Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and internal validation of a multivariable logistic model for pulp disease in children from the central Peruvian Amazon: a secondary analysis of a case–control study

Jhair Alexander Leon-Rodriguez

jhair.leon@upeu.edu.pe

Peruvian Union University

Javier Linkolk López-Gonzales

Peruvian Union University

Research Article

Keywords: Dental pulp diseases, Dental caries, Anemia, Child, Logistic models, Risk factors, Health inequities, Peru

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11107861/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Pulp disease is a frequent and largely preventable consequence of untreated dental caries in children, yet its determinants are seldom modelled jointly in socioeconomically vulnerable settings, and probability models that could support screening in primary care are rarely developed or validated. We aimed to identify factors independently associated with pulp disease and to develop and internally validate a multivariable logistic model estimating its probability in children from the central Peruvian Amazon.

Methods

This was a secondary analysis of the individual-level dataset of a case–control study conducted in 2023 at the primary health care centre of Pichanaki (Chanchamayo, Junín, Peru), comprising 270 children aged 5 to 11 years: 90 cases with pulpal or periapical pathology diagnosed on clinical and radiographic grounds and 180 controls without pulpal involvement. Biological (anemia, hemoglobin concentration, nutritional status), behavioural (fluoride toothpaste, mouthwash and dental floss use, ferrous sulfate intake) and socioeconomic variables (maternal educational level and occupation, maternal age, household income) were retrieved from a previously collected and validated database. Variables with $p < 0.20$ in bivariable analysis, together with clinically relevant covariates, were entered into a multivariable binary logistic regression model; multicollinearity was assessed with the variance inflation factor. Discrimination was quantified with the area under the receiver operating characteristic curve (AUC), calibration with the Hosmer–Lemeshow test and a calibration curve, and internal validity with a 70/30 training–test split and k-fold cross-validation. The optimal classification threshold was derived from the Youden index, and performance was re-estimated across alternative thresholds.

Results

Dental caries (adjusted odds ratio [aOR] 12.08, 95% confidence interval [CI] 3.85–44.62), ferrous sulfate intake (aOR 11.63, 95% CI 3.85–40.76), anemia (aOR 5.73, 95% CI 1.99–18.14) and non-professional maternal educational level (aOR 3.83, 95% CI 1.36–11.82) remained independently associated with higher odds of pulp disease, whereas fluoride toothpaste use (aOR 0.09, 95% CI 0.03–0.25), dental floss use (aOR 0.13, 95% CI 0.04–0.36) and mouthwash use (aOR 0.30, 95% CI 0.10–0.83) were inversely associated. Maternal age below 20 years was no longer significant after adjustment (aOR 2.34, 95% CI 0.73–7.81). The final model was adequately calibrated (Hosmer–Lemeshow $\chi^2 = 6.11$; $p = 0.635$; Nagelkerke $R^2 = 0.817$) and discriminated well (AUC 0.974, 95% CI 0.958–0.990), with an accuracy of 93.3%, sensitivity of 90.0% and specificity of 95.0% at the 0.50 threshold. Internal validation yielded an AUC of 0.975 in the 70/30 split and 0.966 in cross-validation. At the Youden-derived threshold of 0.31, sensitivity reached 96.7% and specificity 92.2%.

Conclusions

In this vulnerable paediatric population, pulp disease was explained by a combination of caries experience, anemia-related indicators, maternal educational level and everyday preventive behaviours, and these routinely available variables separated affected from unaffected children with high accuracy. Because predictors were recorded concurrently with the outcome within a case–control sample, the model should be regarded as a screening and prioritisation aid rather than a causal or population-calibrated risk tool; external and prospective validation is required before clinical deployment.

Background

Oral diseases affect approximately 3.5 billion people worldwide and constitute one of the most widespread yet most neglected groups of noncommunicable conditions [1]. Untreated caries in deciduous and permanent teeth remains the single most prevalent condition in global burden estimates, and the burden is distributed along a steep social gradient: it concentrates in low- and middle-income countries and, within them, in the households with the fewest material and educational resources [2, 3]. In children, dental disease is not confined to the mouth. Pain, infection and premature tooth loss compromise chewing and nutrition, disturb sleep, provoke school absence and measurably degrade quality of life [4].

Pulp disease represents the clinical point at which untreated caries stops being a localised lesion of hard tissue and becomes an infectious and inflammatory process with systemic potential. Once bacterial invasion reaches the pulp, the outcome is pain, pulpal necrosis, periapical involvement and, frequently, extraction [5, 6]. The magnitude of this transition in disadvantaged populations is well documented: indices developed specifically to capture the consequences of untreated decay, such as the pufa/PUFA index, consistently reveal that a substantial proportion of carious lesions in children with limited access to restorative care progresses to pulpal involvement, ulceration, fistula or abscess [7]. In settings where the dental workforce is concentrated in urban centres, extraction under emergency conditions often becomes the only realistic treatment option, and prevention loses ground to symptomatic relief [8, 9].

The determinants of this progression are multifactorial. Dental caries is the proximal cause, but the probability that a lesion advances to the pulp is modulated by biological, behavioural and structural conditions. Iron deficiency anemia and undernutrition have been associated with oral disease through impaired oxygen delivery to tissues, altered immune competence, reduced salivary flow and diminished buffering capacity [10, 11]. Recent quantitative syntheses support this association in children: a systematic review and meta-analysis of iron deficiency anemia and dental caries [12] and a subsequent umbrella meta-analysis of five previous meta-analyses, encompassing roughly 19,400 participants, reported that children with iron deficiency anemia were more than three times as likely to present dental caries (odds ratio 3.64, 95% CI 2.45–5.40) [13]. Behavioural exposures, particularly the daily use of fluoride toothpaste, and socioeconomic conditions such as maternal education and household income

exert an equally decisive influence, and their effect is largely mediated by the family's capacity to purchase preventive products and to reach dental services [14, 15].

The central Peruvian Amazon concentrates these vulnerabilities. In the province of Chanchamayo, and specifically in the district of Pichanaki, high rates of poverty and childhood anemia coexist with fragmented access to oral health care, and a national supplementation programme distributes ferrous sulfate through primary care facilities [16, 17, 18]. Evidence on pulp disease in this population, however, remains scarce and fragmented. A previous case–control study conducted by our group in this setting established iron deficiency anemia as a risk factor for pulp disease [19], but, like most studies in comparable contexts, it examined exposures largely one at a time and did not quantify how much of the outcome can be explained when biological, behavioural and socioeconomic variables are considered simultaneously.

