International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

A META-REGRESSION AND BAYESIAN REGRESSION FRAMEWORK FOR COMBINING RESULTS OF SCIENTIFIC RESEARCH AND SURVEYS OF PEOPLE'S LIFESTYLES TO MAKE RECOMMENDATIONS ON WHAT INTERVENTIONS WILL HELP THEM LIVE LONGER AND HEALTHIER

Ting Lu, Yawen Yuan, Jai Agarwal, Taruna Agarwal, Hrithik Jain, John Leddo, Saanvi Lamba, Rohan Penmetsa, Helena Gabrial, Rohan Matta, AmulyaGottipati, Deepika Ravi, SathvikRedrouthu, Dillon Michlena, Dev Doshi, Deeya Sharma, Mitra Manikandan, Aneesh Sreedhara, SmaranPasupuleti, Riya Pasupuleti, Vihaan Cherukuri, Sambhav Jain, Manav Sabharwal, Eshwar Dokku, Arin Rahman, Charu Mehta, Sumit Kamath, Nevin Philip, SrikarVaishnapu, SathvikAppana, Nikhil Rao, Siddharth Vijay,YajurviPidugu, Chetan Borra, Yash Karayi, Arnav Majeti, Tanya Singhal, Kunal Singhal, Arjun Erasani, Ranveer Kataria, Aleena Ahmad, Mason Elkas, Pooja Somayajula, Aditya Devireddy, Collin O'Brien, SrikaranYelimati, Pranaya Bhatt, AntoniaGabrial, Riya Srikumar,HasiniJasthi, Sophia Nasibdar, Dylan Sood, Harman Sabharwal, EshanIyer, Saharsh Ranga, IshwaryaRamineni, Siri Yarlagadda

John Leddo is director of research at MyEdMaster

DOI: 10.46609/IJSSER.2023.v08i03.013 URL: https://doi.org/10.46609/IJSSER.2023.v08i03.013

Received: 2 March. 2022 / Accepted: 20 March. 2023 / Published: 30 March 2023

ABSTRACT

There has been increasing research into interventions that people can do in areas of diet, exercise, nutritional supplements, sleep, and stress management to improve their health and even slow down or reverse their aging process. Challenges that people face applying these research findings to their daily lives include the fact that this research is evolving and not yet "settled science," most research looks at single interventions in order to maintain experimental control (whereas people are interested in maximizing their benefits by adopting multiple interventions), and different interventions or combinations thereof may be differentially effective for different people. The present paper presents a methodology that uses meta-regression and Bayesian regression to combine results from the scientific literature and surveys of people's lifestyles into an overall predictive model that can make recommendations to people based on their individual characteristics and lifestyles as to what interventions are likely to produces the largest gains in wellness and lifespan extension.

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

Introduction

One of the characteristics that the medical conditions, such as acne or jaundice, described in our previous work (Boina et al. 2021a, b, c, d) had in common is that the knowledge about these conditions was fairly stable, and that there were agreed upon criteria for diagnosing them.

However, this is not true for all medical topics. For example, while professionals have studied the effects of aging on people and the effects of factors like diet and exercise on health for years, recently the two fields have converged and researchers are now looking at whether lifestyle choices regarding such things as diet and exercise can actually slow down or even reverse the aging process.

As with any new field of inquiry, such investigation still involves tracking a moving target. The optimal diet to slow down aging has not been fully articulated and it may, in fact, depend on the particulars of the person eating that diet (e.g., age, gender, lifestyle, presence of co-morbidities).

In the area of diet, research has centered on three main categories: what to eat, when to eat and how much to eat.

For example, when it comes to what to eat, many diets have become popular. One example of this is the ketogenic diet, a diet high in fat and protein, but low in carbohydrates. A study with mice on a ketogenic diet has shown that this diet extended their longevity. Motor functions, memory, and muscle mass were preserved in ketogenic mice. Also, protein acetylation was increased in the liver and skeletal muscle of the ketogenic mice (Roberts et al., 2017).

Multiple studies have also been done on the Mediterranean diet, a diet low in protein and higher in carbohydrates and fats (predominantly, olive oil) and have concluded that a diet that adheres to the principles of the traditional Mediterranean one is associated with longer lifespans (Trichopulou and Vasilopoulou, 2000).

