activity program [12], our data indicate that we can use DEA to predict the efficiency of interventions that are designed to prevent obesity. Therefore, the results of this study highlight the potential applications of DEA in personalized preventative medicine.

Obesity is caused by various factors [19], including environmental (diet and physical activity), genetic, and epigenetic factors [19,28–33]. These genetic factors include common variants (e.g., SNPs), rare variants, and gene-environment interactions [3,34]. In this context, GPS is currently the optimal risk score for evaluating genetic factors, although it can only be measured using a large-scale GWAS, and can only elucidate a subset of the genetic factors. In theory, DEA can calculate an efficiency score that includes all factors, besides the input factors (e.g., food intake and energy expenditure in this study). This theory is strongly supported by our

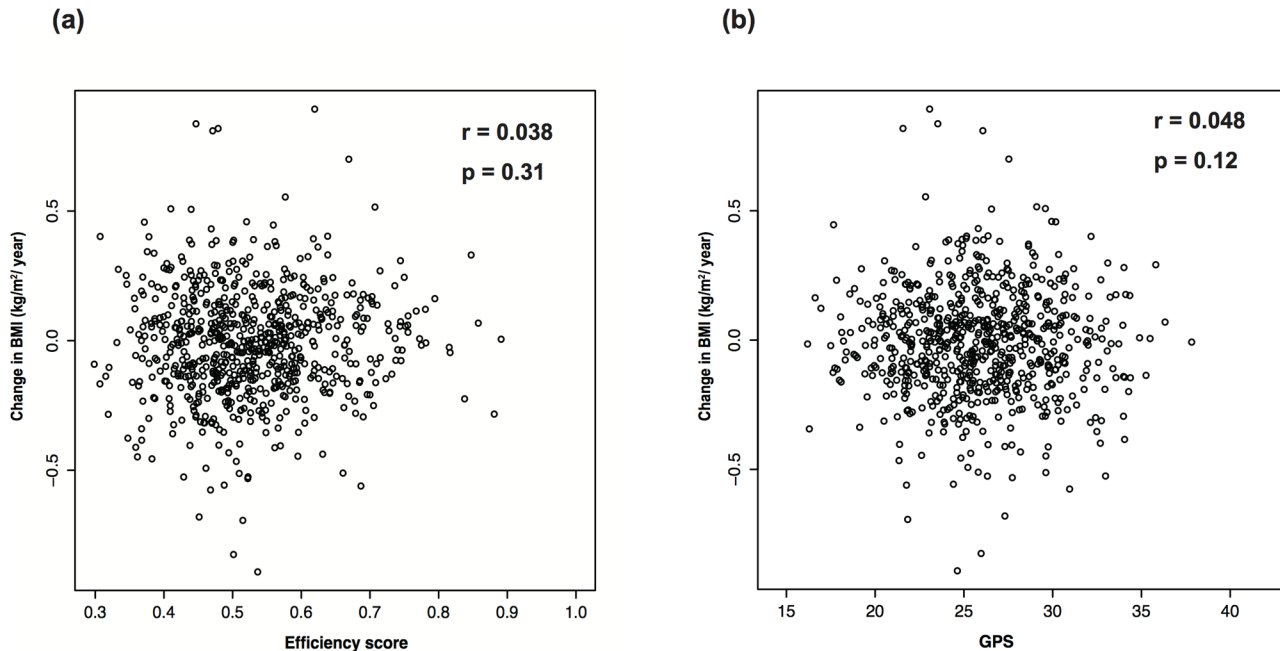


Fig 3. Correlations between changes in body mass index (BMI), genomic predisposition score (GPS), and efficiency score. Correlations are shown for (a) yearly change in BMI and efficiency score ($r = -0.012$, $p = 0.80$) and (b) yearly change in BMI and GPS ($r = 0.055$, $p = 0.21$).

doi:10.1371/journal.pone.0126443.g003

finding that the determination coefficient for the model that included the efficiency score was higher than that for the model that included the GPS. Thus, the efficiency score from DEA is a useful measure for the risk of obesity, given that most genetic factors for the onset of obesity remain unknown. Furthermore, we calculated these efficiency scores using existing clinical data, which is much more cost-effective than personal genomic analyses. Therefore, our methodology can be used as a foundational approach to establishing personalized preventive medicine.

