








Article

Development and Validation of a Novel Pre-Pregnancy Score Predictive of Preterm Birth in Nulliparous Women Using Data from Italian Healthcare Utilization Databases

Ivan Merlo ¹, Anna Cantarutti ^{1,2,*}, Alessandra Allotta ³, Elisa Eleonora Tavormina ^{3,4}, Marica Iommi ⁵, Marco Pompili ⁶, Federico Rea ^{1,2}, Antonella Agodi ⁷, Anna Locatelli ^{8,9}, Rinaldo Zanini ¹⁰, Flavia Carle ^{2,5}, Sebastiano Pollina Addario ^{2,3}, Salvatore Scondotto ^{2,3} and Giovanni Corrao ^{1,2,11} on behalf of the CHRP-Sicily Region Working Group

- ¹ Department of Statistics and Quantitative Methods, University of Milano-Bicocca, 20126 Milan, Italy; ivan.merlo@unimib.it (I.M.); federico.rea@unimib.it (F.R.); giovanni.corrao@unimib.it (G.C.)
 - ² National Centre for Healthcare Research and Pharmacoepidemiology, University of Milano-Bicocca, 20126 Milan, Italy; f.carle@univpm.it (F.C.); walter.pollina.ext@regione.sicilia.it (S.P.A.); salvatore.scondotto@gmail.com (S.S.)
 - ³ Department of Health Activities and Epidemiological Observatory, Regional Health Authority, Sicily Region, 90145 Palermo, Italy; aleallotta@gmail.com (A.A.); elisaeleonora.tavormina@regione.sicilia.it (E.E.T.)
 - ⁴ National Research Council of Italy, Institute for Biomedical Research and Innovation, 90146 Palermo, Italy
 - ⁵ Center of Epidemiology, Biostatistics and Medical Information Technology, Department of Biomedical Sciences and Public Health, Marche Polytechnic University, 60020 Ancona, Italy; m.iommi@staff.univpm.it
 - ⁶ Regional Epidemiological Observatory, Regional Health Agency of Marche, 60125 Ancona, Italy; marco.pompili@regione.marche.it
 - ⁷ Department of Medical and Surgical Sciences and Advanced Technologies “GF Ingrassia”, University of Catania, 95123 Catania, Italy; antonella.agodi@unict.it
 - ⁸ Department of Mother and Child, ASST Vimercate, 20871 Vimercate, Italy; anna.locatelli@unimib.it
 - ⁹ School of Medicine and Surgery, University of Milano Bicocca, 20900 Monza, Italy
 - ¹⁰ Past Director of Woman and Child Health Department, Azienda Ospedaliera della Provincia di Lecco, 23900 Lecco, Italy; rinaldozanini@gmail.com
 - ¹¹ Directorate General for Health, Lombardy Region, 20124 Milan, Italy
- * Correspondence: anna.cantarutti@unimib.it; Tel.: +39-02-6448-5828; Fax: +39-02-6448-5899



Citation: Merlo, I.; Cantarutti, A.; Allotta, A.; Tavormina, E.E.; Iommi, M.; Pompili, M.; Rea, F.; Agodi, A.; Locatelli, A.; Zanini, R.; et al. Development and Validation of a Novel Pre-Pregnancy Score Predictive of Preterm Birth in Nulliparous Women Using Data from Italian Healthcare Utilization Databases. *Healthcare* **2022**, *10*, 1443. <https://doi.org/10.3390/healthcare10081443>

Academic Editors: Maria Cristina Monti and Ottavia Eleonora Ferraro

Received: 13 June 2022

Accepted: 29 July 2022

Published: 1 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Background:* Preterm birth is a major worldwide public health concern, being the leading cause of infant mortality. Understanding of risk factors remains limited, and early identification of women at high risk of preterm birth is an open challenge. *Objective:* The aim of the study was to develop and validate a novel pre-pregnancy score for preterm delivery in nulliparous women using information from Italian healthcare utilization databases. *Study Design:* Twenty-six variables independently able to predict preterm delivery were selected, using a LASSO logistic regression, from a large number of features collected in the 4 years prior to conception, related to clinical history and socio-demographic characteristics of 126,839 nulliparous women from Lombardy region who gave birth between 2012 and 2017. A weight proportional to the coefficient estimated by the model was assigned to each of the selected variables, which contributed to the Preterm Birth Score. Discrimination and calibration of the Preterm Birth Score were assessed using an internal validation set (i.e., other 54,359 deliveries from Lombardy) and two external validation sets (i.e., 14,703 and 62,131 deliveries from Marche and Sicily, respectively). *Results:* The occurrence of preterm delivery increased with increasing the Preterm Birth Score value in all regions in the study. Almost ideal calibration plots were obtained for the internal validation set and Marche, while expected and observed probabilities differed slightly in Sicily for high Preterm Birth Score values. The area under the receiver operating characteristic curve was 60%, 61% and 56% for the internal validation set, Marche and Sicily, respectively. *Conclusions:* Despite the limited discriminatory power, the Preterm Birth Score is able to stratify women according to their risk of preterm birth, allowing the early identification of mothers who are more likely to have a preterm delivery.

Keywords: preterm birth; score; real-world evidence; healthcare utilization database; nulliparous

1. Introduction

Preterm birth (PTB), defined as any live birth occurring before 37 completed weeks of gestation [1], is the leading cause of neonatal mortality, and remains the most common cause of death under 5 years of age, with over 1 million children dying annually worldwide [2]. PTB also increases the newborn's risk of dying due to other causes, mainly from neonatal infections [3], and has long-term medical and social consequences, including an increased risk of altered cardiovascular and renal function, cerebral palsy, mental retardation and disability [4,5].