When it comes to when to eat, the concept of fasting has received great currency in the literature. Intermittent and periodic fasting (IF and PF, respectively) are emerging as safe strategies to affect longevity and healthspan by acting on cellular aging and disease risk factors while causing no or minor side effects. IF lasting from 12 to 48 hours and repeated every 1 to 7 days and PF lasting 2 to 7 days and repeated once per month or less have the potential to prevent and treat disease, but their effect on cellular aging and the molecular mechanisms involved are only beginning to be unraveled. There are also the therapeutic potential and side effects of IF and PF with a focus on cancer, autoimmunity, neurodegeneration, and metabolic and cardiovascular disease (Longo et al., 2021).

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

As for how much to eat, there is a lot of evidence that calorie restriction, limitation of available nutrients without malnutrition, can reduce the incidence of and slow the progression of many age-related pathologies. A broad review of meta-analyses of studies conducted on rodents showed that a low-calorie diet played a vital role in increasing the lifespan of an individual (Ekmekcioglu, 2020).

Similar research has and continues to be done on interventions such as exercise, nutritional supplements, sleep, stress management and even hormesis inducing activities such as sauna bathing and exposure to cold. Typically, for purposes of experimental control, researchers examine one intervention at a time (e.g., type of diet, exercise, or nutritional supplement) and hold other variables constant. With successful interventions, there is usually a reduction in likelihood of encountering a negative event such as contracting a major disease (e.g., cancer, cardiovascular disease) or an increase in the likelihood of obtaining a positive event, such as increased life expectancy or cognitive function.

Results such as these hold great promise for everyday people to increase their lifespans and healthspans (the number of years people remain healthy during their lives). The challenge people face in applying these scientific findings is that, while research typically looks at interventions in isolation, people will naturally want to combine interventions in order to maximize the benefits they receive. This creates an interesting optimization problem. Almost certainly, the effects of interventions are not purely additive. If a proper diet can increase lifespan by, say, five years and exercise can do the same, it may not be the case that doing both will increase lifespan by ten years. While this may not seem like a problem since combining exercise and proper diet is almost certainly better than either one alone, adding other interventions may provide little to no additional benefits.

For example, fish oil is a widely-used nutritional supplement and is considered a “heart-healthy” food that can reduce cardiovascular disease and stroke (American Heart Association, 2022). Olive oil is also widely-used and is associated with reduction in cardiovascular disease (Guasch-Ferre et al., 2022). From a consumer’s point of view, does taking both supplements provide benefits beyond just taking one or is one enough? If taking both is better than taking one, what is the added benefit? Given that people do not have unlimited resources, they may not want to spend money on all possible interventions but maximize the benefits they receive for the money they have to spend.

Unfortunately, these questions are largely unresolved. The purpose of the present paper is to describe a methodology for combining outcomes for different interventions in order to be able to predict what combinations of interventions produce what level of benefits.

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

Methodology

The goal of our methodology is to combine results for different interventions into an overall recommendation for what people can do to optimize their lifespans and healthspans. There are two potential sources of data we can use: results from scientific studies that contain rigorous experimental designs but may have the weakness that only single interventions are investigated and results from people's individual lives that may combine different interventions but lack scientific rigor in their implementation. In order to utilize individual lifestyle results, we design a survey including questions about a person's lifestyle such as what foods they eat, what exercise(s) they do, how well they sleep, what supplements they take, etc. We could use machine learning methods such as Random Forest (Breiman, 2001), which has high prediction accuracy, to process the results. However, we are more concerned with the interpretability of the model to provide the health guidance, so a regression model is better than other machine learning methods. If enough data are not available, using simple linear regression (Aalen, 1989) would lead to overfitting(Hawkins, 2004). Therefore, we can a Bayesian regression (Bishop and Tippin, 2003) method, which doesnot need a lot of data. Bayesian analysis combines prior information (McCarthy and Masters, 2005) and observed data. We can obtain the prior information via meta-regression (van Houwelingen et al., 2002) on scientific articles, which is a linear mixed model (Aalen, 1989) with random effect (Borenstein et al., 2010), capturing the heterogeneity of different results. Because different research papers have different variables such as duration, target population, observational study or randomized control trial(RCT), and they obtained different effect sizes, we need to combine these individual effect sizes into an overall effect size. Meta-regression can summarize an overall effect size on specific outcome for each covariate like diet, physical activity, and sleep index. After that, we obtain an overall effect size estimation for each covariate, and we use them as the prior information. Then, we use the answers from the survey as the observed data, then we conduct the Bayesian regression to estimate the weight of each covariate and interpret their effect on health.

