Journal of Social Science and Humanities, 3 (1): 07-13, 2020
e-ISSN: 2600 - 9056
© RMP Publications, 2020
DOI: 10.26666/rmp.jssh.2020.3.2

The likelihood of Choosing Alternative Source of Collagen among Consumers: Logistic Regression Approach

Jarita Duasa and Afifah Muhamad Husin
Department of Economics, Faculty of Economics and Management Sciences,
International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.

Abstract: Collagen is the main structural protein of hard and soft tissues in animals and human body. It can maintain the biological and structural integrity of extracellular matrix and also provides physical support to tissues. Collagen is much needed nowadays because there are some factors that can deplete the level of collagen within the body. High sugar consumption, exposure to ultraviolet in the sunlight, smoking, genetic changes, autoimmune disorder and aging process are the common factors that can damage the collagen within the human body. Collagen is used in pharmaceutical industries, tissues engineering, food industry, cosmetics, biomedical industry and medical field. The sources of collagens and gelatins are recently being questioned particularly among the Muslim users on the status of syariah permissible (halal). This is because the raw materials used in its manufacture could be sourced from porcine or non-halal slaughtered animals. There are growing concerns regarding the transfer of diseases from animal sources of collagen to human users and the halal issue on the sources and manufacturing processes of it. Though currently collagen can be extracted from marine life, the sources are very limited and this is the main reason why an alternative source is being explored in the current study. The current study provides an alternative source to the available collagen market using recombinant collagen-like protein. The present study aims to analyze the factors that contribute to the probability of using this alternative source of collagen among consumers. Using survey data collected within the area of Wilayah Persekutuan and Selangor (Klang Valley), the data then are tested empirically using Logistic regression. The study finds that the probability consumers opt for the alternative source of collagen is significantly determined by age of consumers and the level of their incomes. The likelihood of the consumers to opt for this alternative source of collagen is higher among elderly and lower income consumers.

Key words: *Collagen, Recombinant collagen-like protein, Logistic Regression, Malaysia*

INTRODUCTION

Collagen is the main structural protein of hard and soft tissues in animals and human body. It can maintain the biological and structural integrity of extracellular matrix and also provides physical support to tissues [1]. Collagen is a potent source of amino acid [2]. Collagen is much needed nowadays because there are some factors that can deplete the level of collagen within the body. High sugar consumption, exposure to ultraviolet in the sunlight, smoking, genetic changes, autoimmune disorder and aging process are the common factors that can damage the collagen within the human body.

There are many types of collagen that have been identified, nearly more than 28 types but they can be categorized into two according to the structure they form; fibrillar and non fibrillar. Fibrillar refers to collagen type I, II, III, X and XI. Non fibrillary refers to all types of collagen other than fibrillar's. But only five of them are the common types, namely collagen type I, type II, type III, type IV and type V. More than 90% of collagen in human body is type I due to the wide prevalence in almost all connective tissues [3]. It can be found in skin, teeth, bone, tendon, ligament and vascular ligature. The collagen is used in pharmaceutical industries, tissues engineering, food industry, cosmetics, biomedical industry and medical field [2]. Collagen can attract new skin cell to the wound site, therefore it helps healing wounds and act as a platform for new skin tissue growth. Collagen can be broken down, changed and absorbed back to body because it is resorbable. In tissue regeneration,

Corresponding Author: Jarita Duasa, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia, +603 6421 4626.

collagen-based membrane has been used to promote the growth of specific types of cell in periodontal and also implant therapy. Osteoarthritis may be treated by using collagen supplements and formulations (*wikipedia*).

One of the major industrial sources of collagen is from the skin and bone of bovine. The collagen from the skin and bone of porcine are also utilized. Recently, researchers look for the other sources than bovine due to the fact that it is a treat to human by spreading diseases especially during the mad cow outbreak. Not to mention 3% of population is allergic to collagen from bovine [2]. Just like bovine, using porcine as source of collagen is risky as it can contaminate the disease like zoonosis. The porcine is also not favored by some as it is forbidden due to the religious constraint [4]. [5] also mentioned collagen that has been extracted from land animal sources like bovine, pig and chicken faces regulatory and quality control difficulties, can contain biological contaminants and poison and other diseases. In addition to these, these land animal sources are not suitable for many religious and ethnics groups.