The rate of PTB has increased over the past decades, reaching the value of 11.1% worldwide, ranging from 9.3% in high-income countries to 11.8% in low-income ones [6]. Several maternal characteristics have been associated with PTB, including gestational and pre-gestational factors. Gestational risk factors include poor nutritional status, stress, substance abuse (smoke, alcohol, drugs), multiple gestation, intra and extra-uterine infections and inflammations [7]. In the past decades, new gestational markers have emerged. Low cervical length and biomarkers, such as fetal fibronectin, as well as maternal hypertensive pathology and fetal underdevelopment, have shown an association with PTB, although predictive accuracy remains limited [8]. Pre-gestational risk factors include the use of assisted medical conception techniques [9], ethnic and socio-demographic characteristics, maternal age, medical disorders and history of preterm delivery, which is considered one of the strongest risk factors, with recurrence rates ranging from 15% to 50% depending on the number and gestational age of previous deliveries [7].

While many studies have evaluated the association between risk factors and preterm delivery, a large proportion of PTBs remains unexplained. A 2016 multinational study involving 4.1 million births of five countries reported that approximately two out of three PTBs lack a plausible biological explanation [10].

Better identification of women at high risk of PTB, already at the beginning of the pregnancy, is a desirable goal as it could lead to a reduction in the rate of preterm delivery by facilitating monitoring and timely intervention. Such an assessment is particularly useful in nulliparous women. Furthermore, being able to predict preterm delivery could be useful to better understand the mechanisms leading to PTB [11]. To date, researchers have mainly focused on studying specific pre-gestational conditions. However, the widespread adoption of electronic health databases now offers the opportunity to consider a large part of the mother's medical history when studying the association between maternal characteristics and preterm delivery, without making *a priori* assumptions about the predictors of PTB.

The aim of the study was to develop and validate a predictive pre-pregnancy score for preterm delivery in nulliparous women using the information collected in healthcare utilization databases of the Italian National Health Service (NHS) during the four years prior to conception.

2. Methods

2.1. Setting

The present study was based on the NHS beneficiaries of three Italian regions that joined the protocol and contributed to the data collection. The regions are located in Northern (Lombardy), Central (Marche) and Southern (Sicily) Italy, covering approximately 16.3 million people (almost 30% of the Italian population).

2.2. Data Sources

All Italian citizens have equal access to healthcare services as part of the NHS. Within each of the 21 Italian regions and autonomous provinces, computerized information systems have been created to collect a variety of data regarding beneficiaries who receive NHS assistance and the provided services. Collected information includes: (i) demographic and administrative data of beneficiaries of the NHS; (ii) hospital discharge records reporting information on diagnoses and procedures received during hospitalization, coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD9-

CM); (iii) drug prescriptions reimbursed by the NHS, coded according to the Anatomical Therapeutic Chemical (ATC) classification system; (iv) records on services provided on an outpatient basis (e.g., outpatient visits, diagnostic exams); (v) health exemptions granted to citizens, identified by a specific national code; (vi) data from Certificates of Delivery Assistance (CedAP) including information self-reported by the mother relating to her socio-economic traits, other medical information relating to pregnancy, childbirth and child health status at delivery. These various types of data are linked using, for each citizen, a single identification code recorded in all databases. To preserve privacy, each identification code is automatically deidentified. Analyses of the regional databases were performed under the rule that the inverse process, that is, patient identification, was allowed only to the Regional Health Authority upon request from the judicial authority.

2.3. Score Development

Since Lombardy has the largest resident population (16% of the entire Italian population), data from this region were used to develop the score. Deliveries occurred from January 2012 to December 2017 were selected through the CedAP registry. Only women with age at delivery between 15 and 55, gestational age between 22 and 42 weeks, and at least 4 years of traceability in healthcare databases before pregnancy were included. Records that lacked important information about the mother or child, as well as incorrect records (i.e., duplicates or records of women for whom no hospital admission reporting an ICD-9-CM code for delivery was found), and deliveries that resulted in no babies born alive were excluded. Nulliparous women were selected combining information from CedAP registry and hospital discharge database. For each woman, the 4 years preceding the beginning of gestation were taken into account. The available information made it possible to outline the profile of women during that period with respect to (i) drugs intake, (ii) health exemptions coverage, (iii) diagnoses and (iv) procedures received during hospitalization, (v) outpatient services used, and (vi) socio-demographic features. Drugs were grouped into categories based on the second level of the ATC code (i.e., first three digits). With regard to exemptions due to chronic diseases, the entire national code was taken into account, while exemptions for disabilities and other socio-economic conditions were grouped according to the first digit of the code. Diagnoses and procedures received during hospitalization were grouped considering the first three digit of the ICD9-CM code. Outpatient services were distinguished according to the code of the medical branch to which they belonged. Each woman was considered exposed to a certain factor (i.e., a variable among those listed above) if the factor code was reported at least once in the healthcare utilization databases during the considered time. Socio-demographic traits were extracted from the CedAP registry and included: (i) age at conception (≤ 25 , 26–35 and ≥ 36 years); (ii) employment status, categorized as unemployed and employed (the latter including both working women and students); (iii) marital status (married or unmarried); (iv) education, measured according to the length of formal education completed and categorized as ≤ 12 (low), from 13 to 15 (intermediate), and ≥ 16 years (high); (v) country of birth (Italy or abroad). Moreover, for each woman the type of conception was taken into account; a binary variable that indicated the use or not of assisted medical conception techniques was assigned to each subject, in accordance with what reported in the CedAP. Deliveries resulting in PTB were identified through the same database.

