RESEARCH ARTICLE

Health Science Reports *Open Access* WILEY

Predicting the risk of infant mortality for newborns operated for congenital heart defects: A population-based cohort (EPICARD) study of two post-operative predictive scores

Nathalie Lelong[1] | Karim Tararbit[1] | Lise-Marie Le Page-Geniller[1] |
Jérémie Cohen[1,2] | Souad Kout[3] | Laurence Foix-L'Hélias[1,4] | Pascal Boileau[5] |
Martin Chalumeau[1,2] | François Goffinet[1,6] | Babak Khoshnood[1] | on behalf of the
EPICARD Study Group

[1]INSERM Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Université de Paris, CRESS, Paris, France

[2]Department of General Pediatrics and Pediatric Infectious Diseases, AP-HP, Hôpital Necker - Enfants malades, Université de Paris, Paris, France

[3]Department of Neonatology, CHI André Grégoire, Montreuil, France

[4]Department of Neonatology, Hôpital Armand Trousseau, Assistance Publique-Hôpitaux de Paris, Medecine Sorbonne University, Paris, France

[5]Department of Neonatology, CHI Poissy Saint-Germain-en-Laye, Poissy, France

[6]Port-Royal Maternity Unit, Cochin University Hospital, Assistance Publique Hôpitaux de Paris, Paris, France

Correspondence
Karim Tararbit, INSERM UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Center for Epidemiology and Statistics, 53 av. de

Abstract

Background: Whereas no global severity score exists for congenital heart defects (CHD), risk (Risk Adjusted Cardiac Heart Surgery-1: RACHS-1) and/or complexity

l'Observatoire, 75014 Paris, France.
Email: karim.tararbit@inserm.fr

(Aristotle Basic Complexity: ABC) scores have been developed for those who undergo surgery. Population-based studies for assessing the predictive ability of these scores are lacking.

Objective: To assess the predictive ability of RACHS-1 and ABC scores for the risk of infant mortality using population-based cohort (EPICARD) data for newborns with structural CHD.

Methods: The study population comprised 443 newborns who underwent curative surgery. We assessed the predictive ability of each score alone and in conjunction with an a priori selected set of predictors of infant mortality. Statistical analysis included logistic regression models for which we computed model calibration, discrimination (ROC), and a rarely used but clinically meaningful measure of variance explained (Tjur's coefficient of discrimination).

Results: The risk of mortality increased with increasing RACHS-1 and the ABC scores and models based on both scores had adequate calibration. Model discrimination was higher for the RACHS-1-based model (ROC 0.68, 95% CI, 0.58-0.79) than the ABC-based one (ROC 0.59, 95% CI, 0.49-0.69), $P = 0.03$. Neither score had the good predictive ability when this was assessed using Tjur's coefficient.

Conclusions: Even if the RACHS-1 score had better predictive ability, both scores had low predictive ability using a variance-explained measure. Because of this limitation and the fact that neither score can be used for newborns with CHD who do not undergo surgery, it is important to develop new predictive models that comprise *all* newborns with structural CHD.

KEYWORDS
congenital heart defects, infant mortality, severity scores, thoracic surgery

1 | INTRODUCTION

Congenital heart defects (CHD) are the most frequent group of major congenital anomalies, accounting for almost 1% of all births.[1-4] Despite considerable progress in the medical and surgical management of CHD,[5-7] they remain the most important cause of infant mortality due to congenital anomalies.[8] Moreover, survivors may have considerable short-term morbidity and long-term adverse neurodevelopmental outcomes.[9-12]

CHD represent a heterogeneous group of anomalies.[13,14] Accounting for this heterogeneity is an important imperative for providing reliable prognostic information to patients and their caretakers, as well as, comparing outcomes across centers or evaluation of alternative diagnostic and treatment strategies.

Whereas no global (all-inclusive) severity score exists for CHD, risk and/or complexity scores have been developed for patients with CHD who undergo surgery. Two commonly used scores aimed at accounting for heterogeneities in surgical procedures for CHD are (a) Risk Adjustment for Congenital Heart Surgery (RACHS-1) and (b) the Aristotle Basic Complexity (ABC) scores.[15-17]

Briefly, RACHS-1 aims to evaluate the risk associated with different surgical procedures whereas the ABC score aims specifically to evaluate the complexity of surgical procedures. Even if these two aspects are related and these scores are thus correlated, they do not evaluate the same concept. Moreover, the available literature has shown that they are not equivalent in terms of their predictive ability for outcomes such as postoperative mortality or length of hospital stay.[18-28]