Multivariable logistic regression addresses precisely this gap: it estimates the independent contribution of each predictor while adjusting for the remainder, and it converts the resulting coefficients into an individual probability of disease [20]. Prediction models built on variables that are already collected in routine care are particularly attractive in resource-limited settings, because they require no additional laboratory or imaging investment and can be used to prioritise children for dental examination. Such models are increasingly reported in oral health — for example, a recent model for severe caries in 12-year-olds from resource-limited regions of China achieved an AUC of 0.711 and was proposed as a preliminary screening instrument [21] — yet they remain almost absent from the Latin American literature on pulpal pathology, and when developed they are seldom evaluated for calibration or subjected to any form of internal validation, as current reporting standards require [22].

Against this background, the present study had two objectives: to identify the biological, behavioural and socioeconomic factors independently associated with pulp disease in children from the central Peruvian Amazon, and to develop and internally validate a multivariable logistic model that estimates the probability of pulp disease from variables routinely available in primary care.

Methods

Study design and reporting

This is a secondary analysis of the individual-level dataset of an observational, analytical case–control study carried out in 2023 in the central Peruvian Amazon, the primary objective of which was to test iron deficiency anemia as a risk factor for pulp disease and which has been reported elsewhere [19]. Whereas the primary report evaluated exposures separately, the present analysis pursues a distinct objective: the simultaneous modelling of biological, nutritional, behavioural and socioeconomic variables, and the development, calibration and internal validation of a probability model for pulp disease. No new participants were recruited and no additional data were collected for this analysis.

The study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [23]; the items specific to the development and validation of a prediction model follow the TRIPOD + AI statement [22].

Setting and participants

The source population comprised children aged 5 to 11 years who attended the dental service of the Primary Health Care Centre of Pichanaki, province of Chanchamayo, department of Junín, Peru. The district is representative of the central Peruvian Amazon and is characterised by high levels of socioeconomic vulnerability, elevated prevalence of childhood anemia and limited availability of restorative dental care [16].

The dataset comprised 270 participants: 90 cases and 180 controls, in a 1:2 ratio. Cases were children diagnosed with pulpal or periapical pathology who had also undergone hemoglobin screening. Controls were children without pulpal involvement, selected from the same population and the same period in order to preserve comparability and temporal homogeneity. The sample size corresponded to that of the original case–control study, which was planned to provide at least ten events per candidate variable ($n = 10 \times [k + 1]$), a conventional criterion intended to limit overfitting and to stabilise the estimation of regression coefficients.

Outcome definition

The outcome was the presence of pulp disease, defined as pulpal or periapical pathology diagnosed on combined clinical and radiographic grounds following the diagnostic terminology of the American Association of Endodontists and the American Board of Endodontics [24]. Examinations in the original study were performed with a standardised odontogram in accordance with the technical norm of the Peruvian Ministry of Health [25]. Examiner calibration, assessed with Cohen's kappa coefficient, reached 1.00 for pulp disease and 0.95 for dental caries, indicating almost perfect agreement.

Candidate predictors

Candidate predictors were selected a priori on the basis of biological plausibility and previous evidence, and were grouped into three domains. Biological variables comprised hemoglobin concentration, anemia, chronic undernutrition and acute undernutrition. Hemoglobin values were retrieved from clinical records, and anemia was defined as a concentration below 11.5 g/dL according to the national technical norm [17]; no altitude adjustment was applied given the low elevation of the study area. Nutritional status was classified from anthropometric measurements following national guidelines [18]. Dental caries experience was recorded with the DMFT/dmft index, which quantifies caries experience in permanent and primary dentition, respectively [26], and was analysed as a dichotomous variable (presence versus absence).

Behavioural variables comprised the use of fluoride toothpaste, mouthwash and dental floss, and the intake of ferrous sulfate; the latter is distributed in Peru through the national programme for the prevention and control of childhood anemia and is therefore administered predominantly to children already identified as anemic [17, 18]. Socioeconomic variables comprised maternal educational level (professional versus non-professional), maternal occupation (professional versus homemaker), maternal age (≥ 20 versus < 20 years) and monthly household income, dichotomised at the national minimum wage in force at the time of data collection (PEN 1,050). Behavioural and socioeconomic data were obtained through structured questionnaires administered to parents or caregivers.

Statistical analysis

All analyses were conducted in R (R Foundation for Statistical Computing, Vienna, Austria). Data cleaning and validation procedures were applied before analysis to verify consistency and completeness. Categorical variables were summarised as absolute and relative frequencies and continuous variables as measures of central tendency and dispersion.

Bivariable associations between each candidate variable and pulp disease were examined with Pearson's chi-square test or Fisher's exact test for categorical variables and with the Wilcoxon rank-sum test for continuous variables. Variables reaching $p < 0.20$ in bivariable analysis, together with covariates considered clinically relevant, were entered into a multivariable binary logistic regression model. Multicollinearity among covariates was evaluated with the variance inflation factor, and variables that introduced redundancy in the linear predictor were not retained in the final model. Associations are reported as crude and adjusted odds ratios with 95% confidence intervals.

Model performance was assessed in the three dimensions recommended for prediction models. Overall fit was examined with the omnibus likelihood-ratio test and the Nagelkerke R^2. Discrimination was quantified with the area under the receiver operating characteristic curve and its 95% confidence interval. Calibration was evaluated with the Hosmer–Lemeshow goodness-of-fit test and by graphical inspection of a calibration curve plotting observed against predicted probabilities across deciles of risk [27]. Classification performance was described through accuracy, sensitivity, specificity, positive and negative predictive values, F1-score and Cohen's kappa coefficient.

Internal validity was examined by two complementary strategies: a 70/30 training–test split and k-fold cross-validation. Finally, because the classification threshold determines the operational behaviour of any probability model, performance was re-estimated across a range of cut-off points, and the optimal threshold was identified with the Youden index [28].