A training set containing 70% of the Lombardy cohort was randomly selected and used to develop the score. With the aim of selecting variables independently able to predict PTB, the least absolute shrinkage and selection operator (LASSO) method was applied, via logistic regression, on the training set. LASSO selects variables correlated to the outcome by shrinking coefficient values, down to zero for the ones not correlated to outcome [12]. The LASSO tuning parameter was chosen using a 10-fold cross-validation, which allowed selecting the parameter that guaranteed the best discriminant power for the model [13]. Only the variables that had an absolute frequency greater than or equal to 5 in the training set were considered as candidate predictors in the selection process. The absolute frequency

of 5 was arbitrary chosen as a compromise between: (i) the need to reduce the number of candidate predictors in order to have acceptable computational times (this criterion made it possible to reduce the number of covariates by 40%, see Results); (ii) avoid the selection of noise variables; and (iii) limit the exclusion of conditions associated with PTB that could positively impact the score performances. The coefficients estimated by the LASSO model were then used to assign an integer weight to each selected condition. In particular, the weights were calculated proportionally to the value of the respective coefficients, making sure the sum of all weights was equal to 100. For each woman, a total aggregate score was identified by sequentially summing the weights of the conditions to which she was exposed. To simplify the system, accounting for excessive heterogeneity of the total aggregate score, the latter was categorized by assigning increasing values of 0, 1, 2 and 3 to the categories of 0–2, 3–6, 7–15 and 16–100, respectively. The index so obtained was termed Preterm Birth Score (PTBS).

Codes used to identify the factors included in the score are shown in the Supplementary Table S1.

The predictive performance of the score was initially evaluated in the training set by constructing the receiver operating characteristic (ROC) curve and calculating the corresponding underlying area (area under the ROC curve (AUC)).

2.4. Score Validation

To assess the reproducibility and generalizability of the results, PTBS was validated under different scenarios, varying temporal and geographical conditions [14]. Firstly, an internal validation procedure was carried out considering the 30% of the cohort from Lombardy region that was excluded in the score development phase. Subsequently, the model was externally validated by considering two different cohorts, respectively, from Marche and Sicily, selected using the same inclusion/exclusion criteria as the original one except for the recruitment period (2015–2019 and 2016–2019 for Marche and Sicily, respectively).

For each validation cohort, the predictive performance was assessed through discrimination and calibration. Discrimination was evaluated by the ROC curves and the corresponding AUCs. Calibration plots displayed observed versus predicted PTB probabilities. With the aim of taking into account the different incidences of the outcome in the different regions, calibration plots were adjusted via a conservative model recalibration by updating the model intercept [15].

2.5. Statistical Software

The variables' selection procedure was performed in R (version 3.5.3) using the 'glmnet' package [16]. All other analyses were performed using the Statistical Analysis System Software (version 9.4; SAS Institute, Cary, NC, USA).

3. Results

3.1. PTBS Development

The 486,400 deliveries that took place in Lombardy between 2012 and 2017 were selected. We sequentially excluded (i) 26 records with maternal age at delivery below 15 or above 55 years, (ii) 489 records with gestational age below 22 or above 42 weeks, (iii) 89,689 deliveries from women without at least 4 years of observation in regional databases before the onset of pregnancy, (iv) 14,186 records that lacked information about the mother, (v) 1594 deliveries from women with no hospital admission, (vi) 1924 records with incorrect linkage or resulted in no babies born alive and (vii) 197,294 deliveries from not nulliparous women (Supplementary Figure S1). Of the 181,198 remaining records, 126,839 (70%) were randomly selected and used as the training set. In the latter, the proportion of preterm births was equal to 7.4%.

A total of 1771 candidate predictors were identified. Among these, 1056 had an absolute frequency of at least 5 and were considered as candidate predictors in the LASSO selection process. This set of variables included (i) 60 drugs, (ii) 482 hospital diagnosis,

(iii) 422 inpatient procedures, (iv) 44 exemptions, (v) 42 outpatient services, (vi) 5 socio-demographic conditions and (vii) the variable that identifies the use of assisted medical conception techniques.

Twenty-six variables were selected and included in the PTBS. Weights and frequencies of the conditions are shown in Table 1, while estimates of the regression coefficients are shown in Supplementary Table S2. Factors most associated with PTB (i.e., factors with weight ≥ 8) were inpatient procedures for other operations on rectum and perirectal tissue, intake of pancreatic hormones, use of assisted medical conception techniques, hospitalization diagnosed with heart failure and presence of an exemption for transplant recipients. Considering their frequency, the variables that contributed the most to the total aggregate score in the training set were intake of sex hormones and modulators of the genital system (weight 4; frequency 20.4%), use of assisted medical conception techniques (weight 10; frequency 6.0%) and age at conception ≥ 36 years (weight 3; frequency 19.6%).

Table 1. Weight and frequencies, in the training set (Lombardy region), of the 26 variables contributing to the Preterm Birth Score (PTBS), selected applying the LASSO method, via logistic regression.