In general, most of the literature on the health outcomes of newborns with CHD are based on data from specialized centers whereas population-based studies remain rare. In particular, none of the studies that have evaluated these scores have compared their predictive ability in population-based cohort data. In addition, previous studies have evaluated in-hospital or postoperative mortality and not all infant (<1-year) deaths, including those after discharge from the hospital. Finally, previous studies have evaluated these scores in the context of all surgical procedures including those for non-structural CHD, notably surgical repair of a patent ductus arteriosus. Hence, the extent to which the predictive ability of the scores may be applicable specifically to surgical procedures for *structural* CHD (exclusively) is not known. We emphasize however that our study was designed to assess the predictive ability of these scores for an outcome that was not the target for these scores (ie, overall infant mortality and not postsurgical mortality) and for a study population for which these scores

were not specifically designed (ie, structural CHD requiring surgery and not all cardiovascular abnormalities including in particular the large proportion of cases of patent ductus arteriosus that require surgery).

The objective of our study was to assess the predictive ability of RACHS-1 and ABC scores for the prediction of infant mortality, using data from a large, prospective, population-based cohort of newborns with structural CHD. An innovative aspect of the study was that in addition to the usual indices of predictive ability, namely calibration (goodness of fit) and discrimination (Receiving Operator Characteristic (ROC) curve), we also calculated a rarely used but intuitive and clinically meaningful measure of predictive ability, the Tjur's Coefficient of Discrimination.[29]

age in the birth cohorts between May 1, 2005 and April 30, 2008 were eligible for inclusion. Cases of patent ductus arteriosus (PDA) and patent foramen ovalae, as well as, cardiac tumors, cardiomyopathy (without structural CHD), and arrhythmias were excluded.

Multiple sources of data including all maternity units, pediatric cardiology and cardiac surgery centers, fetal and neonatal pathology departments, neonatal and pediatric intensive care units, infant units, and outpatient clinics in greater Paris and neighboring tertiary care centers were regularly consulted to attain completeness of case registrations and the information for each case. Informed consent was obtained from study participants, and the study was approved by an ethics committee (French National Committee of information and Liberty, CNIL).

2 | MATERIALS AND METHODS

2.1 | Data

The EPICARD (EPIdémologie des CARDiopathies congénitales) study was a population-based, prospective cohort study with long-term (8-year) follow-up of children with a structural CHD born to women in the Greater Paris area (Paris and its surrounding suburbs). All cases (live births, terminations of pregnancy for fetal anomaly [TOPFA], fetal deaths ≥20 weeks) diagnosed in the prenatal period or up to 1 year of

2.2 | Study population

The total number of cases included in the EPICARD study was 2867. After excluding TOPFA (N = 466) and fetal deaths (N = 53), our initial study population comprised 2348 live births (Figure 1). The total number of live births in the population base was 314 022; hence the live birth prevalence of CHD in the EPICARD study was 74.8 per 10 000.

Of the initial study population of 2348 live births, 1847 newborns were excluded as they did not have a surgical procedure (no procedures or catheter interventions only). Of those who
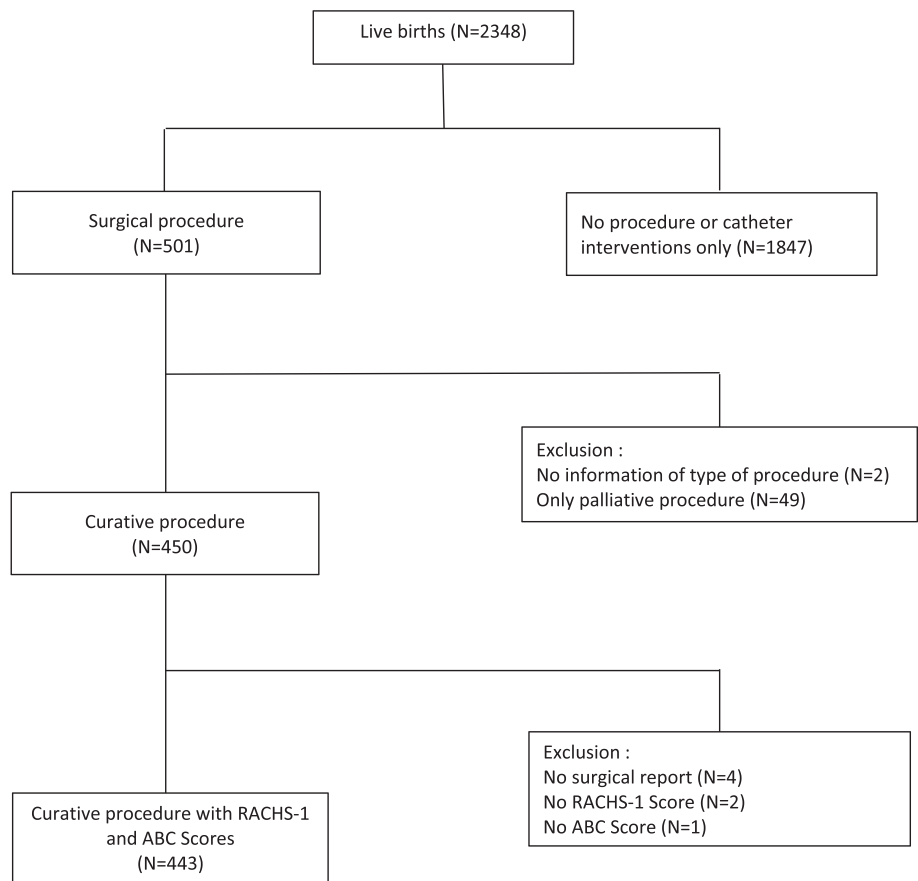