However, there is a major limitation that should be considered when interpreting our findings. For the change in BMI (between the baseline and follow-up surveys), we failed to establish efficiently fitted models, due to the low R^2 values and low partial efficiencies. This finding may be partially explained by the small value of the absolute change in BMI, which may have been affected by several factors. First, the baseline survey of the Takahata study was conducted when the participants received health check-ups. However, based on the results of that health check-up, the obese participants received preventive medical interventions, such as nutritional guidance. Therefore, it is possible that participants with low efficiency scores had received an intervention, which might have affected their change in BMI. Second, the biological significance of BMI differs according to age, as larger BMI increases are observed in the adolescent period, compared to those observed in adulthood, and BMI peaks at the age of 55 years among men and at the age of 60 years among women [30–33]. However, in the Takahata Study, more than half of the participants were >60 years old, and would likely experience a decreasing trend in their BMI. Thus, the role of the efficiency score should be evaluated among all age groups in future studies, including the adolescent period, as this additional data would facilitate more appropriate use of the efficiency score in preventive medicine. Third, the Charnes-Cooper-Rhodes model is one of many possible models that we could have selected for the present study. However, the present study is the first to examine the application of DEA in predicting obesity, and while model selection is an important consideration in DEA, we were unable to

rely on existing studies to justify the model selection. Therefore, we do not reject the possibility that better models may exist for this purpose [15]. Furthermore, we eagerly await future studies in this field, as the better combination of DEA and genetic information would help establish more effective risk models for obesity.

In conclusion, we estimated individuals' susceptibility to obesity using DEA. Although the results of the present study are preliminary, we are planning large-scale prospective studies to confirm the feasibility and usefulness of DEA in this field. Nevertheless, our findings provide novel foundational insights regarding methods to facilitate personalized preventive medicine.

Acknowledgments

We thank Professor Chifumi Kitanaka (Yamagata University) for providing useful discussion. We also thank Editage (www.editage.com) for English language editing.

Author Contributions

Conceived and designed the experiments: HN YN. Performed the experiments: HN YN SN HS. Analyzed the data: HN YN SN HS. Contributed reagents/materials/analysis tools: HN YN RS KO RK IK YU T. Kato HT HY AF T. Kayama. Wrote the paper: HN YN HS. Contributed in data collection: HN RS KO RK IK YU T. Kato HT HY AF T. Kayama. Critical review and approval of the final manuscript: HN YN SN HS RS KO RK IK YU T. Kato HY AF T. Kayama.