Ethical considerations

This study is a secondary analysis of a previously conducted investigation approved by the Ethics Committee of Antenor Orrego Private University (Resolution No. 0762-2023-UPAO) [19]. In the original study, written informed consent was obtained from the parents or legal guardians of all participating

children, and confidentiality of participant information was guaranteed. The present analysis adhered to the ethical principles of the Declaration of Helsinki [29], the Belmont Report [30] and the international ethical guidelines of the Council for International Organizations of Medical Sciences [31]. All records were fully anonymised before analysis, and no identifying information was accessible to the authors of this secondary analysis; consequently, no additional ethical approval was required.

Results

Characteristics of participants

Children with pulp disease had significantly lower hemoglobin concentrations than children without pulp disease (9.94 ± 1.48 versus 11.09 ± 1.25 g/dL; $p < 0.001$). The group with pulp disease also showed a higher frequency of anemia (82% versus 38%), ferrous sulfate intake (88% versus 29%), non-professional maternal educational level (56% versus 25%), maternal age below 20 years (44% versus 11%), low household income (60% versus 29%) and dental caries (78% versus 35%); all differences were statistically significant ($p < 0.001$). Conversely, the use of fluoride toothpaste (17% versus 79%), mouthwash (34% versus 67%) and dental floss (17% versus 73%) was considerably less frequent among children with pulp disease (all $p < 0.001$). No significant differences were observed for maternal occupation ($p = 0.116$), chronic undernutrition ($p = 0.532$) or acute undernutrition ($p = 0.561$) (Table 1).

Table 1
Clinical, behavioural and sociodemographic characteristics according to pulp disease status

Variable	Overall (N = 270)	No pulp disease (n = 180)	Pulp disease (n = 90)	p-value
Hemoglobin, g/dL, mean ± SD	10.70 ± 1.43	11.09 ± 1.25	9.94 ± 1.48	< 0.001[a]
Anemia, n (%)				< 0.001[b]
No	127 (47)	111 (62)	16 (18)	
Yes	143 (53)	69 (38)	74 (82)	
Ferrous sulfate intake, n (%)				< 0.001[b]
Did not receive	139 (51)	128 (71)	11 (12)	
Received	131 (49)	52 (29)	79 (88)	
Maternal educational level, n (%)				< 0.001[b]
Professional	175 (65)	135 (75)	40 (44)	
Non-professional	95 (35)	45 (25)	50 (56)	
Maternal age, n (%)				< 0.001[b]
≥ 20 years	210 (78)	160 (89)	50 (56)	
< 20 years	60 (22)	20 (11)	40 (44)	
Maternal occupation, n (%)				0.116[b]
Professional	184 (68)	117 (65)	67 (74)	
Homemaker	86 (32)	63 (35)	23 (26)	
Household income, n (%)				< 0.001[b]
≥ PEN 1,050	163 (60)	127 (71)	36 (40)	
< PEN 1,050	107 (40)	53 (29)	54 (60)	
Chronic undernutrition, n (%)				0.532[b]
No	233 (86)	157 (87)	76 (84)	

SD standard deviation, *PEN* Peruvian sol. Continuous variables are presented as mean ± standard deviation and categorical variables as n (%). [a]Wilcoxon rank-sum test; [b]Pearson's chi-square test; [c]Fisher's exact test.

Variable	Overall (N = 270)	No pulp disease (n = 180)	Pulp disease (n = 90)	p-value
Yes	37 (14)	23 (13)	14 (16)	
Acute undernutrition, n (%)				0.561[c]
No	256 (95)	172 (96)	84 (93)	
Yes	14 (5)	8 (4)	6 (7)	
Dental caries, n (%)				< 0.001[b]
No	137 (51)	117 (65)	20 (22)	
Yes	133 (49)	63 (35)	70 (78)	
Fluoride toothpaste use, n (%)				< 0.001[b]
No	112 (41)	37 (21)	75 (83)	
Yes	158 (59)	143 (79)	15 (17)	
Mouthwash use, n (%)				< 0.001[b]
No	118 (44)	59 (33)	59 (66)	
Yes	152 (56)	121 (67)	31 (34)	
Dental floss use, n (%)				< 0.001[b]
No	124 (46)	49 (27)	75 (83)	
Yes	146 (54)	131 (73)	15 (17)	

SD standard deviation, *PEN* Peruvian sol. Continuous variables are presented as mean ± standard deviation and categorical variables as n (%). [a]Wilcoxon rank-sum test; [b]Pearson's chi-square test; [c]Fisher's exact test.

Factors associated with pulp disease

In univariable analysis, anemia (odds ratio [OR] 7.44, 95% CI 4.10–14.20), ferrous sulfate intake (OR 17.68, 95% CI 9.03–37.63), non-professional maternal educational level (OR 3.75, 95% CI 2.21–6.45), maternal age below 20 years (OR 6.40, 95% CI 3.47–12.14), low household income (OR 3.59, 95% CI 2.13–6.15) and dental caries (OR 6.50, 95% CI 3.68–11.89) were associated with increased odds of pulp disease (all $p < 0.001$), whereas the use of fluoride toothpaste (OR 0.05, 95% CI 0.03–0.10), mouthwash (OR 0.26, 95% CI 0.15–0.43) and dental floss (OR 0.08, 95% CI 0.04–0.14) was associated with reduced odds (all $p < 0.001$).

In the multivariable model, ferrous sulfate intake (adjusted OR [aOR] 11.63, 95% CI 3.85–40.76), anemia (aOR 5.73, 95% CI 1.99–18.14), dental caries (aOR 12.08, 95% CI 3.85–44.62) and non-professional maternal educational level (aOR 3.83, 95% CI 1.36–11.82) remained independently associated with higher odds of pulp disease. The use of fluoride toothpaste (aOR 0.09, 95% CI 0.03–0.25), mouthwash (aOR 0.30, 95% CI 0.10–0.83) and dental floss (aOR 0.13, 95% CI 0.04–0.36) remained independently protective. Maternal age below 20 years lost statistical significance after adjustment (aOR 2.34, 95% CI 0.73–7.81; p = 0.157), and household income was not retained in the final model (Table 2).