1. Prior information:

We have four choices of prior distribution (Winkler, 1967) that we can consider. After we get the data, we can try different prior setting methods and compare their performance. First, we should list all the covariates including diet, disease history and sleep indexthat we would like to control in our final model. Then, for each covariate, we collect 10~20 academic papers and use meta-regression to calculate the overall effect size. If we are unable to collect enough academic papers results, we can also refer to existing meta-analysis (Hedges, 1992) results from the literature. These results of meta-analysis or meta-regression can be considered as our prior information. Although sometimes we cannot obtain the information about several covariates via the literature, we can set the prior of coefficients in Bayesian regression model equal to the OLS

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

estimate(Pohlmann and Leitner, 2003)of coefficients in a linear regression (Su et al., 2012). We also have other prior setting methods such as uniform prior(Shulman and Feder, 2004), unit information prior and g-prior (Maruyama and George, 2011). In summary, there are many prior setting methods. When we only have a small amount of observed data, the prior information dominates the estimation of this model. However, when we have a large amount of observed data, the prior information only has very small weight in our final estimations because Bayesian regression estimates are the combination of prior information and observed data information.

2. Data information:

We can obtain the observed data from the survey we design and collect. Based on different outcome variables, we can develop different models. If the outcome variable is a continuous variable, we utilize the Bayesian linear regression model. If the outcome variable is a binary variable, we can use Bayesian logistic regression (Pohlmann and Leitner, 2003). If the outcome variable is a discrete and non-negative variable, we can use Bayesian Poisson regression (Chan and Vasconcelos, 2009).

3. Bayesian regression model

3.1 Bayesian linear regression with semi-conjugate prior (Baldwin and Larson, 2017)

To carry out a Bayesian analysis we need to specify prior distribution on regression parameters β and on the variance σ^2.

$$p(\beta,\sigma^2) = p(\beta)\cdot p(\sigma^2) = N_p(\beta_0,\mathbf{\Sigma}_0)\cdot \text{InverseGamma}(\tfrac{\nu_0}{2},\tfrac{\nu_0\sigma_0^2}{2})$$

The joint posterior distribution cannot be derived in closed form, so we can approximate it via Gibbs sampling (George and McCulloch, 1993):

1) Choose a number S of iterations.

2) Choose initial values for β and σ^2 respectively.

3) repeat the following two steps:

$$\text{sample } \sigma^{2(j)} \text{ from } p(\sigma^2|\mathbf{y},\mathbf{X},\beta^{(j-1)})$$
$$\text{sample } \beta^{(j)} \text{ from } p(\beta|\mathbf{y},\mathbf{X},\sigma^{2(j)})$$

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

Discussion

The present methodology is designed to apply to investigations characterized by rapidly changing research results and the need to combine results from different interventions into an overall picture of what outcomes those combinations of interventions will produce, both conditions applying to the exciting field of longevity and anti-aging research. Moreover, the sources of data will be fundamentally different, ranging from rigorous scientific data to collections of data from individual lifestyle logs. The relative ratio of these sources of data are expected to change over time as well. Initially, we would rely largely on scientific research results that focus on individual interventions and single or small numbers of dependent measures. Over time, as people respond to survey data or supply data from daily lifestyle logs that represent a combination of interventions and health outcomes. The proposed methodology is designed to take this into account by beginning with a meta-regression of scientific papers and transitioning to survey and individual log data. Fortunately, the present model is self-correcting, thereby enabling updates to the predictive model as new, and potentially conflicting, data are generated.

References

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907-925. doi:10.1002/sim.4780080803

American Heart Association. Fish and omega-3 fatty acids. American Heart Association. https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/fats/fish-and-omega-3-fatty-acids. Accessed March 25, 2022.