Variable	Frequency (%)			Weight
	Term Birth <i>n</i> = 117,480	Preterm Birth <i>n</i> = 9358	Total <i>n</i> = 126,839	
Drugs				
Pancreatic hormones	63 (0.05)	33 (0.35)	96 (0.08)	10
Agents acting on the renin-angiotensin system	803 (0.68)	173 (1.85)	976 (0.77)	5
Sex hormones and modulators of the genital system	22,664 (19.29)	3216 (34.36)	25,880 (20.40)	4
Endocrine therapy	567 (0.48)	156 (1.67)	723 (0.57)	3
Drugs used in diabetes	459 (0.39)	112 (1.20)	571 (0.45)	3
Immunosuppressants	266 (0.23)	53 (0.57)	319 (0.25)	2
Corticosteroids for systemic use	13,200 (11.24)	1341 (14.33)	14,541 (11.46)	1
Beta blocking agents	1034 (0.88)	163 (1.74)	1197 (0.94)	1
Calcium channel blockers	498 (0.42)	98 (1.05)	596 (0.47)	1
Hospital diagnosis				
Heart failure	3 (0.00)	5 (0.05)	8 (0.01)	8
Chronic renal failure	2 (0.00)	4 (0.04)	6 (0.00)	4
Diffuse diseases of connective tissue	31 (0.03)	13 (0.14)	44 (0.03)	3
Inpatient procedures				
Other operations on rectum and perirectal tissue	7 (0.01)	7 (0.07)	14 (0.01)	12
Diagnostic procedures on liver	23 (0.02)	12 (0.13)	35 (0.03)	4
Lysis of peritoneal adhesions	586 (0.50)	132 (1.41)	718 (0.57)	4
Exemptions				
Transplant recipients	13 (0.01)	12 (0.13)	25 (0.02)	8
Diabetes mellitus	303 (0.26)	83 (0.89)	386 (0.30)	4
Systemic lupus erythematosus	97 (0.08)	27 (0.29)	124 (0.10)	4
Affections of the circulatory system	343 (0.29)	56 (0.60)	399 (0.31)	1
Chronic (active) hepatitis	270 (0.23)	51 (0.54)	321 (0.25)	1
Outpatient services				
Psychiatry	4324 (3.68)	549 (5.87)	4873 (3.84)	1
General consultation	474 (0.40)	86 (0.92)	560 (0.44)	1
Socio-demographic conditions				
Age at conception ≥ 36 years	22,095 (18.81)	2723 (29.09)	24,818 (19.57)	3
Born abroad	14,014 (11.93)	1300 (13.89)	15,314 (12.07)	1
Low education	20,597 (17.53)	1815 (19.39)	22,412 (17.67)	1
Use of assisted medical conception techniques	6005 (5.11)	1587 (16.96)	7592 (5.99)	10

As an example of PTBS calculation, suppose a woman used assisted medical conception techniques (weight = 10), was more than 36 years old at conception (weight = 3) and registered an exemption for affections of the circulatory system (weight = 1) within the previous four years. Her total aggregate score would be 14 and the corresponding PTBS value would be equal to 2.

Overall, 65.2% and 3.7% of the training set women, respectively, had the lowest (0) and the highest (3) PTBS value, and the AUC value was equal to 0.61 (95% CI: 0.60–0.61) (Figure 1). Probability of preterm delivery was 5.5%, 7.9%, 14.2% and 23.1% for PTBS value equal to 0, 1, 2 and 3, respectively.

3.2. PTBS Validation

Preterm birth frequency was equal to 7.2%, 6.2% and 5.7% in the internal validation set (Lombardy), Marche and Sicily, respectively. PTBS distribution in the validation sets was very similar to that observed in the training set, and the AUC value was equal to 0.60 (95% CI: 0.59–0.61), 0.61 (95% CI: 0.59–0.62) and 0.56 (95% CI: 0.55–0.57) in the internal validation set, Marche and Sicily, respectively (Figure 2).

Calibration plots showed that observed preterm birth probabilities for PTBS values in the three validation sets were almost identical to those expected from training set results except for the highest PTBS value in Sicily region, which had observed risk lower than expected, but still higher than the risk observed in lower levels (Figure 3). The interpolation of the calibration curves reflected this situation. Calibration intercept was equal to the ideal value of 0 in both the internal validation set and Marche region, while it was equal to 0.02 in Sicily. Similarly, calibration slope was close to ideal value of 1 in both the internal validation set (0.95) and Marche region (0.96), while it was lower in Sicily (0.67).

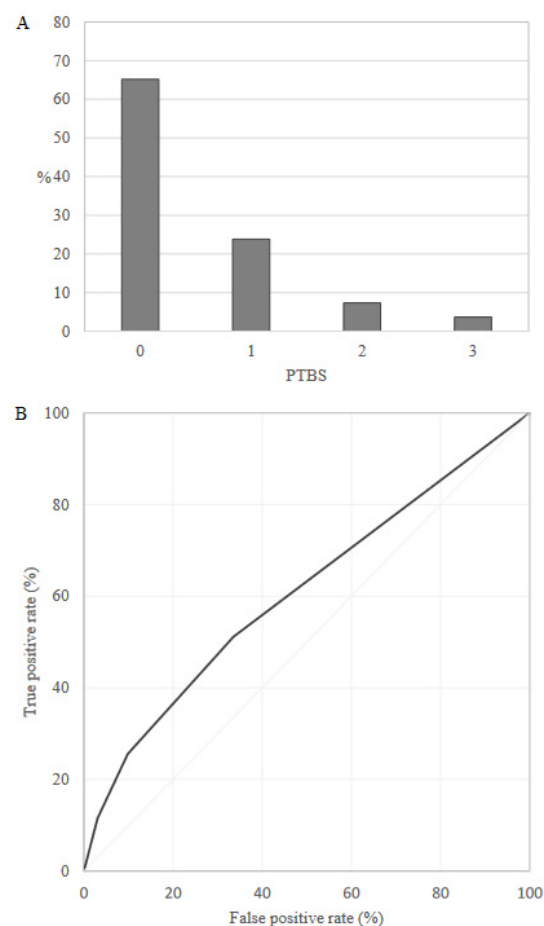


Figure 1. Preterm Birth Score (PTBS) distribution (A) and receiver operating characteristic (ROC) curve (B) in the training set (Lombardy region).