Table 2
Univariable and multivariable logistic regression analysis for pulp disease

Variable	Crude OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
Anemia	7.44 (4.10–14.20)	< 0.001	5.73 (1.99–18.14)	0.002
Ferrous sulfate intake	17.68 (9.03–37.63)	< 0.001	11.63 (3.85–40.76)	< 0.001
Non-professional maternal educational level	3.75 (2.21–6.45)	< 0.001	3.83 (1.36–11.82)	0.014
Maternal age < 20 years	6.40 (3.47–12.14)	< 0.001	2.34 (0.73–7.81)	0.157
Household income < PEN 1,050	3.59 (2.13–6.15)	< 0.001	—	—
Dental caries	6.50 (3.68–11.89)	< 0.001	12.08 (3.85–44.62)	< 0.001
Fluoride toothpaste use	0.05 (0.03–0.10)	< 0.001	0.09 (0.03–0.25)	< 0.001
Mouthwash use	0.26 (0.15–0.43)	< 0.001	0.30 (0.10–0.83)	0.023
Dental floss use	0.08 (0.04–0.14)	< 0.001	0.13 (0.04–0.36)	< 0.001

OR odds ratio, *CI* confidence interval, *PEN* Peruvian sol. Crude odds ratios were obtained from univariable binary logistic regression models; adjusted odds ratios were obtained from the final multivariable binary logistic regression model. Variables with p < 0.20 in bivariable analysis were eligible for the initial multivariable model. Household income was not retained in the final model.

Model performance and internal validation

The final model was globally significant according to the omnibus test (p < 0.001) and showed adequate calibration, with a non-significant Hosmer–Lemeshow statistic ($\chi^2 = 6.11$; p = 0.635). Explanatory capacity was high, with a Nagelkerke R^2 of 0.817.

Discrimination was excellent, with an AUC of 0.974 (95% CI 0.958–0.990). At the conventional 0.50 threshold the model achieved an accuracy of 93.3%, sensitivity of 90.0%, specificity of 95.0%, positive predictive value of 90.0%, negative predictive value of 95.0% and an F1-score of 90.0%; agreement between predicted and observed classifications was strong (Cohen's kappa 0.850). Internal validation supported the stability of these estimates: the 70/30 training–test split yielded an AUC of 0.975 and an accuracy of 90.1%, and k-fold cross-validation produced an ROC value of 0.966 with a sensitivity of 80.0% and a specificity of 93.3% (Table 3).

Table 3
Performance and internal validation metrics of the final logistic regression model

Metric	**Value**
Omnibus test, p-value	< 0.001
Hosmer–Lemeshow χ^2	6.11
Hosmer–Lemeshow, p-value	0.635
Nagelkerke R^2	0.817
AUC	0.974
95% CI for AUC	0.958–0.990
Accuracy	0.933
Sensitivity	0.900
Specificity	0.950
Positive predictive value	0.900
Negative predictive value	0.950
F1-score	0.900
Cohen's kappa coefficient	0.850
Internal validation AUC (70/30 split)	0.975
Internal validation accuracy (70/30 split)	0.901
k-fold cross-validation ROC	0.966
k-fold cross-validation sensitivity	0.800
k-fold cross-validation specificity	0.933

AUC area under the receiver operating characteristic curve, *CI* confidence interval, *ROC* receiver operating characteristic. Classification metrics correspond to the 0.50 probability threshold. Calibration was assessed with the Hosmer–Lemeshow goodness-of-fit test and discrimination with ROC curve analysis. Predictive values depend on the prevalence of pulp disease in the analysed sample.

The receiver operating characteristic curve confirmed the high capacity of the model to distinguish children with pulp disease from those without it, and the narrow confidence interval around the AUC indicates a stable estimate (Fig. 1). The calibration curve showed close agreement between predicted probabilities and observed frequencies across deciles of risk, with most points lying near the diagonal of perfect calibration; minor deviations appeared at the lower end of the predicted probability range, while performance remained consistent over the rest of the spectrum (Fig. 2).

Threshold sensitivity analysis

Performance varied predictably with the classification threshold: lower cut-off points increased sensitivity at the expense of specificity, whereas higher cut-off points had the opposite effect. The threshold identified by the Youden index (0.31) offered the best balance, with a sensitivity of 96.7%, a specificity of 92.2%, an overall accuracy of 93.7% and an F1-score of 0.911. In contrast, a threshold of 0.80 maximised specificity (98.3%) but reduced sensitivity to 70.0% (Table 4).

Table 4
Sensitivity analysis of model performance across classification thresholds

Threshold	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score
0.30	0.933	0.967	0.917	0.853	0.982	0.906
0.31 (Youden index)	0.937	0.967	0.922	0.861	0.982	0.911
0.50	0.933	0.900	0.950	0.900	0.950	0.900
0.80	0.889	0.700	0.983	0.955	0.868	0.808

PPV positive predictive value, *NPV* negative predictive value. The threshold of 0.31 was selected with the Youden index as the cut-off maximising the sum of sensitivity and specificity of the final multivariable logistic regression model.

Discussion

This secondary analysis of a case–control study conducted in the central Peruvian Amazon shows that pulp disease in children aged 5 to 11 years is explained by a small set of variables spanning three domains: caries experience, anemia-related indicators, maternal educational level and everyday preventive behaviours. Modelled jointly, these variables separated affected from unaffected children with an AUC of 0.974 and retained comparable performance under two independent internal validation strategies. To our knowledge, this is the first probability model for pulpal pathology developed in an Amazonian paediatric population, and it uses only variables that Peruvian primary care already records.

Dental caries showed the strongest adjusted association with the outcome (aOR 12.08). This is biologically coherent, since pulp disease is the natural endpoint of a cavitated lesion that is not arrested or restored, and it is consistent with the evidence generated by indices designed to capture the clinical

consequences of untreated decay, which repeatedly document high proportions of pulpal involvement among children with limited access to restorative care [7, 8]. The finding nonetheless requires an interpretive caution that we consider essential: caries lies on the causal pathway leading to pulp disease rather than upstream of it. Its inclusion maximises discrimination and is appropriate for a model intended to identify affected children, but the corresponding odds ratio should not be read as the effect of an independent aetiological factor.

Anemia remained independently associated with pulp disease after adjustment (aOR 5.73), in agreement with the direction and approximate magnitude reported by recent quantitative syntheses of anemia and oral disease in children, including a systematic review and meta-analysis [12] and an umbrella meta-analysis of five previous meta-analyses that estimated a pooled odds ratio of 3.64 (95% CI 2.45–5.40) for the association between iron deficiency anemia and dental caries [13]. Several mechanisms have been proposed, among them reduced oxygen delivery to oral tissues, impaired neutrophil and lymphocyte function, and decreased salivary flow with loss of buffering and remineralising capacity [10, 11]. Two caveats apply. First, anemia and oral disease share upstream determinants — poverty, food insecurity and diets dominated by inexpensive fermentable carbohydrates — so part of the association may reflect common causes rather than a direct biological effect. Second, the relationship is plausibly bidirectional: dental pain and chronic oral infection limit the intake of iron-rich foods and may contribute to anemia of inflammation. A case–control design with concurrent measurement cannot arbitrate between these explanations.