FIGURE 1   Flow chart of the study population – the EPICARD population-based cohort

underwent surgery (N = 501, 21% of all newborns with CHD), 450 patients had a curative, that is, intervention aimed at definitive repair rather than a palliative procedure. Among these, 443 could be assigned both a RACHS-1 and an ABC Score and the models which included only the RACHS-1 or ABC as predictive variables were based on data from these 443 newborns. One infant had missing information on the covariates (other predictor variables noted below). Hence, the models that included both the RACHS-1/ABC scores and the additional predictive variables were based on data from 442 infants.

The outcome variable was infant (≤1 year) mortality. The main predictor variable was the maximum RACHS-1 (or ABC) score, which was assigned by a neonatologist with training in pediatric cardiology (SK). Double coding of the procedures, blinded to the outcome, was conducted by SK. For newborns who had several surgical interventions, the surgical procedure with the highest RACHS-1 (or ABC) score was used to assign the maximum RACHS-1 / ABC for the infant. A few procedures could not be assigned a RACHS-1 (N = 2) and/or ABC score (N = 1) and four cases were excluded as surgical reports were not available.

2.3 | Statistical analysis

For descriptive analyses, we calculated proportions with 95% binomial exact confidence intervals. The chi-square test was used to test the statistical significance of the associations between scores (score categories) and other co-variables with the risk of mortality.

We used logistic regression to assess the predictive ability of the two scores, alone or together with other predictor variables: preterm birth (<37 weeks of gestation), type of malformations (isolated CHD, CHD, and chromosomal anomalies, CHD and anomalies of other systems), small for gestational age (<10th percentile, www.audipog.net), surgery at the first month of life, and sex at birth. These additional variables were chosen a priori as they are both known predictors of infant mortality and comprise the same set of variables as those used by Jenkins et al[15] in their original study for development of the RACHS-1 score.

Model calibration (goodness of fit) was evaluated using the Hosmer-Lemeshow statistic and model discrimination with area under the ROC curve. In addition, we estimated two measures of explained variation (R-square like statistics) developed specifically for logistic regression models. First, we calculated the Efron's R^2 as recommended by Mittlböck[30] in his review of the different R^2 proposed in the literature for logistic regression models. We also calculated the variance-explained measure by Tjur,[29] the "coefficient of discrimination". Tjur's measure has the advantage of having a simple and clinically intuitive interpretation. It ranges from 0, corresponding to a model with no predictive ability to 1 for a model with perfect predictive ability. Tjur's coefficient of discrimination corresponds to the difference between the average model-predicted probability of an event (eg, death) for those who experience the event (newborns who died in our study) minus the average model-predicted probability of an event for those who do not (newborns who survived the first year of life in our study).

A perfect predictive model would assign a probability of one to all those who experience the event and a probability of zero to those who do not. In general, the greater the difference between these two probabilities the greater is the predictive ability of the model.

The numerical results for Tjur's coefficient of discrimination and Efron's R^2 were comparable and we only give the results for Tjur's more intuitive measure here (Efron's measures are available upon request).

The Stata/SE software (version 13; Stata Corp, College Station, TX) was used for data analysis.

3 | RESULTS

Sociodemographic characteristics of the study population are detailed in Table S1 in supplementary material. Briefly, women were more frequently of French origin (43.2%), in the highest occupational category (professional, 27.7%), younger than 30 (40.0%), and primiparous (46.1%).

