**Shanyhin Anton, Babienko Volodymyr, Strakhov Yevhen, Korkhova Arina. Mathematical modeling of the dependence of the risk of vitamin D deficiency on anthropometric and laboratory parameters. Journal of Education, Health and Sport. 2023;13(4):356-366. eISSN 2391-8306. DOI http://dx.doi.org/10.12775/JEHS.2023.13.04.042**
**https://apcz.umk.pl/JEHS/article/view/43825**
**https://zenodo.org/record/7892342**

The journal has had 40 points in Ministry of Education and Science of Poland parametric evaluation. Annex to the announcement of the Minister of Education and Science of December 1, 2021. No. 32343.
Has a Journal's Unique Identifier: 201159. Scientific disciplines assigned: Physical Culture Sciences (Field of Medical sciences and health sciences); Health Sciences (Field of Medical Sciences and Health Sciences).

Punkty Ministerialne z 2019 - aktualny rok 40 punktów. Załącznik do komunikatu Ministra Edukacji i Nauki z dnia 1 grudnia 2021 r. Lp. 32343. Posiada Unikatowy Identyfikator Czasopisma: 201159.
Przypisane dyscypliny naukowe:Nauki o kulturze fizycznej (Dziedzina nauk medycznych i nauk o zdrowiu); Nauki o zdrowiu (Dziedzina nauk medycznych i nauk o zdrowiu).



Received: 10.03.2023. Revised: 24.03.2023. Accepted: 28.04.2023.

# Mathematical modeling of the dependence of the risk of vitamin D deficiency on anthropometric and laboratory parameters

**Anton Shanyhin[1], Volodymyr Babienko[1], Yevhen Strakhov[2], Arina Korkhova[2]**

**[1]Odessa National Medical University**
**[2]Odessa I.I. Mechnikov National University**

**Shanyhin Anton:** ORCID 0000-0003-2644-4542; https://orcid.org/0000-0003-2644-4542
e-mail: anton.shanyhin@onmedu.edu.ua
**Babienko Volodymyr:** ORCID 0000-0002-4597-9908;
https://orcid.org/0000-0002-4597-9908; e-mail: volodymyr.babiyenko@onmedu.edu.ua
**Strakhov Yevhen:** ORCID 0000-0001-8207-8108; https://orcid.org/0000-0001-8207-8108
e-mail: strakhov.e.m@onu.edu.ua
**Korkhova Arina:** ORCID 0009-0003-7234-5358; https://orcid.org/0009-0003-7234-5358
e-mail: askorkhova@gmail.com

## Abstract

Recently, data on the use of mathematical models to determine the status of vitamin D in clinical practice in patients with various diseases appear in the literature, but there is no data on the creation of a mathematical model that allows screening in the general population. The creation of a mathematical model, with the help of which it is possible to identify groups at risk of vitamin D deficiency, will contribute to the reduction of laboratory research costs and allow screening of vitamin D status on a more massive and reasonable basis.

Goal. Study of the relationship between the risk of vitamin D deficiency and anthropometric and laboratory parameters using mathematical modeling methods.

Materials and methods. 928 residents of the southern region of Ukraine between the ages of 19 and 82 (average age - 47.2 years) were surveyed. Based on the correlation analysis of the studied data, the most influential data were selected as the following indicators: age, body mass index, atherogenicity coefficient, high-density lipoproteins, and the ratio of waist volume to hip volume.

Conclusions. The study emphasizes the importance of anthropometric and laboratory parameters in predicting the risk of vitamin D deficiency. The use of mathematical modeling methods makes it possible to identify the most important risk factors and quantify their contribution. The results of the study may have important implications for public health by helping to identify high-risk groups in a timely manner.