The association observed for ferrous sulfate intake (aOR 11.63) deserves particular emphasis, because its apparent direction is counterintuitive and could be seriously misread. In Peru, ferrous sulfate is not distributed at random: it is delivered through the national programme for the prevention and control of childhood anemia, essentially to children who have already been screened and identified as anemic [17, 18]. Intake of the supplement therefore functions as a marker of previously diagnosed anemia and of contact with the health system, not as an exposure that competes on equal terms with the remaining covariates. This is a textbook example of confounding by indication combined with reverse temporality [32], and it is reinforced in our data by the strong correlation between supplementation and anemia (88% of cases received ferrous sulfate and 82% were anemic). We therefore interpret this variable as a proxy for the severity and duration of iron deficiency, and we explicitly warn against any causal reading: nothing in these data suggests that iron supplementation harms pulp vitality, and these findings must not be used to question a public health intervention of demonstrated value. Its retention in the model is justified on predictive, not aetiological, grounds — it is a readily available flag that identifies children who warrant dental examination.

Non-professional maternal educational level remained independently associated with pulp disease (aOR 3.83), while household income, significant in univariable analysis, was not retained in the final model, most plausibly because both variables capture overlapping dimensions of the same socioeconomic position. The finding is consistent with the meta-analytic evidence on socioeconomic inequality and caries, which documents a graded relationship between indicators of social position and caries

experience across age groups and settings [14]. Maternal education operates through several channels that are particularly relevant in the Amazonian context: health literacy and recognition of early signs of disease, the ability to establish and supervise daily oral hygiene routines, the capacity to purchase fluoride toothpaste regularly, and the time and transport resources needed to attend a dental appointment in a district where care is concentrated in a single facility [15]. Maternal age below 20 years, strongly associated in univariable analysis, lost significance after adjustment, which suggests that its apparent effect was largely mediated by maternal education and household income rather than being independent.

Consistent inverse associations were observed for the three preventive behaviours examined. The magnitude estimated for fluoride toothpaste (aOR 0.09) is directionally coherent with the causal evidence from randomised trials synthesised in Cochrane reviews of fluoride toothpaste [33] and fluoride mouthrinses [34], although our estimate is considerably larger than the effect sizes reported in those syntheses. We interpret this difference cautiously. Reported use of toothpaste, mouthwash and floss was obtained from caregivers after the child's diagnosis and is therefore susceptible to recall and social desirability bias; in addition, these behaviours are unlikely to act in isolation, and they plausibly operate as markers of a broader family oral health culture and of the material capacity to sustain it. Our design cannot separate the pharmacological effect of fluoride from the socioeconomic and behavioural context in which its use is embedded.

The discriminative performance of the model (AUC 0.974) is substantially higher than that reported by comparable models in paediatric oral health; a recent prediction model for severe caries developed in 2,437 Chinese 12-year-olds from resource-limited regions achieved an AUC of 0.711 in the training set and 0.701 in the validation set [21]. We regard this contrast as informative rather than favourable, and three features of our study account for most of it. First, a case–control sample compares two groups selected precisely because they differ in the outcome, which yields greater separation of the covariate distributions than a population-based cohort would produce. Second, the model incorporates predictors that are proximal to, and in the case of caries partly constitutive of, the outcome. Third, most performance metrics were estimated in the same data used for model development, and the 70/30 split reuses that sample rather than providing an independent test. An AUC approaching 0.97 in field epidemiology should therefore prompt caution rather than confidence, and we present these figures as internal performance, not as an expectation of what the model would achieve in routine practice.

The same reasoning applies to calibration. The non-significant Hosmer–Lemeshow statistic and the proximity of the calibration curve to the diagonal indicate agreement between predicted and observed frequencies within this sample, but the predicted probabilities are conditional on the 1:2 case–control sampling ratio and consequently do not represent absolute risks in the source population; the same limitation affects the positive and negative predictive values, which depend on prevalence. Before any field use, the model intercept would need to be recalibrated to the local prevalence of pulp disease, a step that is well described for models developed under non-representative sampling [27]. With that caveat, the threshold analysis has an immediate operational reading. At the Youden-derived cut-off of

0.31 the model behaves as a screening instrument (sensitivity 96.7%, specificity 92.2%), which is the appropriate posture where the cost of overlooking a child with pulpal involvement — pain, infection and eventual extraction — exceeds the cost of an unnecessary dental examination. Where treatment capacity is the binding constraint, a higher threshold such as 0.80 maximises specificity (98.3%) and concentrates scarce appointments on the children most likely to require them.

These findings have practical implications for primary care in the Peruvian Amazon. All variables in the model — caries status from the standardised odontogram, hemoglobin from routine screening, supplementation from programme records, and maternal education from the intake interview — are already collected during ordinary paediatric visits, so the model imposes no additional cost, equipment or laboratory requirement. Its most immediate application is bidirectional: children attending anemia screening or receiving ferrous sulfate could be systematically referred for dental examination, and children presenting with caries could be screened for anemia. This argues for the explicit integration of oral health into the national strategy for the reduction of childhood anemia, two programmes that currently operate in parallel within the same facilities and the same population.

Several strengths support the internal validity of this analysis. The outcome was diagnosed with combined clinical and radiographic criteria by examiners with almost perfect calibration (kappa 1.00 for pulp disease and 0.95 for caries), hemoglobin was obtained from laboratory records rather than self-report, and dental examination followed a standardised national odontogram. Analytically, collinearity was formally assessed, performance was characterised across the three dimensions of fit, discrimination and calibration rather than by discrimination alone, two complementary internal validation strategies were applied, and the behaviour of the model was examined across the full range of classification thresholds. Finally, the study addresses a population that is almost absent from the international literature on paediatric oral health.