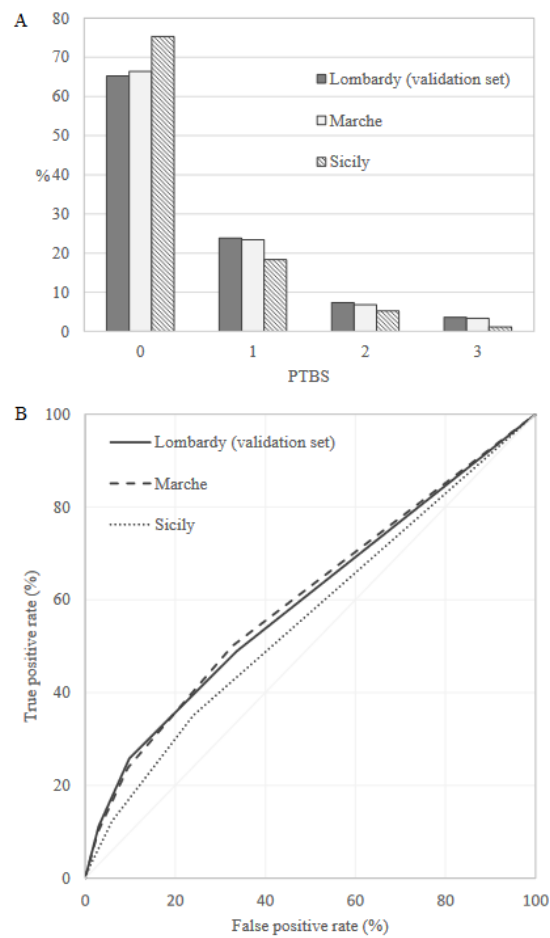


Figure 2. Preterm Birth Score (PTBS) distributions (A) and receiver operating characteristic (ROC) curves (B) comparing discriminant power in the validation sets.

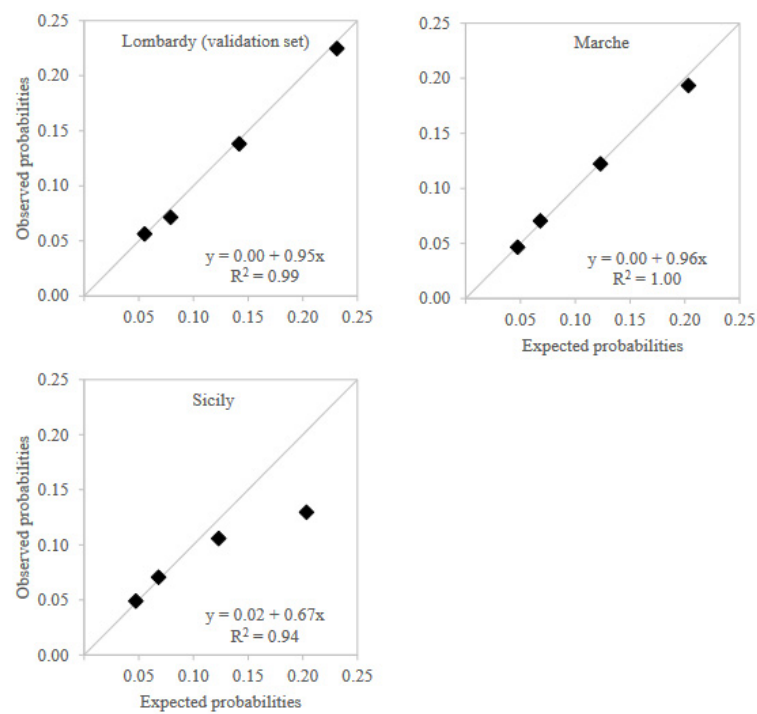


Figure 3. Calibration plots across validation territories.

Notes: Expected probabilities for Marche and Sicily regions were adjusted via a conservative model recalibration by updating the model intercept to take into account the different incidences of preterm birth between regions.

Descriptive statistics of factors included in the PTBS for validation sets were reported in the Supplementary Tables S3–S5.

4. Comment

4.1. Principal Findings

In this study, a new predictive score for preterm delivery was developed using the Italian healthcare utilization databases, considering both the woman's medical history and her socio-demographic conditions. Only nulliparous women were considered and the variables contributing to the score were selected from a large number of candidate predictors already available at the time of conception. These characteristics make PTBS a tool applicable to every woman from the very beginning of pregnancy. Furthermore, the absence of *a priori* hypotheses on the nature of predictors is an innovative approach to the topic and represents an element of novelty in the present work.