3.1 | Distribution of scores in the study population

Table 1 shows the distribution of RACHS-1 score categories for the newborns who had curative surgery. Approximately half of the surgical interventions were in category 2, about a third in category 3 and 15% in category 4. Less than 1% of cases were in categories 1 (lowest risk) or in categories 5 and 6 (highest risk). Table 1 also shows the distribution of ABC scores. Approximately 2% of cases were in category 1 (least complex operations), a third in category 2, 38% in category 3, and 26% in category 4 (most complex operations).

TABLE 1 Distribution of scores categories and infant mortality for the Risk Adjustment for Congenital Heart Surgery (RACHS-1) and Aristotle Basic Complexity (ABC) scores of patients with a curative procedure for congenital heart defects, the EPICARD study

			Infant Mortality	
Scores	N	%	N	% (95% CI)
RACHS-1				
1	3	0.7	0	0.0 (0.0, 70.8[a])
2	217	49.0	7	3.2 (1.3, 6.5)
3	149	33.6	8	5.4 (2.3, 10.3)
4	68	15.3	10	14.7 (7.3, 25.4)
5	3	0.7	2	66.7 (0.09, 99.2)
6	3	0.7	0	0.0 (0.0, 70.8[a])
ABC				
1	8	1.8	0	0.0 (0.0, 36.9[a])
2	149	33.6	6	4.0 (1.5, 8.6)
3	170	38.4	11	6.5 (3.3, 11.3)
4	116	26.2	10	8.6 (4.2, 15.3)

[a]One sided.

3.2 | Infant mortality

Of the total number of newborns (N = 443) with CHD who had a curative surgical intervention, 27 (6.1%, 95% CI, 4.1-8.7) died before 1 year of age. The risk of mortality increased significantly (P < 0.001), and in a "dose-response" manner across the categories of RACHS-1 score; from a low of 3.2% for the category combining 1 and 2 scores to 33.3% for the category combining 5 and 6 scores (Table 1). The risk of infant mortality also increased across categories of ABC; however, this increase was lower in magnitude and was not statistically significant.

3.3 | Predictive ability of models based on RACHS-1

Table 2 shows the results of the predictive models based on the RACHS-1 score alone and in combination with other predictor variables. The model with the RACHS-1 score model had moderate discrimination (Area under the ROC curve = 0.68, 95% CI, 0.58-0.79). After the addition of other predictors known to be associated with the risk of infant mortality (Table 2), the model discrimination increased (Figure 2, online Appendix, area under the ROC curve = 0.77, 95% CI, 0.67-0.87). This model also had adequate calibration (Hosmer-Lemeshow lack of fit test not statistically significant, P = 0.79).

Estimates of the predictive ability measure proposed by Tjur (coefficient of discrimination) were low for both the model with RACHS-1 alone (Tjur's coefficient = 3%) and in the full model with RACHS-1 and other predictors (Tjur's coefficient = 9%). The low value of Tjur's coefficient reflected in particular the low average model-predicted probability of mortality for infants who died (14.4% in the

best model, that is, RACHS-1 + additional predictors), which was fairly close to the average predicted probability of mortality for infants who survived (5.6%).

3.4 | Predictive ability of models based on ABC score

Table 3 shows the results of the predictive models based on the ABC score alone and in combination with other predictor variables. The model with the ABC score alone had fairly low discrimination (Area under the ROC = 0.59, 95% CI, 0.49-0.69); and in particular lower than the model based on RACHS-1 (P = 0.03). With the addition of other predictors, model discrimination increased (Figure 3, online
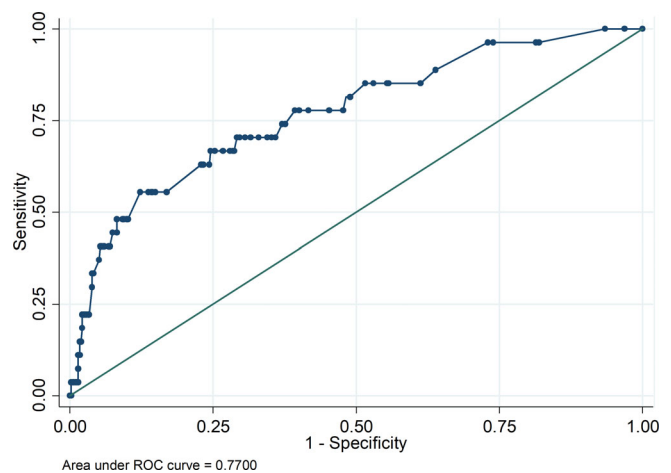