Presently, marine source is considered the safest source to obtain collagen for. Marine source of collagen is favored than land animals as it shown the adverse inflammatory and immunologic response and prevalence of various diseases [6]. Type I collagen is mainly obtained from the skins, bones, fins and scales of fresh and salt water fishes.

There is a need to find a new type of source due to the problem of collagen extracted from animal sources. Recombinant human collagen is a new and promising way for mass production of collagen [1]. Recently, collagen-like protein was found in bacteria and it may represent as an alternative biosynthetic collagen material and was hope to be a complement to existing sources [7]. This current research is aimed to produce ***a recombinant collagen-like protein*** that suitable for food and nutraceutical industries. The first part of the current project, thus, aims to improve the productivity of bacterial expression system through cloning, characterization and fermentation process optimization as shown by Figure 1.

Having said this, the present study attempts to explore the possibility that consumers in Malaysia will opt for this alternative source of collagen using recombinant collagen-like protein via molecular biology and biochemistry.

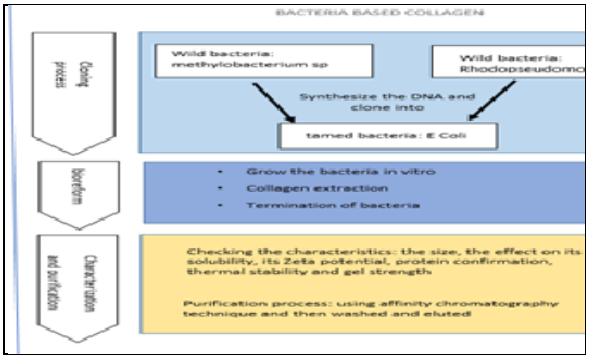

Fig 1 Process of synthesizing recombinant collagen-like protein from bacteria.

DATA AND METHODOLOGY

Data

In assessing the acceptance level or probability of opting the alternative source of collagen among consumers, a survey is conducted. The scope of this research is focused on consumers in Klang Valley, which is recorded at 1884500 as a total population, similar to total population in both areas. In order to determine the appropriate sample size for the study, this research applies the widely used formula by [8]. From the calculation, using the population size of 1,884,500 and 5 per cent level of precision or sampling error, it is suggested that the minimum sample size is 399.

When estimating the sample size for the pilot trial, the simplest methods to apply are sample size rules of thumb. [9] cites a general flat rule to 'use at least 30 subjects or greater to estimate a parameter', whereas [10] suggests a minimum sample size of 12 subjects per treatment arm. [11] recommend a pilot trial sample size of 70 in order to reduce the imprecision around the estimate of the standard deviation. The present study is able to collect 77 responses for the pilot study in order to do estimation of the model.

Methodology

The present study adopts two methods of analysis on the collected data. First, descriptive analysis is conducted with the use of frequencies. The frequencies involve analysis on the number of respondents for each variable or item such as demographic variables.

Second, the logistic model is developed to analyze the probability that respondents are willing to opt for the recombinant collagen-like protein as an alternative source of collagen. This is a nonlinear regression model specifically designed for binary dependent variables. Unlike linear probability model, this model adopts a nonlinear formulation that forces the predicted values to be between 0 and 1 by using cumulative

probability distribution function (c.d.f.) which is denoted by *F*. The logistic cumulative distribution function has a specific functional form, defined in terms of the exponential function. The population logit model of the binary dependent variable Y with multiple regressors could be expressed as:

$$\Pr(Y=1|X_1, X_2.....) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+....+\beta_k X_k)}} \quad (1)$$

The main reason for using logit regression is that the logistic c.d.f. could be computed faster than the normal c.d.f. [12]. Logistic regression model is estimated to predict a categorical (usually dichotomous) variable from a set of predictor variables. For the purpose of this study, the dependent or the outcome variable of interest was constructed as dichotomous indicator based on the response to survey question: "The recombinant collagen-like protein can be an alternative for the existing collagen product". Respondents who answered 'Agree/strongly agree' is coded as 1 and those who answer 'Disagree/strongly disagree' is coded as 0.