Albeit PTBS showed limited discriminatory power, the score is able to stratify women according to their risk of PTB allowing the early identification of mothers who are more likely to have a preterm delivery. The predictive performance proved to be comparable in Northern and Central Italy, while a lower discriminatory power was observed in the Sicily region. This could be partially explained by a misclassification of women with respect to the use of assisted medical conception techniques, which is one of the factors that contributes most to the score. In fact, comparing the group of women who experienced PTB with the one who did not, the proportion of women who used assisted reproduction techniques in Sicily did not differ as much as in the other regions (Lombardy: 16.9% vs. 5.1%; Marche: 14.9% vs. 4.1%; Sicily: 3.9% vs. 2.3%). The lack of a substantial difference between groups, considering that medically assisted conception is a well-known risk factor, could be due to errors in the regional database. This misclassification might have generated a conservative estimate of the PTBS performance in Sicily.

4.2. Results in the Context of What Is Known

Although moderate, the discrimination power of PTBS is consistent with that of other previously published predictive models. A recently developed neural network-based algorithm for PTB prediction in nulliparous women showed an AUC of 0.60 during the first trimester of pregnancy [17]. Another study reported an AUC of 0.66 for a predictive model based on a set of well-known risk factors for PTB which, unlike the present study, also included information on ongoing pregnancy and history of preterm births [10]. Furthermore, an independent external validation on a Dutch cohort of five pre-existing prediction models (all considering previous preterm deliveries in addition to other maternal characteristics) reported AUCs ranging from 0.54 to 0.67 [18].

4.3. Clinical Implications—The Meaning of the Study

During the variables' selection process, many predictors were identified. While socio-demographic conditions and the use of assisted medical conception techniques are known risk factors, interpreting the association between PTB and the other factors is more challenging. In fact, these identify the use of certain health services, but no causal relationship with the outcome was investigated. However, some of the selected variables are representative of conditions and risk factors studied extensively in the literature. The intake of pancreatic hormones (i.e., glucagon) and drugs used in diabetes, as well as the exemption for diabetes mellitus, identify people affected by diabetes, a condition that has been shown to be associated with PTB [19,20], while agents acting on the renin–angiotensin system, beta-blocking and calcium channel blockers are drugs commonly used for the treatment of hypertension [21], another strong predictor of preterm delivery [22]. Furthermore, both the transplant recipients and systemic lupus erythematosus exemptions identify condi-

tions known to be associated with a high preterm delivery rate [23–26]. There were very few women admitted to the hospital diagnosed with chronic renal failure; however, their increased risk of PTB is consistent with what was reported in a meta-analysis [27]. The most frequent component of PTBS is the intake of sex hormones and modulators of the genital system. This includes widely used hormonal contraceptives, some of which were associated with PTB in a previous study [28], as well as a variety of other hormonal drugs, that deserve careful investigation in future studies.

4.4. Strengths and Limitations

The present study has several strengths. First, the score was developed by monitoring a large cohort of women in a real-world setting. Second, the score performance was validated under different settings, varying in both temporal and geographical conditions. Third, by using electronic health data, the whole medical history of the mothers was taken into account without limiting the analysis to specific risk factors. Finally, the score can be used in combination with information on previous pregnancies and variables detected during gestation for a better PTB risk assessment.

The study also has a number of potential limitations. First, spontaneous deliveries could not be distinguished from programmed deliveries due to data unreliability. To overcome this limit, and attempt to restrict the evaluation to spontaneous deliveries only, the PTBS was validated in the Lombardy region using more restrictive definitions of preterm delivery (<36 and <32 weeks of gestation). In both cases, the discriminatory power was approximately equal to that observed in the main analysis (data not shown). Second, the analysis was restricted to nulliparous women. Albeit our strict inclusion criteria reduced the potential for confounding by including women in their first experience with the pregnancy, the generalizability of our findings to women with previous births requires extreme caution. However, because the history of preterm delivery is considered one of the most reliable criteria for identifying high-risk pregnancies for women with previous births, our score is a useful tool for women for whom there is greater uncertainty about pregnancy. Third, although we attempted to avoid false-positive signals by excluding variables with a frequency lower than 5 and applying the LASSO method, some high weights were assigned to very rare conditions/procedures in our population (e.g., heart failure, other operations on the rectum and perirectal tissue). Nevertheless, the good validation of the score in a different sample of the Lombardy region, as well as in other Italian regions, corroborates the goodness of the developed tool. Finally, the administrative purpose for which the healthcare utilization databases were instituted limits the completeness and accuracy of the medical information reported. For example, no data on smoking habits, BMI and lifestyle were available, and the severity of the comorbidities that lead women to use the services of the NHS cannot be evaluated.

4.5. Conclusions

In summary, a new predictive score, able to stratify nulliparous women according to their risk of preterm delivery, was developed and validated using data routinely collected in Italian healthcare utilization databases. Despite the limited discriminatory power, PTBS can be a useful tool for several healthcare professionals (e.g., general practitioner, obstetrician/gynecologist) to identify women at high risk of PTB early in pregnancy as well as for policymakers to guide health planning.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare10081443/s1>, Figure S1: Flow-chart of the cohort; Table S1: List of factors included in the Preterm Birth Score (PTBS), and corresponding codes; Table S2: Estimates of LASSO logistic regression coefficients for the 26 variables contributing to the Preterm Birth Score (PTBS); Table S3: Frequency of the 26 variables contributing to the Preterm Birth Score (PTBS) in the internal validation set (Lombardy); Table S4: Frequency of the 26 variables contributing to the Preterm Birth Score (PTBS) in Marche regio; Table S5: Frequency of the 26 variables contributing to the Preterm Birth Score (PTBS) in Sicily region.