**Key words**: **mathematical modeling; vitamin D; anthropometry; blood lipids; prevention**

**Introduction**. Vitamin D is a fat-soluble vitamin and has a pleiotropic effect in many tissues of the body, including the musculoskeletal system and adipocytes. [1, 2, 3] Numerous epidemiological, experimental and clinical studies indicate the presence of correlations between the level of vitamin D in blood serum and the risk of developing dyslipidemias, carbohydrate metabolism disorders and diseases of the cardiovascular system. [4, 5, 8] These diseases create a significant burden on public health, increasing the costs of the health care system, contributing to the growth of morbidity and mortality. [1, 4, 6] The relationship of vitamin D levels with high-density lipoproteins (HDL), low-density lipoproteins (LDL), very low-density lipoproteins (VLDL), triglycerides (TG) and the atherogenic coefficient (AC) has been proven. [1, 3, 7]

Separately, it was established that the cause of low concentration of 25(OH)D in blood serum is overweight and obesity. [7] Correlations between body mass index (BMI) and levels of 25-hydroxyvitamin D are described in the literature. Such changes are due to a decrease in the bioavailability of vitamin D due to sequestration in fat deposits. [6, 8]

Today, most of the world's scientists recognize the need for early diagnosis and prevention of vitamin D deficiency conditions, especially in risk groups. [3, 4, 9] Thus, timely screening for deficiency and insufficiency of vitamin D becomes relevant, since correction of 25(OH)D status is a pain and easy task, than treatment of diseases associated with its low level. Because of this, the number of laboratory requests for determination of serum 25(OH)D is increasing, which additionally causes an increase in costs to the health care system. [10, 11]

Recently, data on the use of mathematical models to determine the status of vitamin D in clinical practice in patients with various diseases appear in the literature, but there is no data on the creation of a mathematical model that allows screening in the general population.

[4, 12] The creation of a mathematical model, with the help of which it is possible to determine groups at risk of vitamin D deficiency, will contribute to the reduction of laboratory research costs and allow screening of vitamin D status more massively and reasonably.

**Goal**. Study of the relationship between the risk of vitamin D deficiency and anthropometric and laboratory parameters using mathematical modeling methods.

**Materials and methods**. In the course of the study, 928 residents (507 women, 421 men) of the southern region of Ukraine aged from 19 to 82 years (average age — 47.2 years) were examined. All patients who participated in the study previously provided written consent to participate in the study. The research was carried out with the provision of safety measures for life and health, with observance of human rights and moral and ethical norms, which corresponds to the principles of the Helsinki Declaration of Human Rights, the order of the Ministry of Health of Ukraine No. 693 of 01.10.2015, the Convention of the Council of Europe on Human Rights and of biomedicine (ETS-164) dated 04/04/1997 and was approved by the bioethics commission of Odesa National Medical University (protocol No. 12 dated 12/23/2019).

In order to determine the degree of obesity, BMI was determined using the formula body weight/height (kg/m2) according to the recommendations of the International Group on Obesity of the WHO (WHO, 1997). A waist-to-hip ratio (WHR) greater than 0.8 was considered a criterion for abdominal obesity. Determination of the level of vitamin 25(OH)D total (estimation of the total level of 25(OH)D2 and 25(OH)D3) was performed using an automatic immunochemical analyzer Architech i2000sr (Abbott, USA). The lipid spectrum of the blood was determined according to generally accepted indicators (total cholesterol, TG, HDL, VLDL, LDL, AC). The concentration of total cholesterol, TG, and HDL was determined by the enzymatic-colometric method on an automatic biochemical analyzer Cobas 6000; Roche Diagnostics.

Based on the paired quantitative correlation analysis of the studied data, the following indicators were selected as indicators that can influence the level of vitamin D: age, BMI, AC, TG, VLDL, HDL, LDL, total cholesterol, WHR.

Adequacy of linear regression models was assessed using adjusted coefficients of determination. The influence of model factors was evaluated using standardized regression coefficients and elasticity coefficients. Metrics based on the error matrix and ROC analysis were used to assess the quality of logistic regression models. The results were considered statistically significant at $p<0.001$. Mathematical data are presented in the text in the form of tables and graphs.

**Results and Discussion**. Multifactor linear regression models were built with various combinations of factors, which were used to determine the influence of certain factors on the formation of the value of the studied indicator:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n,$$

where y is the value of the modeled indicator; $\alpha_0$, …, $\alpha_n$ are the coefficients of the linear multiple regression equation; $x_1$, ..., $x_n$ are factors that potentially affect the value of the indicator y. The sign of the coefficient $\alpha_i$ indicates the direction of the relationship between the factor with index i and the studied indicator, that is, the tendency to increase or decrease the value of the modeled indicator due to the increase of the selected factor.