The predictor variables consist of dummy variables, categorical variables and continuous variables. Among dummy variables included are "gender", "citizenship", "religion", "know what is collagen". The categorical variables are such as "education level", "marital status", "ethnicity", "occupation" and "monthly income". Meanwhile, the continuous variables is "age". To simplify, we develop a Logit Model as follows:

$$L_i = ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 (AGE)_i + \beta_2 (DUMMY\ GENDER)_i + \beta_3 (DUMMY\ CITIZEN)_I + \cdots + \vartheta_I \quad (2)$$

where L_i is a dummy variable with value of 0 or 1. L_i = 0, if there is disagreement that the recombinant collagen-like protein can be an alternative for the existing collagen product and L_i = 1 if there is agreement that the recombinant collagen-like protein can be an alternative for the existing collagen product.

Logistic regression allows us to assess how well the set of predictor variables explains the categorical dependent variable. In other words, it provides an indication of the relative importance of each predictor variable or the interaction among the predictor variables. There are several steps used for the method to be conducted. First, we have to check the details of sample size data by looking at 'Case Processing Summary' table. It will show the number of cases that we expect. Besides, the table of 'Dependent Variable Encoding' tells how dependent variable is coded. In the present study, those who state that 'they agree or strongly agree that 'the recombinant collagen-like protein can be an alternative for the existing collagen product' represented by 1 and those who state that 'they disagree or strongly disagree that the recombinant collagen-like protein can be an alternative for the existing collagen product' represented by 0.

Second, 'Classification Table' is being analyzed for the heading of 'Block 0'. These results are analysis without any of independent variables used in the model. This table of results serves as a baseline later for comparing the model without predictor variables included. Overall percentage of correctly classified cases is reported here with the hope that later, when a set of predictor variables is entered, it will be able to improve the accuracy of the predictions. Third, the Omnibus test of Model Coefficients in Block 1 gives the overall indication of how well the model performs, over and above the results for Block 0, with no predictors entered into the model. This is referred to as a 'goodness of fit' test. For this result, we expect a highly significant value (the sig. value should be less than 0.05).

Fourth, Hosmer and Lemeshow test is another test that can be used to describe whether the model used is fit or not to the data. For this test, poor fit indicated by a significance value less than 0.05 for the Chi-square statistic. Fifth, the table headed 'Model Summary' gives another piece of information about the usefulness of the used model. Two *R*-squares are reported, namely, the Cox and Snell *R*-square and the Nagelkerke *R*-square. They provide an indication of the amount of variation in the dependent variables explained by the model (from the minimum value of 0 to a maximum approximately value 1). This is called pseudo R square statistics. Lastly, the importance of each of predictor variables is displayed in 'Variables in the Equation' table. The test that used here is the Wald test. At the column sig (p-value), those variables, which have values less than 0.05, are the variables that contribute significantly to the predictive ability of the model. The odd ration of each predictor is reported under the column "Exp(B)' and the coefficient of each predictor is reported under the column of 'B'. B values can be positive or negative to inform the direction of the relationship between predictor and dependent variables. In other words, it could inform which factors increase the likelihood of an 'agree' answer and which factors decrease the likelihood of a 'disagree' answer.

The odd ratio represents the change in odds of being in one of the categories of outcome when the value of a predictor increases by one unit. In the model (equation 2), if we take the antilog of the *j*th slope coefficients

(β's), subtract one from it, and multiply the result by 100, we will obtain the *percent change* in the odds for a unit increase in the *j*th regressor. The percentage change could be interpreted as probability that opt for the alternative source will change (increase or decrease) due to a unit increase in independent variables.

RESULTS AND DISCUSSION

This section provides the findings from the analysis of data. Data are analyzed descriptively as well as empirically using logistic regression.

Descriptive Analysis

As depicted in Table 1, the number of female (66.2 percent) respondents exceeds the male one (33.8 percent). Majority of them (72.7 percent) are not from main ethnicities of Malaysia in which it is suspected that most of the respondents are not Malaysians. This could be proven true as the non-Malaysian respondents are about 74 percent from the sample size. Surprisingly, most respondents are Muslims (98.7 percent) and looking at the respondents' education level, more than 85 percent of them have higher level. of education, with first degree or post-graduate degrees.