Author Contributions: A.C., F.R. and G.C.: conceptualization; I.M., A.A. (Alessandra Allotta) and M.I.: data curation; I.M., A.A. (Alessandra Allotta), M.I.: formal analysis; G.C.: funding acquisition; E.E.T., M.P., A.A. (Antonella Agodi), A.L., R.Z., F.C., S.P.A., S.S. and G.C.: investigation; I.M., A.C., F.R. and G.C.: methodology; G.C.: project administration; F.C., S.S. and G.C.: supervision; E.E.T., M.P., A.A. (Antonella Agodi), A.L., R.Z., F.C., S.P.A., S.S. and G.C.: validation; E.E.T., M.P., A.A. (Antonella Agodi), A.L., R.Z., F.C., S.P.A., S.S. and G.C.: visualization; I.M., A.C., F.R. and G.C.: writing—original draft; A.A. (Antonella Agodi), E.E.T., M.I., M.P., A.A. (Antonella Agodi), A.L., R.Z., F.C., S.P.A. and S.S.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by grants from the Sicily Region [“Messa a punto di un modello di valutazione della qualità dei percorsi di gestione integrata per alcune condizioni di cronicità” project, grant number H45J19001550002], Italian Ministry of Health [“Modelli per il monitoraggio e la valutazione delle cure integrate (CI) nell’ambito del Nuovo Sistema di Garanzia dell’assistenza sanitaria” project, grant number J59H06000160001] and Italian Ministry of the Education, University and Research [“Fondo d’Ateneo per la Ricerca”, year 2020, grant number 2020-ATE-0541].

Institutional Review Board Statement: According to the rules from the Italian Medicines Agency (available at: http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf, accessed on 31 July 2022) retrospective studies without direct contact with patients do not need a written consent to process personal data when they are used for research aims.

Informed Consent Statement: We confirm that we had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data Availability Statement: The data that support the findings of this study are available from the Lombardy region, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the Lombardy region upon reasonable request.

Acknowledgments: This study was funded by grants from the Sicily Region [“Messa a punto di un modello di valutazione della qualità dei percorsi di gestione integrata per alcune condizioni di cronicità” project, grant number H45J19001550002], Italian Ministry of Health [“Modelli per il monitoraggio e la valutazione delle cure integrate (CI) nell’ambito del Nuovo Sistema di Garanzia dell’assistenza sanitaria” project, grant number J59H06000160001] and Italian Ministry of the Education, University and Research [“Fondo d’Ateneo per la Ricerca”, year 2020, grant number 2020-ATE-0541]. The Italian Ministry of the Education, University and Research and the Italian Ministry of Health had no role in the design of the study, the collection, the analysis, the interpretation of the data, or the decision to approve publication of the finished manuscript.

Conflicts of Interest: Giovanni Corrao received research support from the European Community (EC), the Italian Agency of Drug (AIFA), and the Italian Ministry of Education, University and Research (MIUR). He took part to a variety of projects that were funded by pharmaceutical companies (i.e., Novartis, GSK, Roche, AMGEN and BMS). He also received an honoraria as member of Advisory Board from Roche. Other authors declare that they have no conflicts of interest.

References

1. Lawn, J.E.; Gravett, M.G.; Nunes, T.M.; Rubens, C.E.; Stanton, C. GAPPS Review Group. Global report on preterm birth and stillbirth (1 of 7): Definitions, description of the burden and opportunities to improve data. *BMC Pregnancy Childbirth* **2010**, *10* (Suppl. S1), S1. [[CrossRef](#)]
2. Liu, L.; Oza, S.; Hogan, D.; Chu, Y.; Perin, J.; Zhu, J.; Lawn, J.E.; Cousens, S.; Mathers, C.; Black, R.E. Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* **2016**, *388*, 3027–3035. [[CrossRef](#)]
3. Lawn, J.E.; Cousens, S.; Zupan, J.; Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: When? Where? Why? *Lancet* **2005**, *365*, 891–900. [[CrossRef](#)]
4. Chehade, H.; Simeoni, U.; Guignard, J.P.; Boubred, F. Preterm Birth: Long Term Cardiovascular and Renal Consequences. *Curr. Pediatric Rev.* **2018**, *14*, 219–226. [[CrossRef](#)]
5. Moster, D.; Lie, R.T.; Markestad, T. Long-term medical and social consequences of preterm birth. *N. Engl. J. Med.* **2008**, *359*, 262–273. [[CrossRef](#)] [[PubMed](#)]