References

1. Willett WC (2002) Balancing life-style and genomics research for disease prevention. *Science* 296: 695–698. PMID: [11976443](https://pubmed.ncbi.nlm.nih.gov/11976443/)
2. NHGRI Expert Panel Recommendations for a populationbased cohort. Available: <http://www.genome.gov/Pages/About/OD/ReportsPublications/PotentialUSCohort.pdf>. Accessed Sep 16 2010.
3. Manolio TA, Bailey-Wilson JE, Collins FS (2006) Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 7: 812–820. PMID: [16983377](https://pubmed.ncbi.nlm.nih.gov/16983377/)
4. Yamagata University Genomic Cohort C, Narimatsu H (2013) Constructing a contemporary gene-environmental cohort: study design of the Yamagata Molecular Epidemiological Cohort Study. *J Hum Genet* 58: 54–56. doi: [10.1038/jhg.2012.128](https://doi.org/10.1038/jhg.2012.128) PMID: [23171998](https://pubmed.ncbi.nlm.nih.gov/23171998/)
5. Yoo C, Ramirez L, Liuzzi J (2014) Big data analysis using modern statistical and machine learning methods in medicine. *Int Neurolog J* 18: 50–57. doi: [10.5213/inj.2014.18.2.50](https://doi.org/10.5213/inj.2014.18.2.50) PMID: [24987556](https://pubmed.ncbi.nlm.nih.gov/24987556/)
6. Cooper WW, Seiford LM, Tone K (2007) Basic CCR model. In: Cooper WW, Seiford LM, Tone K, editors. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2nd ed. New York: Springer. pp. 21–40.
7. Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) Data envelopment analysis. In: Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE, editors. *An Introduction to Efficiency and Productivity Analysis*. New York: Springer. pp. 161–182.
8. Bogetoft P, Otto L (2011) Data envelopment analysis DEA. In: Bogetoft P, Otto L, editors. *Benchmarking with DEA, SFA, and R*. New York: Springer. pp. 81–114.
9. Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2: 429–444.
10. Banker RD, Conrad RF, Strauss RP (1986) A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production. *Management Science* 32: 30–44.
11. Sherman HD (1984) Hospital efficiency measurement and evaluation. Empirical test of a new technique. *Med Care* 22: 922–938. PMID: [6436590](https://pubmed.ncbi.nlm.nih.gov/6436590/)
12. Alburquerque-Sendin F, Barberio-Mariano E, Brandao-Santana N, Rebelatto DA, Rebelatto JR (2012) Effects of an adapted physical activity program on the physical condition of elderly women: an analysis of efficiency. *Rev Bras Fisioter* 16: 328–336. PMID: [22699693](https://pubmed.ncbi.nlm.nih.gov/22699693/)
13. Farrell MJ (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A (General)* 120: 253–281.