Area under ROC curve = 0.7700

FIGURE 2　Receiver operating characteristic (ROC) curve for the predictive model based on the RACHS-1 and other predictive variables (see text and Table 2) of infant mortality

TABLE 2　Risk on infant death according to the curative Risk Adjustment for Congenital Heart Surgery (RACHS-1) score, the EPICARD study

	OR (95% CI)	ORa (95% CI)
Score RACHS-1		
1–2	1.0 (Reference)	1.0 (Reference)
3	1.7 (0.6, 4.9)	2.0 (0.6, 6.2)
4	5.3 (1.9, 14.6)	9.8 (3.0, 32.4)
5–6	15.2 (2.4, 97.4)	19.4 (2.7, 140.6)
Preterm birth		2.6 (1.0, 7.0)
Type of malformations		
Isolated CHD		1.0 (Reference)
CHD and chromosomal anomalies		2.8 (0.8, 9.8)
CHD and anomalies of other systems		1.5 (0.5, 4.1)
Small for gestational age		2.6 (1.0, 6.6)
Surgery at first month of life		0.7 (0.3, 1.9)
Male		0.8 (0.3, 1.8)

TABLE 3　Risk on infant death according to the curative Aristotle basic comprehensive (ABC) score, the EPICARD study

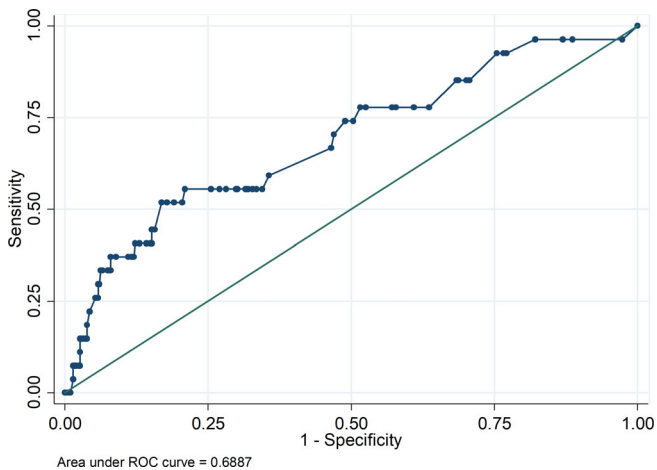|	OR (95% CI)	ORa (95% CI)
Score ABC		
1–2	1.0 (Reference)	1.0 (Reference)
3	1.7 (0.6, 4.8)	1.8 (0.6, 5.3)
4	2.4 (0.8, 6.8)	4.0 (1.1, 14.0)
Preterm birth		1.9 (0.7, 5.0)
Type of malformations		
Isolated CHD		1.0 (Reference)
CHD and chromosomal anomalies		2.8 (0.8, 9.5)
CHD and anomalies of other systems		1.6 (0.6, 4.3)
Small for gestational age		2.4 (1.0, 5.9)
Surgery at first month of life		0.8 (0.3-2.2)
Male		0.7 (0.3-1.6)

FIGURE 3 Receiver operating characteristic curve for the predictive model based on the Aristotle Basic Complexity score and other predictor variables (see text and Table 3) of infant mortality

Appendix, Area under the ROC curve = 0.69, 0.58-0.80) but remained less than that of the comparable model for RACHS-1 at the limit of statistical significance ($P = 0.06$).

Model calibration was adequate for the ABC score plus other predictors (Hosmer-Lemeshow lack of fit test not statistically significant, $P = 0.83$). The Tjur's coefficient for the ABC score-based models were even less than those for the RACHS-1, 0.6% and 4% for the ABC only and ABC + other predictors, respectively. As was the case for the RACHS-1 based models, the main reason for the low values of the Tjur's coefficient for ABC-based models was that the average model-predictive probability of death for infants who died was quite low (9.7% in the model with ABC + other predictors) and close to that of infants who survived (5.9%).

4 | DISCUSSION

4.1 | Principal findings

Using data from a prospective, population-based cohort study of 443 newborns with a structure congenital heart defect who had undergone curative (non-palliative) surgery, we assessed the predictive ability of two commonly used predictive scores of pediatric cardiac surgery, RACHS-1, and the ABC scores. We assessed the predictive ability of the two scores using both the most commonly used indices of predictive ability, namely discrimination (area under the ROC curve) and calibration (Hosmer-Lemeshow test for lack of fit), as well as, a clinically meaningful and intuitive measure of predictive ability, the Tjur's Coefficient of Discrimination. To our knowledge, this is the first study to use Tjur's coefficient in our field.