The limitations must be weighed with equal candour. First, the design is a secondary analysis of a case–control study with concurrent measurement of predictors and outcome; temporality cannot be established, and no association reported here should be interpreted causally. Second, case–control sampling means that predicted probabilities are not population-calibrated, as discussed above. Third, participants were recruited non-probabilistically among children who sought care at a single primary health care centre, which introduces selection towards symptomatic children and limits generalisability to the wider district and region. Fourth, behavioural and socioeconomic variables were reported by caregivers and are subject to recall and social desirability bias. Fifth, with 90 events and nine candidate variables, the analysis meets the classical criterion of ten events per variable but not the more demanding sample size criteria currently recommended for the development of prediction models [35]; the width of several confidence intervals, notably for caries and ferrous sulfate, reflects this constraint. Sixth, internal validation relied on a single 70/30 split, a strategy known to be statistically inefficient and unstable in samples of this size [36], and optimism-corrected estimates through bootstrap resampling were not obtained. Seventh, clinical usefulness was not formally quantified through decision curve analysis [37]. Finally, several plausible determinants were not available in the original dataset, including

dietary sugar frequency, fluoride concentration in the water supply, previous utilisation of dental services and parental oral health status; residual confounding by these factors cannot be excluded.

Future work should proceed in three directions. The model requires external validation in independent primary care facilities of Junín and other Amazonian regions, followed by recalibration to local prevalence. A prospective cohort design, in which candidate predictors are measured before the onset of pulpal involvement, would allow the temporal ordering that the present design cannot establish, and would permit the aetiological and the predictive questions to be separated. Finally, the addition of bootstrap-based optimism correction, decision curve analysis and a graphical tool such as a nomogram would make the model directly usable by non-specialist personnel in the facilities where these children are actually seen.

Conclusions

In children aged 5 to 11 years from the central Peruvian Amazon, pulp disease was independently associated with dental caries, anemia, ferrous sulfate intake and non-professional maternal educational level, and inversely associated with the use of fluoride toothpaste, dental floss and mouthwash. A multivariable logistic model built exclusively from variables available in routine primary care discriminated affected from unaffected children with high accuracy and remained stable under internal validation. Because the model was developed in a case–control sample with concurrent measurement of predictors and outcome, it should be understood as a screening and prioritisation aid rather than as a causal explanation or a population-calibrated risk calculator. Its main value is programmatic: it identifies anemia screening and iron supplementation as practical entry points for the early detection of pulpal disease in children who would otherwise reach the dental service only when pain has already made extraction inevitable.

Abbreviations

aOR
adjusted odds ratio
AUC
area under the receiver operating characteristic curve
CI
confidence interval
DMFT/dmft
decayed, missing and filled teeth index for permanent and primary dentition
NPV
negative predictive value
OR
odds ratio
PEN

Peruvian sol
PPV
positive predictive value
ROC
receiver operating characteristic
SD
standard deviation
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD
Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
VIF
variance inflation factor.

Declarations

Ethics approval and consent to participate

This study is a secondary analysis of anonymised data from a case–control study approved by the Ethics Committee of Antenor Orrego Private University, Trujillo, Peru (Resolution No. 0762-2023-UPAO). Written informed consent was obtained from the parents or legal guardians of all participating children in the original study, and assent was obtained from the children according to national requirements. All procedures were performed in accordance with the Declaration of Helsinki, the Belmont Report and the international ethical guidelines of the Council for International Organizations of Medical Sciences. Because the present analysis used a fully anonymised dataset and involved no additional contact with participants, no further ethical approval was required.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available from the corresponding author on reasonable request. The primary results of the parent case–control study from which these data derive have been published previously and are cited in this manuscript [19].

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgements

The authors thank the staff of the Primary Health Care Centre of Pichanaki, Chanchamayo, Junín, and the families who participated in the original study.

Authors' contributions

JALR conceived the study, curated the dataset, performed the statistical analysis and drafted the manuscript. JLLG supervised the statistical modelling and the internal validation strategy, contributed to the interpretation of the results and critically revised the manuscript for important intellectual content. Both authors read and approved the final manuscript.

Conceptualization: J.A.L.R. and J.L.L.G. Data curation: J.A.L.R. and J.L.L.G. Formal analysis: J.A.L.R. and J.L.L.G. Funding acquisition: not applicable; this research received no external funding. Investigation: J.A.L.R. and J.L.L.G. Methodology: J.A.L.R. and J.L.L.G. Project administration: J.A.L.R. and J.L.L.G. Resources: J.A.L.R. and J.L.L.G. Software: J.A.L.R. and J.L.L.G. Supervision: J.L.L.G. Validation: J.A.L.R. and J.L.L.G. Visualization: J.A.L.R. and J.L.L.G. Writing – original draft: J.A.L.R. and J.L.L.G. Writing – review and editing: J.A.L.R. and J.L.L.G. All authors reviewed the manuscript and approved the submitted version.

Acknowledgements

The authors thank the staff of the Primary Health Care Centre of Pichanaki, Chanchamayo, Junín, and the families who participated in the original study.

References

1. Peres MA, Macpherson LMD, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. Lancet. 2019;394(10194):249–60.
2. Wen PYF, Chen MX, Zhong YJ, Dong QQ, Wong HM. Global burden and inequality of dental caries, 1990 to 2019. J Dent Res. 2022;101(4):392–9.
3. World Health Organization. Global oral health status report: towards universal health coverage for oral health by 2030. Geneva: World Health Organization; 2022.
4. Klarić Puđa I, Goršeta K, Jurić H, Soldo M, Marks LAM, Majstorović M. A cohort study on the impact of oral health on the quality of life of adolescents and young adults. Clin Pract. 2025;15(4):76.
5. Maldupa I, Al-Yaseen W, Giese J, Ahmed Elagami R, Raggio DP. Recommended procedures for managing carious lesions in primary teeth with pulp involvement: a scoping review. BDJ Open. 2024;10:74.