Baldwin, S.A. & Larson, M.J. (2017). An introduction to using Bayesian linear regression with clinical data. *Behaviour Research and Therapy*, 98:58-75. doi:10.1016/j.brat.2016.12.016

Bishop, C. M. &Tipping, M.E. (2003). Bayesian regression and classification. *Nato Science Series sub Series III Computer And Systems Sciences*,190:267-288.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*. 1(2):97-111. doi:10.1002/jrsm.12

Boina, N., Agarwal, J., Agarwal, T., Leddo, J. et al**.** (2021a). A Novel Meta-Machine Learning Approach to Diagnose Stress from Environmental Factors Using Automated Knowledge Graphs, International Journal of Social Science and Economic Research, *6(12), 4961-4970.*

Boina, N., Agarwal, J., Agarwal, T., Leddo, J. et al. (2021b). A Novel Meta-Machine Learning

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

Approach to Diagnose Stress from Individual Factors Using a Self-retrieved Dataset and then Provide Directed Treatment, International Journal of Social Science and Economic Research, *6(12), 4933-4944.*

Boina, N. et al. (2021c). A Novel Meta-Machine Learning Platform Able to Autonomously Learn How to Diagnose Acne and Jaundice. International Journal of Social Science and Economic Research, 6(10), 4151-4158.

Boina, N. et al. (2021d). A Novel Meta-Machine Learning Platform Able to Autonomously Learn How to Diagnose Autism, Breast Cancer, Melanoma Mole Cancer and Pink Eye. International Journal of Social Science and Economic Research, 6(10), 4159-4171.

Breiman, L. (2001). Random Forests. *Machine Learning*. 45(1):5-32. doi:10.1023/A:1010933404324

Chan, A.B. & Vasconcelos, N. (2009). Bayesian Poisson regression for crowd counting. In: *2009 IEEE 12th International Conference on Computer Vision*, 545-551. doi:10.1109/ICCV.2009.5459191

Ecmekcioglu, C. (2020). Nutrition and Longevity – From mechanisms to uncertainties. Critical Reviews in Food and Science Nutrition, *60(18), 3063-3082.*

George, E.I. & McCulloch, R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423):881-889. doi:10.1080/01621459.1993.10476353

Guasch-Ferré, M., Li, Y., Willett, W., et al. (2022). Consumption of Olive Oil and Risk of Total and Cause-Specific Mortality Among U.S. Adults, *J Am Coll Cardiol.*, 79 (2) 101–112.**https://doi.org/10.1016/j.jacc.2021.10.041**

Hawkins, D.M. (2004).The Problem of Overfitting,*J Chem Inf Comput Sci*., 44(1):1-12. doi:10.1021/ci0342472

Hedges, L.V. (1992). Meta-Analysis. *Journal of Educational Statistics*, 17(4):279-296. doi:10.3102/10769986017004279

Longo, V.D., Di Tano, M., Mattson, M.P. &Guidi, M. P. (2021). Intermittent and periodic fasting, longevity and disease. Nature Aging, *1, 47-59.*

Maruyama Y, George EI. Fully Bayes factors with a generalized g-prior. *The Annals of Statistics*. 2011;39(5):2740-2765. doi:10.1214/11-AOS917

ISSN: 2455-8834

Volume:08, Issue:03 "March 2023"

McCarthy, M.A. & Masters, P. (2005). Profiting from Prior Information in Bayesian Analyses of Ecological Data. *Journal of Applied Ecology,*42(6):1012-1019.

Pohlmann, J.T. & Leitner, D.W. (2003). A comparison of ordinary least squares and logistic regression (1). *The Ohio Journal of Science*, 103(5):118-126.

Roberts, M.N., Wallace, M.A., Tomilov, A.A., Cortopassi, G.A., Ramsey, J.J.& Lopez-Dominguz, J.A. (2017). A Ketogenic Diet Extends Longevity and Healthspan in Adult Mice. Cell Metabolism, *26, 539-546.*

Shulman, N. & Feder, M. (2004). The uniform distribution as a universal prior. *IEEE Transactions on Information Theory*, 50(6):1356-1362. doi:10.1109/TIT.2004.828152

Su, X., Yan, X. & Tsai, C.L. (2012). Linear regression. *WIREs Computational Statistics*, 4(3):275-294. doi:10.1002/wics.1198

Trichopoulou, A. &Vasilopoulou, E. (2000). Mediterranean Diet and Longevity. British Journal of Nutrition, Dec;84 Suppl 2:S205-9. doi: 10.1079/096582197388554. PMID: 11242471.

van Houwelingen, H.C., Arends, L.R. &Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21(4):589-624. doi:10.1002/sim.1040

Winkler, R.L. (1967). The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association*, 62(319):776-800. doi:10.1080/01621459.1967.10500894