We found that both the models based on the RACHS-1 and ABC scores had adequate calibration but that the models based on RACHS-1 score had better discrimination. We also found that the addition of predictor variables known to be associated with the risk of infant mortality increased model discrimination, particularly for the ABC-based model.

We also found that both models had fairly low values of the Tjur's coefficient, which represents the difference between the average model-based predicted probability of death for newborns who die minus the average model-based predicted probability of death for those who survived. In general, all models had the shortcoming of having a low average predicted probability of death for those who died, which in turn resulted in low measures of predictive ability in terms of the Tjur's coefficient.

4.2 | Strengths of the study

Using data from a prospective cohort (EPICARD), our study represents the first population-based assessment of the predictive ability of the two commonly used scores (RACHS-1 and ABC Score) for predicting mortality of newborns with CHD who undergo corrective surgery. In addition to the usual indices of model performance, discrimination (area under the ROC) and calibration (Hosmer-Lemeshow test for lack of fit), we used, for the first time in our field (perinatal epidemiology), an intuitive and clinically meaningful measure of predictive ability, the Tjur's Coefficient of Discrimination.[29]

4.3 | Limitations of the data

Our study has certain limitations and caveats. We did not evaluate the performance of the Aristotle Comprehensive Complexity Score or the newer, empirically-derived STS (Society for Thoracic Surgery) - EACTS (European Association for Cardiothoracic Surgery (EACTS) Score.[28] The Aristotle Comprehensive Complexity Score is an expert-based score, which reflects complexity through three components: the potential for mortality, the potential for morbidity, and surgical technical difficulty.

The STS-EACT score is based on procedure-specific mortality rate estimates, which were calculated using data from the STS and EACTS databases for the period 2002 to 2007 and a Bayesian model that adjusted for small denominators. Each procedure is assigned a numeric score ranging from 0.1 to 5.0 based on the estimated mortality rate. Procedures are also sorted by increasing risk and grouped into five categories that were chosen to be optimal with respect to minimizing within-category variation and maximizing between-category variation. Both the Aristotle Comprehensive Complexity score and the STS-EACTS have been shown to have better predictive ability than the ABC Score and comparable or better predictive ability compared with the RACHS-1 score.

Nevertheless, the RACHS-1 and the ABC scores are the most common ones used in our population and evaluated in the literature. All the existing scores are intended to predict postoperative (hospital) mortality and not infant mortality. Hence, we evaluated these scores for an outcome that strictly speaking they were not intended to predict. For some categories of RACHS-1 and ABC score, we had very few cases. This was because we looked at the predictive scores for surgical interventions in the case of structural
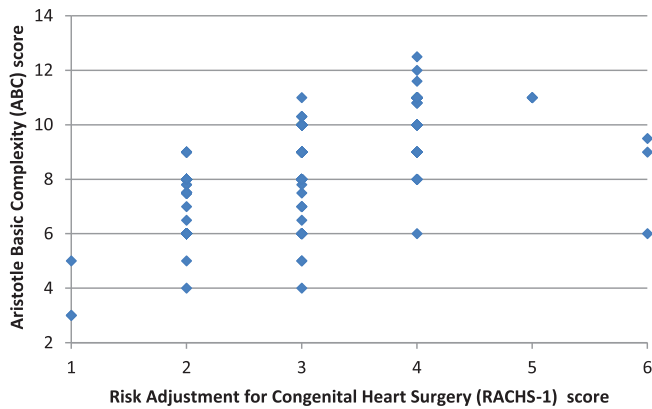
FIGURE 4 Scatterplot of Risk Adjustment for Congenital Heart Surgery (RACHS-1) and Aristotle Basic Complexity (ABC) scores, the Epicard study

CHD as explained above. Hence, we had to combine low (and high) score categories, which resulted in lower estimates of models' discrimination.

4.4 | Interpretations

As in previous studies, we found model calibration (Hosmer-Lemeshow test) to be adequate with no statistically significant evidence of lack of fit. However, we found a somewhat lower discrimination (smaller area under the ROC curves) for the two scores than those reported in the literature. Previous, hospital-based studies reported model discrimination of 70%-80% with areas under the ROC curve that were higher for RACHS-1.