6. Lin LM, Ricucci D, Saoud TM, Sigurdsson A, Kahler B. Vital pulp therapy of mature permanent teeth with irreversible pulpitis from the perspective of pulp biology. Aust Endod J. 2020;46(1):154–66.
7. Monse B, Heinrich-Weltzien R, Benzian H, Holmgren C, van Palenstein Helderman W. PUFA – an index of clinical consequences of untreated dental caries. Community Dent Oral Epidemiol. 2010;38(1):77–82.
8. Dhillon IK, Hong CHL, Hu S, Sim YF, Goh BKC, Duggal MS, et al. Accuracy of the American Association of Endodontists diagnostic criteria for assessing pulp health in primary teeth. Clin Oral Investig. 2023;27(10):6043–53.
9. Laculli F, Rodríguez-Lozano FJ, Briseño-Marroquín B, Wolf TG, Spagnuolo G, Rengo S. Vital pulp therapy of permanent teeth with reversible or irreversible pulpitis: an overview of the literature. J Clin Med. 2022;11(14):4016.
10. Kazemipoor M, Moradi H, Mokhtari F, Kheirollahi K. Assessing the occurrence and severity of pre- and postendodontic pain in anemic and healthy female patients. Clin Exp Dent Res. 2023;9(6):1122–8.
11. Balcázar C, Isidro L, Nájera A, Hernández G, Rueda M, Garrido S. Factores de riesgo asociados a la patología periapical y pulpar. Horiz Sanit. 2017;16(2):111–9.
12. Aguirre-Ipenza R, Nieto-Gutiérrez W, Contreras W, Contreras PJ, Curioso WH. Iron deficiency anemia and dental caries: a systematic review and meta-analysis. Glob Pediatr Health. 2024;11:2333794X241273130.
13. Fatih MT, Mahmood MK, Ali YMM, Rasheed TA, Noori ZF, Kurda HA, et al. Association between iron deficiency anemia and dental caries in children: a systematic umbrella meta-analysis. J Hum Nutr Diet. 2025;38(6):e70149.
14. Schwendicke F, Dörfer CE, Schlattmann P, Foster Page L, Thomson WM, Paris S. Socioeconomic inequality and caries: a systematic review and meta-analysis. J Dent Res. 2015;94(1):10–8.
15. Villalba N, Jacquett N, Cabañas A, Chirife M. Pérdida prematura de dientes temporarios en niños que acudieron a un hospital público de la ciudad de Luque, Paraguay. Rev Odontopediatr Latinoam. 2013;3(1):25–34.
16. Instituto Nacional. de Estadística e Informática. Junín: compendio estadístico. Lima: INEI; 2017.
17. Ministerio de Salud. Norma técnica para el manejo preventivo, terapéutico y control de la anemia. Lima: MINSA; 2017.
18. Ministerio de Salud. Plan nacional para la reducción y control de la anemia materno infantil y la desnutrición crónica infantil en el Perú 2017–2021. Lima: MINSA; 2017.
19. Leon-Rodriguez JA, Espinoza-Salcedo M, Gutierrez-Polanco YM, Leon-Rodriguez JD, Lopez-Tisnado AA, Rivera-Cruz OS. Iron deficiency anemia as a risk factor for pulp disease in children from the central Peruvian jungle: a case–control study. J Appl Oral Sci. 2024;32:e20240014.
20. García A, Bujaldón A, Rodríguez A. Lesiones periapicales: diagnóstico y tratamiento. Av Odontoestomatol. 2015;31(1):31–42.

21. Xu Y, Yang L, Wu J, Fathaniard AH, Zhao N, Gao H, et al. Development and validation of a predictive model for severe caries in 12-year-olds from resource-limited regions: a cross-sectional study in China. BMC Oral Health. 2026;26:550.
22. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9.
24. American Association of Endodontists. American Board of Endodontics. Endodontic diagnosis. Chicago: American Association of Endodontists; 2013.
25. Ministerio de Salud. Norma técnica de salud para el uso del odontograma, NTS N° 188-2022-MINSA/DGIESP. Lima: MINSA; 2022.
26. Kamiab N, Mohammadi Kamalabadi Y, Sheikh Fathollahi M. DMFT of the first permanent molars, dmft and related factors among first-grade primary school students in Rafsanjan urban area. J Dent (Shiraz). 2021;22(2):109–17.
27. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.
28. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.
29. World Medical Association. WMA Declaration of Helsinki: ethical principles for medical research involving human participants. Ferney-Voltaire: WMA; 2024.
30. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report. J Am Coll Dent. 2014;81(3):4–13.
31. Council for International Organizations of Medical Sciences. International ethical guidelines for health-related research involving humans. Geneva: CIOMS; 2017.
32. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. JAMA. 2016;316(17):1818–9.
33. Walsh T, Worthington HV, Glenny AM, Marinho VCC, Jeroncic A. Fluoride toothpastes of different concentrations for preventing dental caries. Cochrane Database Syst Rev. 2019;3(3):CD007868.
34. Marinho VCC, Chong LY, Worthington HV, Walsh T. Fluoride mouthrinses for preventing dental caries in children and adolescents. Cochrane Database Syst Rev. 2016;7(7):CD002284.
35. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441.
36. Steyerberg EW. Validation in prediction research: the waste by data splitting. J Clin Epidemiol. 2018;103:131–3.
37. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.