Nonetheless, the data shows that more than 70 percent of sample respondents are students. Less than 10 per cent are employed in private sector and self-employed. As of income earned, about 61 per cent of the respondents are earned less than RM500 per month. If the range of income is widened to less and equal to RM1000 per month, almost 71.4 per cent of them are in this range of income. Since Malaysia's national poverty line income (PLI) is RM800 per month, it could be inferred that majority of the respondents is fall under the category of below national poverty line or absolute poverty.

Table 1 Distribution of Respondents by Gender, Marital Status, Ethnicity, Education levels, Occupation, Citizenship, Religion and Income levels

Item	**Category**	**Frequency**	**Valid Percent**
Gender	Male	26	33.8
	Female	51	66.2
Ethnicity	Malay	17	22.1
	Chinese	2	2.6
	Indian	0	0
	Bumiputra	2	2.6
	Others	56	72.7
Citizenship	Malaysian	20	26.0
	Non-Malaysian	57	74.0
Religion	Muslim	76	98.7
	Non-Muslim	1	1.3
Marital Status	Single	54	70.1
	Married	21	27.3
	Divorced	2	2.6
Educational Level	No education	0	0
	Primary education	1	1.3
	Secondary education	4	5.2
	Diploma/pre-university	4	5.2
	First Degree	32	41.6
	Postgraduate education (Master/PhD)	36	46.8
Occupation	Public sector	0	0
	Private sector	6	7.8
	Self-employed	7	9.1
	Housewife/homemaker	3	3.9
	Retired	0	0
	Unemployed	2	2.6
	Student	59	76.6
Monthly income	≤ RM500	47	61.0
	RM501 –RM1,000	8	10.4
	RM1,001 – RM1,500	5	6.5
	RM1,501 – RM2,000	11	14.3
	RM2,001 – RM2,500	3	3.9
	RM2,501 – RM3,000	2	2.6
	> RM3,000	1	1.3

In survey conducted, respondents (consumers) are asked on the concept of recombinant collagen-like protein (RCLP) that is suggested in the current study to be an alternative source of the current source of collagen. Surprisingly, more than 70 percents of respondents are familiar with the concept of recombinant collagen-like protein which implies than they are aware of current method of extracting collagen which has been used in health care and cosmetic industries. About 58.4 percent of respondents have heard about RCLP in Malaysia and about 44.2 percent of them have heard about RCLP in other countries. They are also asked on suitability of RCLP as an alternative of collagen for Malaysian community that consists of Muslims and non-Muslims. 89.6 percent of respondents agree and strongly agree that RCLP is suitable for the community even though 10.4 percent says it is not suitable. Bigger percentage of responses supports the suitability of RCLP for Malaysian community as this might be a better choice for Muslim population who are doubtful on the *halal* source of the

products. In fact, 92.2 percent agrees that RCLP could be an alternative for the existing collagen product.

Table 2: Information on Recombinant Collagen-Like Protein

Item	Strongly Disagree (SDA)	Disagree (DA)	Agree (A)	Strongly Agree (SA)
I am familiar with concept of recombinant collagen-like protein	1 (1.3)	21 (27.3)	52 (67.5)	3 (3.9)
I have heard about recombinant collagen-like protein in Malaysia	0 (0.0)	32 (41.6)	42 (54.5)	3 (3.9)
I have heard about recombinant collagen-like protein in other countries	0 (0.0)	43 (55.8)	31 (40.3)	3 (3.9)
The collagen using recombinant collagen-like protein is suitable for our community (Muslims and non-Muslims)	0 (0.0)	8 (10.4)	63 (81.8)	6 (7.8)
The recombinant collagen-like protein can be an alternative for the existing collagen product.	0 (0.0)	6 (7.8)	64 (83.1)	7 (9.1)

Logistic regression

The present study develops logistic model to further analyze the probability that the respondents are willing to opt for the recombinant collagen-like protein as an alternative for the existing collagen product. Using the survey feedback, responses on the question whether 'the recombinant collagen-like protein can be an alternative for the existing collagen product' are converted into binary numbers '0' and '1' in order to develop logistic regression. These data are for the dependent variable. The responses of Strong Agree and Agree are computed as '1' and the responses of Strongly Disagree and Disagree are computed as '0'. The set of predictors includes in the model are categorical variables, continuous variables and dummy variables as mentioned earlier in methodology section. The results of logistic regression are displayed on Table 3.