In the table 1 presents non-standardized coefficients and indicators of adequacy of multivariable linear models built on factors affecting the value of the 25(OH)D level.

*Table 1*

**Models for estimating the value of the 25(OH)D level based on laboratory test data (unstandardized coefficients)**

| Model (y) | $R^2_T$ | $R^2_A$ | Factors of influence (x), coefficients ($\alpha$) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Age, $\alpha_1$ | BMI, $\alpha_2$ | AC, $\alpha_3$ | TG, $\alpha_4$ | VLDL, $\alpha_5$ | HDL, $\alpha_6$ |
| 1 | 0.223 | 0.217 | -0.340 | -0.405 | -0.369 | -1.615 | 2.906 | 1.192 |
| 2 | 0.224 | 0.219 | -0.340 | -0.407 | -0.357 | -0.290 | - | 1.204 |
| 3 | 0.224 | 0.219 | -0.342 | -0.398 | -0.592 | -0.235 | - | - |
| **4** | **0.224** | **0.220** | **-0.342** | **-0.407** | **-0.631** | **-** | **-** | **1.004** |
| 5 | 0.224 | 0.220 | -0.343 | -0.399 | -0.788 | - | - | - |
| 6 | 0.219 | 0.216 | -0.351 | -0.427 | - | - | - | - |
| 7 | 0.199 | 0.197 | -0.367 | - | - | - | - | - |

Table 1 includes models with the highest values of adequacy indicators and a different number of factors. Based on the $R^2_T$ indicator, we can see that models 2, 3, 4 and 5 are the best, and according to the $R^2_A$ indicator, models 4 and 5. For further analysis, we will choose model 4, since it includes the HDL indicator, which directly affects the atherogenic factor and risk development of atherosclerosis. According to correlation analysis, patients with higher HDL levels had higher serum 25(OH)D levels.

From a practical point of view, it is more appropriate to predict not the level of vitamin D, but its status, that is, the risk of deficiency or insufficiency. For this purpose, the previously obtained indicator of 25(OH)D was transformed into a binary variable according to its values. According to the classification developed by the International Institute of Medicine and the Committee of Endocrinologists for the creation of guidelines for clinical practice and Methodological recommendations for the treatment and prevention of vitamin D deficiency in the population of Central European countries, concentrations <30 ng/ml were defined as an insufficient level of vitamin D. [13]

For a preliminary assessment of the level of influence of the studied anthropometric and laboratory indicators on the risk of vitamin D deficiency, biserial correlation coefficients were calculated, which demonstrate the closeness and direction of the relationship between the factor and the risk of vitamin D deficiency. Their values and confidence levels are shown in Table 2.

*Table 2*

**Correlations between the risk of 25-hydroxyvitamin D deficiency and other indicators**

| Indicator | r | p |
|---|---|---|
| Age | 0.417 | <0.001 |
| BMI | 0.150 | <0.001 |
| AC | 0.128 | <0.001 |
| TG | 0.075 | 0.02 |
| VLDL | 0.071 | 0.03 |
| HDL | -0.117 | <0.001 |
| LDL | 0.091 | 0.006 |
| Total cholesterol | 0.031 | 0.34 |
| Waist-to-hip ratio (WHR) | 0.144 | <0.001 |

Therefore, it can be concluded that the most influential factors are age, BMI, CA, HDL and the WHR ratio.

Mathematical models were then constructed using a logistic regression-based classification algorithm to predict the risk of serum 25(OH)D deficiency. Within logistic regression analysis, the probability of an insufficient level of 25(OH)D, expressed through logistic regression, can be expressed as the following equation:

$$OR = \frac{P}{(1-P)}, P = \frac{1}{1+e^{-y}},$$

де $log(OR) = y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$, $P$ – the possibility of insufficient vitamin D levels.

Table 3 presents the logistic regression coefficients for the corresponding indicators for previously obtained models 1-6.