6. Blencowe, H.; Cousens, S.; Oestergaard, M.Z.; Chou, D.; Moller, A.-B.; Narwal, R.; Adler, A.; Garcia, C.V.; Rohde, S.; Say, L.; et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *Lancet* **2012**, *379*, 2162–2172. [[CrossRef](#)]
7. Goldenberg, R.L.; Culhane, J.F.; Iams, J.D.; Romero, R. Epidemiology and causes of preterm birth. *Lancet* **2008**, *371*, 75–84. [[CrossRef](#)]
8. Son, M.; Miller, E.S. Predicting preterm birth: Cervical length and fetal fibronectin. *Semin. Perinatol.* **2017**, *41*, 445–451. [[CrossRef](#)]
9. Jackson, R.A.; Gibson, K.A.; Wu, Y.W.; Croughan, M.S. Perinatal outcomes in singletons following in vitro fertilization: A meta-analysis. *Obstet. Gynecol.* **2004**, *103*, 551–563. [[CrossRef](#)]
10. Ferrero, D.M.; Larson, J.; Jacobsson, B.; Di Renzo, G.C.; Norman, J.E.; Martin, J.N., Jr.; Simpson, J.L. Cross-Country Individual Participant Analysis of 4.1 Million Singleton Births in 5 Countries with Very High Human Development Index Confirms Known Associations but Provides No Biologic Explanation for 2/3 of All Preterm Births. *PLoS ONE* **2016**, *11*, e0162506. [[CrossRef](#)]
11. Goldenberg, R.L.; Goepfert, A.R.; Ramsey, P.S. Biochemical markers for the prediction of preterm birth. *Am. J. Obstet. Gynecol.* **2005**, *192*, S36–S46. [[CrossRef](#)] [[PubMed](#)]
12. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Society. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
13. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-validation. *Encycl. Database Syst.* **2009**, *5*, 532–538.
14. Steyerberg, E.W.; Harrell, F.E., Jr. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol.* **2016**, *69*, 245–247. [[CrossRef](#)] [[PubMed](#)]
15. Janssen, K.J.; Vergouwe, Y.; Kalkman, C.J.; Grobbee, D.E.; Moons, K.G. A simple method to adjust clinical prediction models to local circumstances. *Can. J. Anaesth. J. Can. Danesthesie* **2009**, *56*, 194–201. [[CrossRef](#)] [[PubMed](#)]
16. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)] [[PubMed](#)]
17. Arabi Belaghi, R.; Beyene, J.; McDonald, S.D. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PLoS ONE* **2021**, *16*, e0252025. [[CrossRef](#)] [[PubMed](#)]
18. Meertens, L.; van Montfort, P.; Scheepers, H.; Van Kuijk, S.M.; Aardenburg, R.; Langenveld, J.; Van Dooren, I.M.; Zwaan, I.M.; Spaanderman, M.E.; Smits, L.J. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: A systematic review and independent external validation. *Acta Obstet. Gynecol. Scand.* **2018**, *97*, 907–920. [[CrossRef](#)] [[PubMed](#)]
19. Köck, K.; Köck, F.; Klein, K.; Bancher-Todesca, D.; Helmer, H. Diabetes mellitus and the risk of preterm birth with regard to the risk of spontaneous preterm birth. *J. Matern. Fetal Neonatal Med. Off. J. Eur. Assoc. Perinat. Med. Fed. Asia Ocean. Perinat. Soc. Int. Soc. Perinat. Obstet.* **2010**, *23*, 1004–1008. [[CrossRef](#)] [[PubMed](#)]
20. Wahabi, H.A.; Esmaeil, S.A.; Fayed, A.; Al-Shaikh, G.; Alzeidan, R.A. Pre-existing diabetes mellitus and adverse pregnancy outcomes. *BMC Res. Notes* **2012**, *5*, 496. [[CrossRef](#)] [[PubMed](#)]
21. Ram, C.V. Antihypertensive drugs: An overview. *Am. J. Cardiovasc. Drugs Drugs Devices Other Interv.* **2002**, *2*, 77–89. [[CrossRef](#)]
22. Bramham, K.; Parnell, B.; Nelson-Piercy, C.; Seed, P.T.; Poston, L.; Chappell, L.C. Chronic hypertension and pregnancy outcomes: Systematic review and meta-analysis. *BMJ* **2014**, *348*, g2301. [[CrossRef](#)] [[PubMed](#)]
23. Shah, S.; Venkatesan, R.L.; Gupta, A.; Sanghavi, M.K.; Welge, J.; Johansen, R.; Kean, E.; Kaur, T.; Gupta, A.; Grant, T.; et al. Pregnancy outcomes in women with kidney transplant: Metaanalysis and systematic review. *BMC Nephrol.* **2019**, *20*, 24. [[CrossRef](#)] [[PubMed](#)]
24. Marson, E.J.; Kamarajah, S.K.; Dyson, J.K.; White, S.A. Pregnancy outcomes in women with liver transplants: Systematic review and meta-analysis. *HPB Off. J. Int. Hepato Pancreato Biliary Assoc.* **2020**, *22*, 1102–1111. [[CrossRef](#)] [[PubMed](#)]
25. Acuna, S.; Zaffar, N.; Dong, S.; Ross, H.; D’Souza, R. Pregnancy outcomes in women with cardiothoracic transplants: A Systematic review and meta-analysis. *J. Heart Lung Transplant. Off. Publ. Int. Soc. Heart Transplant.* **2020**, *39*, 93–102. [[CrossRef](#)] [[PubMed](#)]
26. Wei, S.; Lai, K.; Yang, Z.; Zeng, K. Systemic lupus erythematosus and risk of preterm birth: A systematic review and meta-analysis of observational studies. *Lupus* **2017**, *26*, 563–571. [[CrossRef](#)]
27. Zhang, J.J.; Ma, X.X.; Hao, L.; Liu, L.J.; Lv, J.C.; Zhang, H. A Systematic Review and Meta-Analysis of Outcomes of Pregnancy in CKD and CKD Outcomes in Pregnancy. *Clin. J. Am. Soc. Nephrol. CJASN* **2015**, *10*, 1964–1978. [[CrossRef](#)]
28. Jensen, E.T.; Daniels, J.L.; Stürmer, T.; Robinson, W.; Williams, C.J.; Vejrup, K.; Magnus, P.; Longnecker, M.P. Hormonal contraceptive use before and after conception in relation to preterm birth and small for gestational age: An observational cohort study. *BJOG Int. J. Obstet. Gynaecol.* **2015**, *122*, 1349–1361. [[CrossRef](#)]