Table 3: Logistic regression results

Independent variables	***Binary logistic*** *Dependent variable:* 'The recombinant collagen-like protein can be an alternative for the existing collagen product' *(Agree=1, Disagree = 0)*	
	B	*Exp(B)*
Constant	24.91 (48839)	7E+010
Dummy_Gender	-1.046 (1.96)	0.351
Age	0.359** (0.181)	1.432
Dummy_citizenship	18.22 (40192)	5E+012
Dummy_religion	-17.506 (40192)	0.000
Ethnicity_Malay		
Ethnicity_Chinese	29.216 (40192)	2291412.8
Ethnicity_Bumiputra	14.645 (4759.62)	5E+012
Ethnicity_others	69.269 (47052)	1E+030
Education_primary		
Education_secondary	18.432 (40192)	1E+008
Education_diploma/pre-U	19.014 (57661)	2E+008
Education_degree	3.550 (5.316)	34.821
Education_postgraduate	1.314 (1.634)	3.723
Marital status_single		
Marital status_married	-12.911 (27745)	0.000
Marital status_divorce	52.905 (29595)	9E+022
Income_<RM500		
Income_up to RM1000	-2.140 (2.147)	0.118

Income_up to RM1500	-3.616* (2.116)	0.027
Income_up to RM2000	17.487 (26985)	39319877
Income_up to RM2500	17.573 (33186)	42849848
Income_up to RM3000	-35.757 (6731)	0.000
Income_>RM3000	2.825 (75512)	16.854
Occupation_private sector		
Occupation_self-employed	-2.731 (29475)	0.065
Occupation_home maker	-70.299 (28216)	0.000
Occupation_unemployed	-74.404 (10299)	0.000
Occupation_student	-67.227 (62635)	0.000
% correct classification (Block 1)	94.8	
Omnibus Chi-square stat.	26.016(sig = 0.30)	
Hosmer & Lemeshow Test stat.	0.405(sig.= 0.999)	
Cox &Snell R-square	0.287	
Negelkerke R-square	0.680	

From Table 3, the regression results show that the significant predictors in the model are 'age' and 'Monthly Income between RM1001 to RM1500'. Other independent variables are not significantly contributing to the likelihood that consumers choose recombinant collagen-like protein as an alternative of the existing source of collagen. The independent variables of the categorical variables such as income level, education level, marital status and occupation are interpreted in such a way that the coefficients or betas are the difference between the focus category as compared to the base category.

The negative value of beta (β) for 'Income level between RM1001 and RM1500 per month', that is -3.616, indicates that an increase in independent variable score as compared to the base category (income level less than RM500 per month) will result in a decrease in probability of the case recording a score of 1 or 'yes' in the dependent variable. The odd ratio (Exp (β)) for this variable is 0.027. This could be interpreted that the odds of a person answering 'yes', that recombinant collagen-like protein can be an alternative of the existing source of collagen, is 0.027 times lower for someone who earn income between RM1001 and RM1500 per month as compared to those who earn less than RM500 per month, all other factors being equal.

Positive and significant β values for 'age' variable suggests that elderly consumers are more likely to state that recombinant collagen-like protein can be an alternative of the existing source of collagen, other things equal. Observing the odd ratios of this variable, they represent the change in odds of being in one of the categories of outcome when the value of each predictor increases by one unit. The odds of a person agrees that recombinant collagen-like protein can be an alternative of the existing source of collagen is 0.36 times higher among older consumer than the young one, all other factors being equal.

In the classification table, with no predictor (Block 0), the overall percent of correctly classified cases is 92.2%. When a set of predictor variables is entered (Block 1), it improves the accuracy of this prediction to 94.8%. The Omnibus tests of Model coefficients (the Chi-square statistic) in Block 1 however is not significant (p-value > 0.05) but the Hosmer & Lemeshow test supports the 'goodness of fit' of the model with the Chi-square statistics of 0.405 and significance level of 0.99. The pseudo R-square statistics (Cox & Snell R-square and Nagelkerke R-square) show that between 29% and 68% of the variability in the dependent variable is explained by the set of predictor variables. In general, the model obtained is considered as a good model.