*Table 3*

**Logistic regression coefficients of indicators of models 1-6**

| Model | Coefficients ($\beta$) | | | | | |
|---|---|---|---|---|---|---|
| | Age | BMI | AC | TG | VLDL | HDL |
| 1 | 0.064 | 0.036 | -0.059 | 0.261 | -0.593 | -0.664 |
| 2 | 0.064 | 0.037 | -0.062 | -0.009 | - | -0.665 |
| 3 | 0.065 | 0.043 | 0.067 | -0.041 | - | - |
| **4** | **0.064** | **0.037** | **-0.069** | **-** | **-** | **-0.668** |
| 5 | 0.065 | 0.044 | 0.035 | - | - | - |
| 6 | 0.065 | 0.045 | - | - | - | - |

To assess the quality of the obtained models, several metrics were used based on the conjugation table, also known as the error matrix, which is built on the basis of the classification results (table 4). The rows of this matrix contain the true values of the classes, and the columns are the classes predicted by the model. In this study, class 0 is associated with a sufficient level of vitamin D, class 1 with an insufficient level or deficiency of vitamin D.

*Table 4*

**Error matrix based on classification results**

| | | Intended classes | |
|---|---|---|---|
| | | Negative class (0) | Positive class (1) |
| Real classes | Negative class (0) | $TN$ (true-negative cases) | $FP$ (false positive cases) |
| | Positive class (1) | $FN$ (false-negative cases) | $TP$ (true positive cases) |

Based on the error matrix, the following quality indicators of the obtained models were calculated:

1. Accuracy
2. Precision
3. Sensitivity
4. Specificity
5. F1 Score

Indicators 2-5 are recommended for use in class imbalances, which was the case for this study (33.4% of patients had adequate vitamin D levels and 66.4% were insufficient or deficient).

Table 5 shows the aforementioned metrics for estimating the risk of 25(OH)D deficiency using models 1-6.

*Table 5*

**Metrics for risk assessment of 25(OH)D deficiency levels in models**

| Mmodel | Accuracy | Precision | Sensitivity | Specificity | F1 Score |
|---|---|---|---|---|---|
| 1 | 0.750 | 0.791 | 0.878 | 0.441 | 0.832 |
| 2 | 0.746 | 0.787 | 0.878 | 0.426 | 0.830 |
| 3 | 0.737 | 0.791 | 0.854 | 0.456 | 0.821 |
| **4** | **0.746** | **0.787** | **0.878** | **0.426** | **0.830** |
| 5 | 0.741 | 0.792 | 0.860 | 0.456 | 0.825 |
| 6 | 0.741 | 0.789 | 0.866 | 0.441 | 0.826 |

According to Table 5, the most accurate model according to all criteria is model 1, which includes six indicators. However, model 4, which includes only four measures, has very close accuracy rates. Therefore, the use of model 4 with fewer laboratory parameters seems more appropriate.

In addition, ROC analysis using ROC curves (Receiver Operator Characteristic) was used to assess the quality of the models. The ROC curve shows the dependence of the number of correctly classified positive cases on the number of incorrectly classified negative cases, that is, this graph describes the relationship between the sensitivity of the model and its specificity.

In fig. 1 shows the ROC curves of models 1-6.