For each level in RACHS-1 there was a heterogeneous distribution of the ABC scores (Figure 4). This may be at least explained by the fact that these two scores are intended to measure different concepts which are not fully correlated. This also can explain in part why the two scores had different discrimination and predictive ability. One or more of the following could explain this discrepancy. Our outcome was infant (< one-year) mortality whereas previous studies looked at postoperative hospital mortality (usually 30-days post-op). Nevertheless, we conducted analyses with the deaths restricted to hospital/30 days post-op mortality. The results were comparable to those we reported for infant mortality (detailed results available upon request). The scores were initially conceived and developed in the United States and there may be differences in practice or patient populations that explain why these models showed better model discrimination in studies that evaluated these scores at the time of their development and validation vs those found in our study.

Nevertheless, it is likely that the most important reason for the difference in the model discrimination in our study vs those in the literature is related to the selection criteria for our study population. We excluded from our study population all non-structural CHD whereas all cases of pediatric cardiac surgery, including the commonly performed PDA ligations were included in the previous studies based on all surgical cases. In contrast, our study included only *structural* congenital heart defects. Hence, cases of "functional" CHD, most

notably PDA, which comprise a large proportion of pediatric cardiac surgery cases, were excluded from our study population. This explains why we had so few cases in the lowest risk (RACHS-1) and complexity (ABC) score categories. In turn, given the few cases in the lowest risk/complexity for the two scores, we had to combine categories 1 and 2 (as well as 5 and 6 for the RACHS-1) in our models. This in turn reduces the predictive ability of the models, as combining categories results in loss of prognostic information.

An important limit of the currently available scores is that they can only be used for infants who undergo surgery. However, these infants accounted for only one-fifth of all newborns in our cohort of newborns with structural CHD. Moreover, many newborns with severe CHD may die prior to surgery either because surgery is deemed futile and compassionate care is given, or because death occurs before surgery can be performed. In our cohort, surgical cases accounted for only 1/3 of all infant deaths. Hence, it is important to develop new predictive scores for the risk of mortality and other health and developmental outcomes, which can be used for *all* newborns with structural CHD.

5 | CONCLUSIONS

In this population-based cohort study, we found that of the two commonly used scores for predicting the outcome of cardiac surgery for infants with CHD, the RACHS-1 score had higher discrimination than the ABC score. However, neither score had a good predictive ability when this was assessed using a measure (Tjur's "coefficient of discrimination") based on the difference in the average model-predictive probability of death for newborns who died vs the average model-predicted probability of death for those who survived. In particular, both models had low average model-predicted probabilities of death for infants who died.

CONFLICT OF INTEREST
The authors have no competing interests to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The EPICARD study was approved by the French National Committee of Information and Freedom (*Commission nationale de l'informatique et des libertés*). In accordance with French laws, at the time of recruitment for the study (observational/non-interventional) parents received an information letter in which it was noted that they could refuse to participate in the study.