Figures

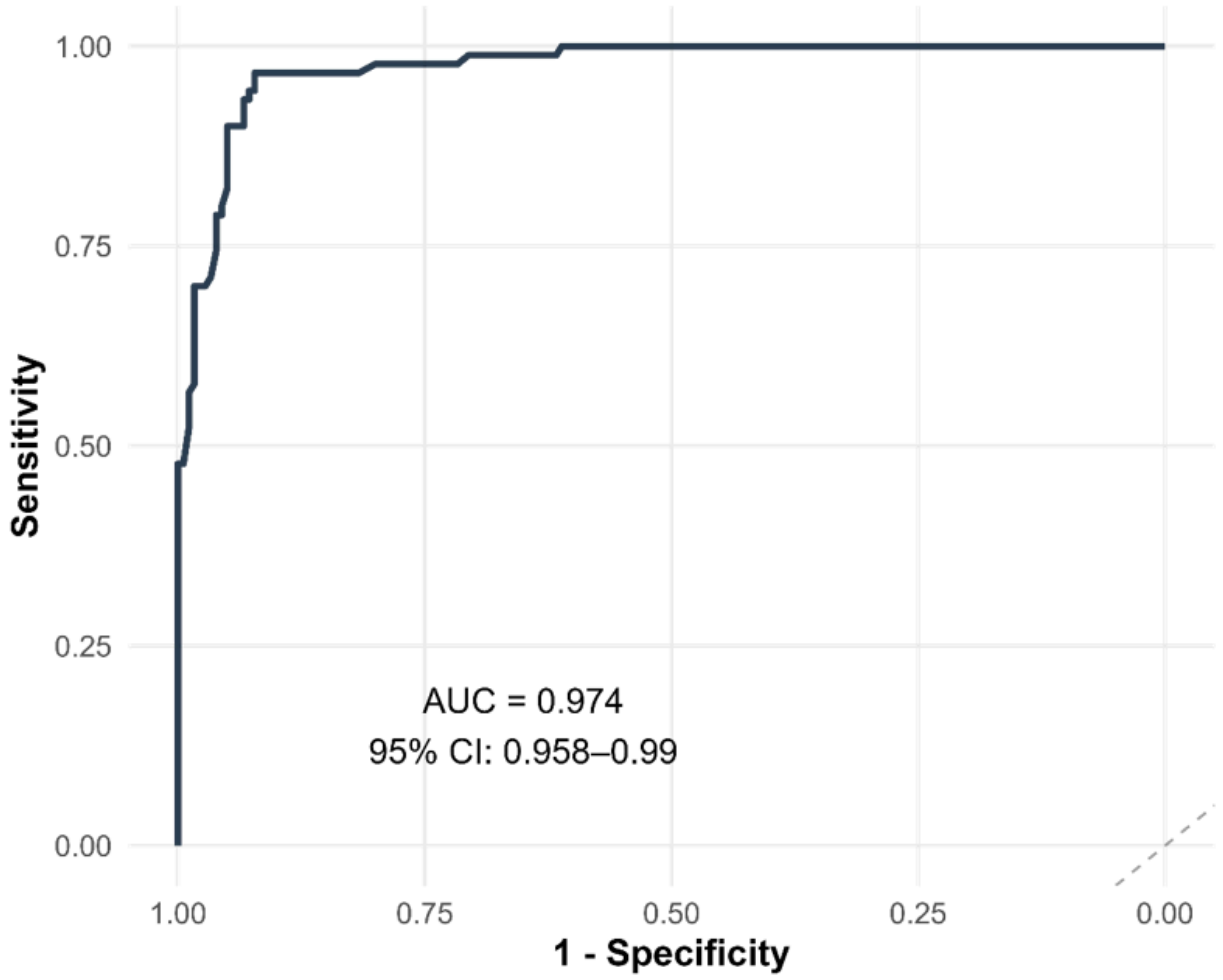


Figure 1

Receiver operating characteristic curve of the final logistic regression model for pulp disease

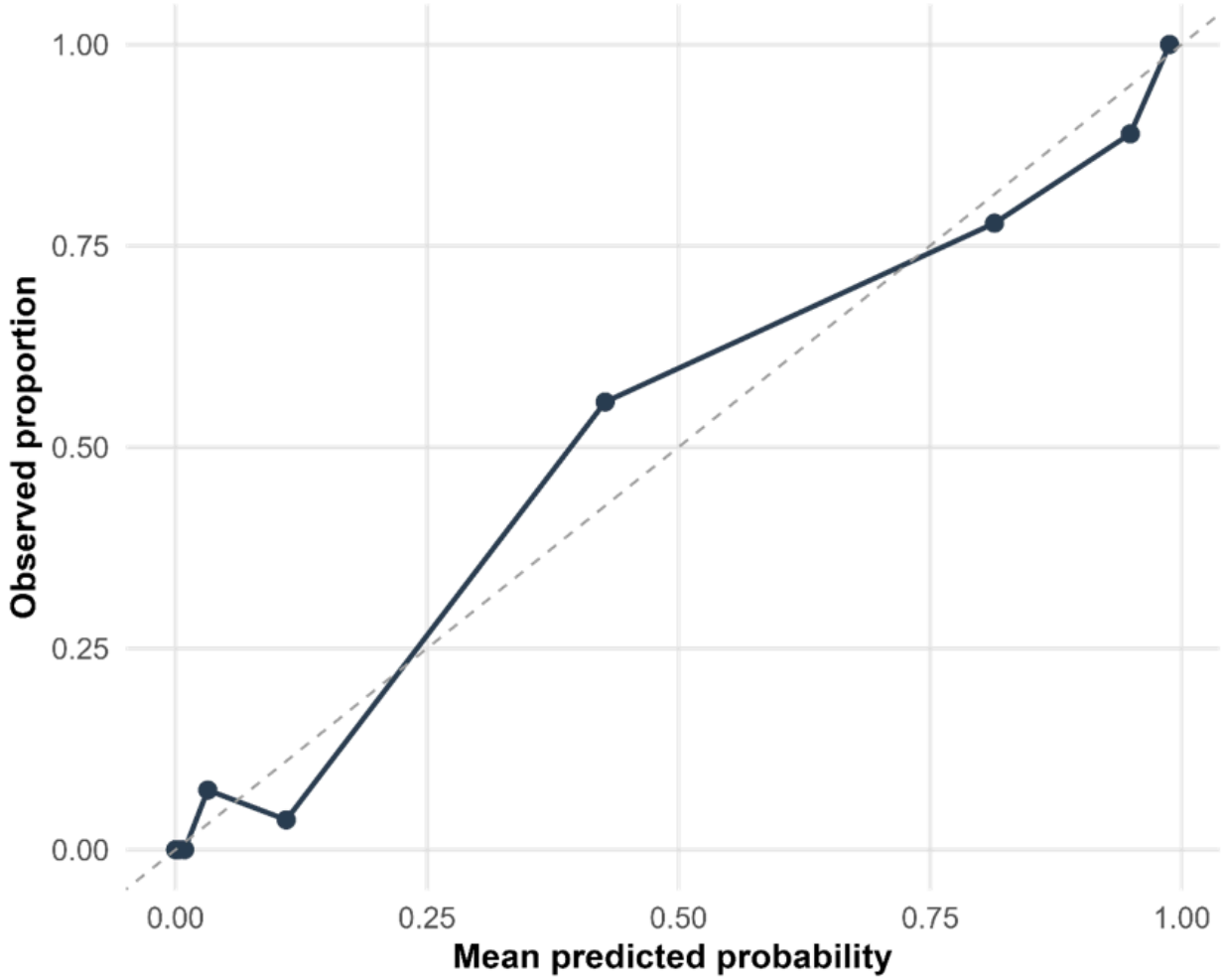


Figure 2

Calibration curve of the final logistic regression model for pulp disease