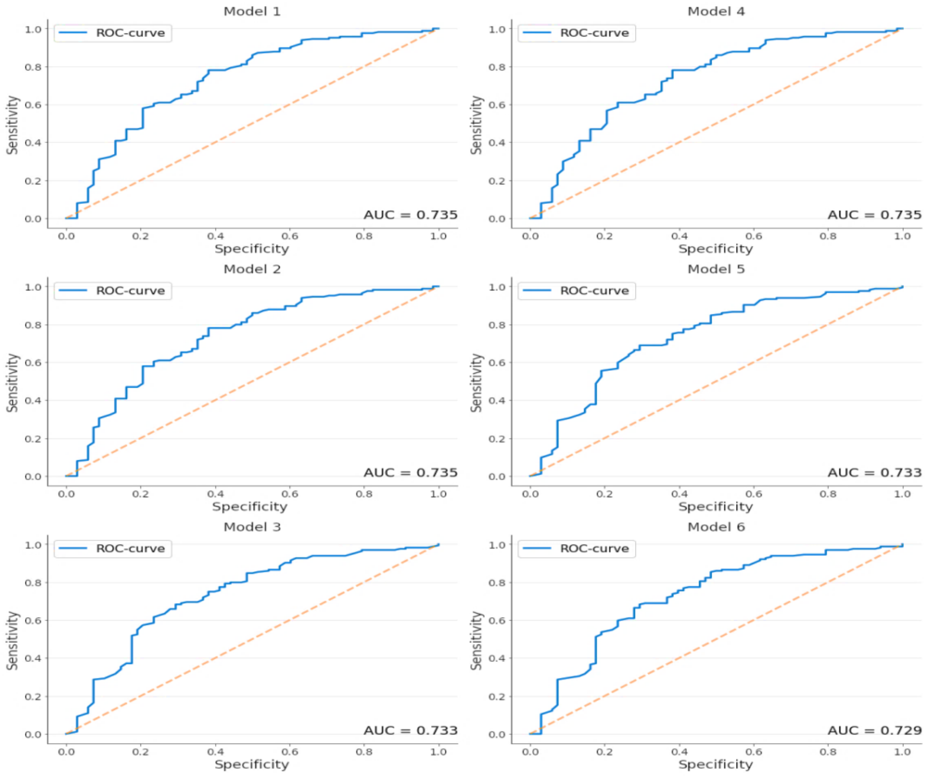


**Fig. 1. ROC curves of logistic regression for models 1-6**

The quality of the classification model is determined by the area of the figure under the ROC curve, which is called AUC (Area Under Curve). The higher the AUC value, the higher the predictive value of the model. The AUC values of the constructed models are in the range from 0.729 to 0.735, therefore, they can be considered quite effective. Note that the AUC of model 4 coincides with the AUC of model 1, that is, according to this indicator, they have the same quality, which is another argument in favor of using model 4 in practice.

It should be noted that it is the area under the ROC curve, in contrast to the classic indicators of the adequacy of the logistic regression model, that allows the most accurate assessment of the quality of each of the models.

Mathematical models are used as a tool not only for understanding the true causes of certain problems in real practice, but also for developing preventive measures to prevent risks to public health. In this study, linear multivariate models with all possible combinations of factors were analyzed in order to build the optimal model both from the point of view of statistical quality indicators and from the perspective of the specificity of the subject area, which is the main value of the work. Thus, the obtained mathematical model reasonably coincides with real data.

Given that, according to statistical analysis, the majority of the population of Ukraine has a deficiency and insufficiency of vitamin D, caused by insufficient endogenous synthesis and limited intake of it with food products, as well as the important role of vitamin D levels in the development of disorders of lipid metabolism, there is a need to implement a global vitamin prevention system D deficiency states. The use of lipidogram, age, and BMI data allows identifying risk groups for vitamin D deficiency in the population and providing more directed laboratory control of the 25(OH)D level, reducing the number of unfounded laboratory tests of vitamin D, as well as providing correction of vitamin D status in the early stages.

**Conclusions**

1. Correlation analysis and mathematical modeling of the dependence of the risk of vitamin D deficiency on anthropometric and laboratory indicators were carried out. The most influential factors were age, BMI, AC, HDL and the WHR ratio.

2. With the help of multiple logistic regression models, the risk of vitamin D deficiency in residents of the southern regions of Ukraine was assessed. The assessment of the quality of the models was carried out according to five indicators based on the error matrix, as well as the ROC analysis. According to the data of the conducted research, the most optimal is the use of model 4, which includes age, BMI, AC and HDL.

Thus, mathematical modeling makes it possible to estimate the level of influence of selected anthropometric and laboratory indicators on the level of 25(OH)D in blood serum and to predict the risk of vitamin D deficiency states. Despite the fact that laboratory control of 25(OH)D is the most informative indicator of determining the status of vitamin D in the body, the use of a mathematical model makes it possible to identify individuals at risk for further diagnosis and reduce the number of laboratory tests among people who, according to this assessment, have sufficient vitamin D level.

**References**


1. Abboud M, Liu X, Fayet-Moore F, et al. Effects of Vitamin D Status and Supplements on Anthropometric and Biochemical Indices in a Clinical Setting: A Retrospective Study. Nutrients. 2019;11(12):3032. Published 2019 Dec 12. doi: https://doi.org/10.3390/nu11123032

2. Bouillon R., Marcocci C., Carmeliet G., Bikle D., White J., Dawson-Hughes B., Lips P. Skeletal and extra-skeletal actions of vitamin D: Current evidence and outstanding questions. Endocr. Rev. 2019;40 doi: https://doi.org/10.1210/er.2018-00126