14. Thanassoulis E, Portela MCS, Despic O (2008) Data envelopment analysis: the mathematical programming approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS, editors. *The measurement of productive efficiency and productivity growth*. New York: Oxford University Press. pp. 251–420.
15. Cooper WW, Seiford LM, Tone K (2007) In: Cooper WW, Seiford LM, Tone K, editors. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2nd ed. New York: Springer. pp. 114–115.
16. Hollingsworth B (2008) The measurement of efficiency and productivity of health care delivery. *Health Econ* 17: 1107–1128. doi: [10.1002/hec.1391](https://doi.org/10.1002/hec.1391) PMID: [18702091](https://pubmed.ncbi.nlm.nih.gov/18702091/)
17. Liu EY, Li MY, Wang W, Li Y (2013) MaCH-Admix: Genotype Imputation for Admixed Populations. *Genetic Epidemiology* 37: 25–37. doi: [10.1002/gepi.21690](https://doi.org/10.1002/gepi.21690) PMID: [23074066](https://pubmed.ncbi.nlm.nih.gov/23074066/)
18. Qi Q, Chu AY, Kang JH, Jensen MK, Curhan GC, et al. (2012) Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med* 367: 1387–1396. doi: [10.1056/NEJMoa1203039](https://doi.org/10.1056/NEJMoa1203039) PMID: [22998338](https://pubmed.ncbi.nlm.nih.gov/22998338/)
19. Lu Y, Loos RJ (2013) Obesity genomics: assessing the transferability of susceptibility loci across diverse populations. *Genome Med* 5: 55. doi: [10.1186/gm459](https://doi.org/10.1186/gm459) PMID: [23806069](https://pubmed.ncbi.nlm.nih.gov/23806069/)
20. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948. doi: [10.1038/ng.686](https://doi.org/10.1038/ng.686) PMID: [20935630](https://pubmed.ncbi.nlm.nih.gov/20935630/)
21. Daimon M, Sato H, Oizumi T, Toriyama S, Saito T, et al. (2008) Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population. *Biochemical and Biophysical Research Communications* 365: 466–471. PMID: [17991425](https://pubmed.ncbi.nlm.nih.gov/17991425/)
22. Konta T, Hao Z, Abiko H, Ishikawa M, Takahashi T, et al. (2006) Prevalence and risk factor analysis of microalbuminuria in Japanese general population: the Takahata study. *Kidney International* 70: 751–756. PMID: [16807548](https://pubmed.ncbi.nlm.nih.gov/16807548/)
23. Karasawa S, Daimon M, Sasaki S, Toriyama S, Oizumi T, et al. (2010) Association of the common fat mass and obesity associated (FTO) gene polymorphism with obesity in a Japanese population. *Endocr J* 57: 293–301. PMID: [20051647](https://pubmed.ncbi.nlm.nih.gov/20051647/)
24. Kohno K, Narimatsu H, Shiono Y, Suzuki I, Kato Y, et al. (2014) Management of erythropoiesis: cross-sectional study of the relationships between erythropoiesis and nutrition, physical features, and adiponectin in 3519 Japanese people. *Eur J Haematol* 92: 298–307. doi: [10.1111/ejh.12250](https://doi.org/10.1111/ejh.12250) PMID: [24329589](https://pubmed.ncbi.nlm.nih.gov/24329589/)
25. Miyabara Y, Onoe Y, Harada A, Kuroda T, Sasaki S, et al. (2007) Effect of physical activity and nutrition on bone mineral density in young Japanese women. *J Bone Miner Metab* 25: 414–418. PMID: [17968494](https://pubmed.ncbi.nlm.nih.gov/17968494/)
26. Sasaki S, Yanagibori R, Amano K (1998) Self-administered diet history questionnaire developed for health education: a relative validation of the test-version by comparison with 3-day diet record in women. *J Epidemiol* 8: 203–215. PMID: [9816812](https://pubmed.ncbi.nlm.nih.gov/9816812/)
27. (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet* 363: 157–163. PMID: [14726171](https://pubmed.ncbi.nlm.nih.gov/14726171/)
28. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, et al. (2014) DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 383: 1990–1998. doi: [10.1016/S0140-6736\(13\)62674-4](https://doi.org/10.1016/S0140-6736(13)62674-4) PMID: [24630777](https://pubmed.ncbi.nlm.nih.gov/24630777/)
29. Booth KM, Pinkston MM, Poston WS (2005) Obesity and the built environment. *J Am Diet Assoc* 105: S110–117. PMID: [15867906](https://pubmed.ncbi.nlm.nih.gov/15867906/)
30. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, et al. (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 384: 766–781. doi: [10.1016/S0140-6736\(14\)60460-8](https://doi.org/10.1016/S0140-6736(14)60460-8) PMID: [24880830](https://pubmed.ncbi.nlm.nih.gov/24880830/)
31. Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, et al. (2013) Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum Mol Genet* 22: 3597–3607. doi: [10.1093/hmg/ddt205](https://doi.org/10.1093/hmg/ddt205) PMID: [23669352](https://pubmed.ncbi.nlm.nih.gov/23669352/)
32. Graff M, North KE, Mohlke KL, Lange LA, Luo J, et al. (2012) Estimation of genetic effects on BMI during adolescence in an ethnically diverse cohort: The National Longitudinal Study of Adolescent Health. *Nutr Diabetes* 2: e47. doi: [10.1038/nutd.2012.20](https://doi.org/10.1038/nutd.2012.20) PMID: [23168566](https://pubmed.ncbi.nlm.nih.gov/23168566/)
33. Funatogawa I, Funatogawa T, Nakao M, Karita K, Yano E (2009) Changes in body mass index by birth cohort in Japanese adults: results from the National Nutrition Survey of Japan 1956–2005. *Int J Epidemiol* 38: 83–92. doi: [10.1093/ije/dyn182](https://doi.org/10.1093/ije/dyn182) PMID: [18782894](https://pubmed.ncbi.nlm.nih.gov/18782894/)
34. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–753. doi: [10.1038/nature08494](https://doi.org/10.1038/nature08494) PMID: [19812666](https://pubmed.ncbi.nlm.nih.gov/19812666/)