CONCLUSION

The present study attempts to proposed an alternative source to the available collagen market using recombinant collagen-like protein and to assess the probability that the members of community (consumers) will opt for the alternative source of collagen. To achieve the objectives, survey questions are designed and primary data are collected through a survey on a sample of consumers in Klang Valley, Malaysia. Data are tested empirically using Logistic regression.

The study finds that the probability consumers opt for the alternative source of collagen is significantly determined by age of consumers and the level of their incomes. The likelihood of the consumers to opt for

this alternative source of collagen is higher among elderly and lower income consumers. Aging group of consumers is starting to realize the importance of collagen in maintaining skin health, improving immunity, activating muscles and treating joint pain. Collagen finds extensive application in the healthcare industry and it is also used widely in the food industry as a food supplement in capsule, beverage and tablet forms. Since it is considered as necessities in consumption of most consumers, it is unsurprised to find that the probability to choose way of extracting collagen using an alternative method is quite high particularly when the cost of buying them could be lessened through this alternative method of extraction. Nonetheless, recently, the concerns regarding the transfer of diseases from certain raw material sources have hindered market growth of collagen. Thus, it is recommended that the industry should be regulated by laws and regulations governing animal-based raw materials and ingredients used in collagen production in order to maintain its market growing.

ACKNOWLEDGMENTS

This study is extracted from of a research project funded by the Trans-disciplinary Research Grant Scheme (TRGS), TRGS/1/2018/UIAM/01/1/2, Ministry of Higher Education, Malaysia. The authors would like to thank the Ministry for generously funding the research.

REFERENCES

[1] Dong, C. and Lv, Y. 2016. Application of Collagen Scaffold in Tissue Engineering: Recent Advances and New Perspectives, *Polymers* 8 (42).

[2] Silvipriya, K., Krishna, Kumar K., Bhat, A., Dinesh Kumar B., John, A. and Lakshmanan, P. 2015. Collagen: Animal Sources and Biomedical Application, Journal of Applied Pharmaceutical Science 5(3), 123-127.

[3] Cheah, K.S.E.1985. Collagen Genes and Inherited Connective Tissue Disease. Biochem. J. 229, 287-303.

[4] Cortial, D., Gouttenoire, J., Rousseau, CF., Ronziere, MC., Piccardi, N., Msika, P., Herbage, D., Mallein-Gerin, F. and Freyria, AM. 2006. Activation by IL-1 of Bovine Articular Chondrocytes in Culture within a 3D Collagen-based Scaffold. An in vitro Model to Address the Effect of Compounds with Therapeutic Potential in Osteoarthritis. Osteoarthritis and Cartilage 14, 631-640.

[5] Aberoumand, A. 2012. Comparative Study between Different Methods of Collagen Extraction from Fish and its Properties. World Applied Sciences Journal 16(3). 316-319.

[6] Addad, S., Jean-Yves, E., J., Faye, C., Richard-Blum, S. and Lethias, C. 2011. Isolation, Characterization and Biological Evaluation of Jellyfish Collagen for the Use in Biomedical Applications. Marine Drugs 9(6), 967-983.

[7] An B., Kaplan D. and Brodsky B. 2014. Engineered Recombinant Bacterial Collagen as an Alternative Collagen-based Biomaterial for Tissue Engineering. Frontiers in Chemistry 2, pp 1-5.

[8] Yamane, T. 1967. Statistics: An Introductory Analysis, 2nd edn. Harper and Row, New York.

[9] Browne R.H. 1995. On the use of a pilot sample for sample size determination. Stat Med 14, 1933–1940.

[10] Julious S.A. 2005. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Stat 4, 287–291.

[11] Teare M.D., Dimairo M, Shephard N. 2014. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials 15, 264–264.

[12] Stock, J. and Watson, M. 2007. Introduction to Econometrics, Pearson Education Inc., Boston.