3. Shanyhin A.V. Znachennia ratsionu kharchuvannia ta rivnia insoliatsii v zabezpechenosti vitaminom D. Suchasni aspekty profilaktyky. Health of Society. 2022;11(1):16–22. https://doi.org/10.22141/2306-2436.11.1.2022.288 (in Ukrainian)

4. Prutiyan T.L., Dobrovol′s′ka O.O. Otsinka yakosti zhyttya v zhinok u postmenopauzi z arterial′noyu hipertenziyeyu, ozhyrinnyam ta osteoporozom. Suchasni aspekty modernizatsiyi nauky: stan problemy, tendentsiyi rozvytku: Materialy XXIII Mizhnarodnoyi nauk.-prakt. konf.(m. Dikirkh (Lyuksemburh) 7 serp. 2022 r.). Dikirkh, 2022. S.105-110 https://repo.odmu.edu.ua:443/xmlui/handle/123456789/11608 (in Ukrainian)

5. Arredondo A., Azar A., Recamán A.L. Diabetes, a global public health challenge with a high epidemiological and economic burden on health systems in Latin America. *Glob. Public Health.* 2018;13:780–787. doi: https://doi.org/10.1080/17441692.2017.1316414

6. Pan′kiv V.I., Povoroznyuk V.V., Pan′kiv I.V., Boyko V.I., Hluhovs′ka S.V. Stan zabezpechennya vitaminom D naselennya Zakhidnoho rehionu Ukrayiny. International Journal of Endocrinology. 2019;15(3):268-271 doi: https://doi.org/10.22141/2224-0721.15.3.2019.172115 (in Ukrainian)

7. Cimmino G, Morello A, Conte S, et al. Vitamin D inhibits Tissue Factor and CAMs expression in oxidized low-density lipoproteins-treated human endothelial cells by modulating NF-κB pathway. *Eur J Pharmacol*. 2020;885:173422. doi: https://doi.org/10.1016/j.ejphar.2020.173422

8. Park C.Y., Han S.N. The Role of Vitamin D in Adipose Tissue Biology: Adipocyte Differentiation, Energy Metabolism, and Inflammation. *J Lipid Atheroscler*. 2021;10(2):130-144. doi: https://doi.org/10.12997/jla.2021.10.2.130

9. Setayesh L, Amini A, Bagheri R, Moradi N, Yarizadeh H, Asbaghi O, Casazza K, Yekaninejad MS, Wong A, Suzuki K, Mirzaei K. Elevated Plasma Concentrations of Vitamin D-Binding Protein Are Associated with Lower High-Density Lipoprotein and Higher Fat Mass Index in Overweight and Obese Women. *Nutrients*. 2021; 13(9):3223. doi: https://doi.org/10.3390/nu13093223

10. Holick MF. The vitamin D deficiency pandemic: Approaches for diagnosis, treatment and prevention. *Rev Endocr Metab Disord*. 2017;18(2):153-165 doi: https://doi.org/10.1007/s11154-017-9424-1

11. Sizar, O.; Khare, S.; Goyal, A.; Bansal, P.; Givler, A. *Vitamin D Deficiency*; Updated 21 July 2021; StatPearls Publishing: Treasure Island, FL, USA, 2021. Available online: https://www.ncbi.nlm.nih.gov/books/NBK532266/ (accessed on 19 February 2023).

12. Khmil′ S.V., Franchuk U.YA., Malanchuk L.M., Franchuk M.V. Matematychna model′ prohnozuvannya vynyknennya preeklampsiyi seredn′oho stupenya u zhinok hrupy vysokoho ryzyku rozvytku pizn′oho hestozu na tli metabolichnoho syndromu. Zdobutky klinichnoyi i eksperymental′noyi medytsyny. 2021;3:157-161 doi: https://doi.org/10.11603/1811-2471.2021.v.i3.12530 (in Ukrainian)

13. Holick, M.F., Binkley, N.C., Bischoff-Ferrari, H.A., Gordon, C.M., Hanley, D.A, Heaney, R.P., Murad, M.H., & Weaver, C.M. (2011). Evaluation, treatment, and prevention of vitamin D deficiency. An Endocrine Society Clinical Practice Guideline. J. Clin. Endocrinol. Metab., 96 (7), 1911-1930.