ORCID

Karim Tararbit https://orcid.org/0000-0002-3188-0536

Lise-Marie Le Page-Geniller https://orcid.org/0000-0002-1772-2195

REFERENCES

1. Dolk H, Loane M, Garne E. Congenital heart defects in Europe: prevalence and perinatal mortality, 2000 to 2005. *Circulation*. 2011;123:841-849.
2. Hoffman JI. Incidence of congenital heart disease: I. Postnatal incidence. *Pediatr Cardiol*. 1995;16:103-113.
3. Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol*. 2002;39:1890-1900.
4. Hoffman JI, Kaplan S, Liberthson RR. Prevalence of congenital heart disease. *Am Heart J*. 2004;147:425-439.
5. Khoshnood B, De Vigan C, Vodovar V, et al. Trends in prenatal diagnosis, pregnancy termination, and perinatal mortality of newborns with congenital heart disease in France, 1983-2000: a population-based evaluation. *Pediatrics*. 2005;115:95-101.
6. Bonnet D, Coltri A, Butera G, et al. Detection of transposition of the great arteries in fetuses reduces neonatal morbidity and mortality. *Circulation*. 1999;99:916-918.
7. Allan LD. The outcome of fetal congenital heart disease. *Semin Perinatol*. 2000;24:380-384.
8. Lee K, Khoshnood B, Chen L, Wall SN, Cromie WJ, Mittendorf RL. Infant mortality from congenital malformations in the United States, 1970-1997. *Obstet Gynecol*. 2001;98:620-627.
9. Limperopoulos C, Majnemer A, Shevell MI, Rosenblatt B, Rohlicek C, Tchervenkov C. Neurodevelopmental status of newborns and infants with congenital heart defects before and after open heart surgery. *J Pediatr*. 2000;137:638-645.
10. Wernovsky G, Newburger J. Neurologic and developmental morbidity in children with complex congenital heart disease. *J Pediatr*. 2003;142:6-8.
11. Brown MD, Wernovsky G, Mussatto KA, Berger S. Long-term and developmental outcomes of children with complex congenital heart disease. *Clin Perinatol*. 2005;32:1043-1057.
12. Wernovsky G, Shillingford AJ, Gaynor JW. Central nervous system outcomes in children with complex congenital heart disease. *Curr Opin Cardiol*. 2005;20:94-99.
13. Franklin RC, Jacobs JP, Krogmann ON, et al. Nomenclature for congenital and paediatric cardiac disease: historical perspectives and the international pediatric and congenital cardiac code. *Cardiol Young*. 2008;18(Suppl 2):70-80.
14. Houyel L, Khoshnood B, Anderson RH, et al. Population-based evaluation of a suggested anatomic and clinical classification of congenital heart defects based on the International Paediatric and Congenital Cardiac Code. *Orphanet J Rare Dis*. 2011;6:64. https://doi.org/10.1186/1750-1172-6-64.
15. Jenkins KJ, Gauvreau K, Newburger JW, Spray TL, Moller JH, Iezzoni LI. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg*. 2002;123:110-118.
16. Welke KF, Shen I, Ungerleider RM. Current assessment of mortality rates in congenital cardiac surgery. *Ann Thorac Surg*. 2006;82:164-170.
17. Lacour-Gayet F, Clarke D, Jacobs J, et al. The Aristotle score: a complexity-adjusted method to evaluate surgical results. *Eur J Cardiothorac Surg*. 2004;25:911-924.
18. Kang N, Cole T, Tsang V, et al. Risk stratification in paediatric open-heart surgery. *Eur J Cardiothorac Surg*. 2004;26:3-11.
19. Boethig D, Jenkins KJ, Hecker H, et al. The RACHS-1 risk categories reflect mortality and length of hospital stay in a large German pediatric cardiac surgery population. *Eur J Cardiothorac Surg*. 2004;26:12-17.
20. Larsen SH, Pedersen J, Jacobsen J, Johnsen SP, Hansen OK, Hjortdal V. The RACHS-1 risk categories reflect mortality and length of stay in a Danish population of children operated for congenital heart disease. *Eur J Cardiothorac Surg*. 2005;28:877-881.
21. Mildh L, Pettilä V, Sairanen H, et al. Predictive value of paediatric risk of mortality score and risk adjustment for congenital heart surgery score after paediatric open-heart surgery. *Interact Cardiovasc Thorac Surg*. 2007;6:628-631.
22. O'Brien SM, Jacobs JP, Clarke DR, et al. Accuracy of the aristotle basic complexity score for classifying the mortality and morbidity potential of congenital heart surgery operations. *Ann Thorac Surg*. 2007;84:2027-2037.
23. Bojan M, Gerelli S, Gioanni S, et al. The Aristotle comprehensive complexity score predicts mortality and morbidity after congenital heart surgery. *Ann Thorac Surg*. 2011;91:1214-1221. https://doi.org/10.1016/j.athoracsur.2011.04.056.
24. Kang N, Tsang VT, Elliott MJ, de Leval MR, Cole TJ. Does the Aristotle score predict outcome in congenital heart surgery? *Eur J Cardiothorac Surg*. 2006;29:986-988.
25. Al-Radi OO, Harrell FE Jr, Caldarone CA, et al. Case complexity scores in congenital heart surgery: a comparative study of the Aristotle basic complexity score and the risk adjustment in congenital heart surgery (RACHS-1) system. *J Thorac Cardiovasc Surg*. 2007;133:865-875.
26. DeCampli WM, Burke RP. Interinstitutional comparison of risk-adjusted mortality and length of stay in congenital heart surgery. *Ann Thorac Surg*. 2009;88:151-156.
27. Bojan M, Gerelli S, Gioanni S, Pouard P, Vouhé P. Comparative study of the Aristotle comprehensive complexity and the risk adjustment in congenital heart surgery scores. *Ann Thorac Surg*. 2011;92:949-956.
28. O'Brien SM, Clarke DR, Jacobs JP, et al. An empirically based tool for analyzing mortality associated with congenital heart surgery. *J Thorac Cardiovasc Surg*. 2009;138:1139-1153.
29. Tjur T. Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. *Am Stat*. 2009;63(4):366-372.
30. Mittlböck M, Schemper M. Explained variation for logistic regression. *Stat Med*. 1996;15:1987-1997.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.