

Original article

Predicting serious complication risks after bariatric surgery: external validation of the Michigan Bariatric Surgery Collaborative risk prediction model using the Dutch Audit for Treatment of Obesity

Erman O. Akpinar, M.D.^{a,b,*}, Amir A. Ghaferi, M.D.^c, Ronald S. L. Liem, M.D.^{d,e},
Aaron J. Bonham, M.Sc.^c, Simon W. Nienhuijs, M.D., Ph.D.^f,
Jan Willem M. Greve, M.D., Ph.D.^{a,g,h}, Perla J. Marang-van de Mheen, Ph.D.ⁱ, on behalf of
the Dutch Audit for Treatment of Obesity (DATO) Research Group

^aDepartment of Surgery, Maastricht University Medical Center, NUTRIM School for Nutrition and Translational Research in Metabolism, Maastricht, The Netherlands

^bScientific Bureau, Dutch Institute for Clinical Auditing, Leiden, The Netherlands

^cDepartment of Surgery, University of Michigan, Ann Arbor, Michigan

^dDepartment of Surgery, Groene Hart Hospital, Gouda, The Netherlands

^eDutch Obesity Clinic, The Hague & Gouda, The Netherlands

^fDepartment of Surgery, Catharina Hospital, Eindhoven, The Netherlands

^gDepartment of Surgery, Zuyderland Medical Centre, Heerlen, The Netherlands

^hDutch Obesity Clinic South, Heerlen, The Netherlands

ⁱDepartment of Biomedical Data Sciences, Medical Decision Making, Leiden University Medical Centre, Leiden, The Netherlands

Received 13 June 2022; accepted 10 September 2022

Abstract

Background: Risk-prediction tools can support doctor–patient (shared) decision making in clinical practice by providing information on complication risks for different types of bariatric surgery. However, external validation is imperative to ensure the generalizability of predictions in a new patient population.

Objective: To perform an external validation of the risk-prediction model for serious complications from the Michigan Bariatric Surgery Collaborative (MBSC) for Dutch bariatric patients using the nationwide Dutch Audit for Treatment of Obesity (DATO).

Setting: Population-based study, including all 18 hospitals performing bariatric surgery in the Netherlands.

Methods: All patients registered in the DATO undergoing bariatric surgery between 2015 and 2020 were included as the validation cohort. Serious complications included, among others, abdominal abscess, bowel obstruction, leak, and bleeding. Three risk-prediction models were validated: (1) the original MBSC model from 2011, (2) the original MBSC model including the same variables but updated to more recent patients (2015–2020), and (3) the current MBSC model. The following predictors from the MBSC model were available in the DATO: age, sex, procedure type, cardiovascular disease, and pulmonary disease. Model performance was determined using the area under the curve (AUC) to assess discrimination (i.e., the ability to distinguish patients with events from those without events) and a graphical plot to assess calibration (i.e., whether the predicted absolute risk for patients was similar to the observed prevalence of the outcome).

*Correspondence: Erman O. Akpinar, M.D., Department of Surgery, Maastricht University Medical Centre, 6229 HX Maastricht, The Netherlands.

E-mail address: e.akpinar@maastrichtuniversity.nl (E.O. Akpinar).

Results: The DATO validation cohort included 51,291 patients. Overall, 986 patients (1.92%) experienced serious complications. The original MBSC model, which was extended with the predictors “GERD (yes/no),” “OSAS (yes/no),” “hypertension (yes/no),” and “renal disease (yes/no),” showed the best validation results. This model had a good calibration and an AUC of .602 compared with an AUC of .65 and moderate to good calibration in the Michigan model.

Conclusion: The DATO prediction model has good calibration but moderate discrimination. To be used in clinical practice, good calibration is essential to accurately predict individual risks in a real-world setting. Therefore, this model could provide valuable information for bariatric surgeons as part of shared decision making in daily practice. (*Surg Obes Relat Dis* 2023;19:212–221.) © 2023 American Society for Metabolic and Bariatric Surgery. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords:

Bariatric surgery; Risk-prediction model; Serious complication; External validation; Real-world; DATO; MBSC

Clinicians increasingly use risk-prediction tools to accurately estimate an individual’s risk profile to guide their doctor–patient (shared) decision making and to inform patients about the risks of surgery as we move to clinical care offering individualized treatments, care, and monitoring [1–3]. For instance, a recent systematic review reported on 16 models developed to predict diabetes remission after bariatric surgery [4], including 11 scoring systems and 5 logistic regression models. Clinicians caring for patients in other populations, for example, may use these scoring systems, provided that the prediction model is correct for that patient population; that is, it needs external validation because otherwise clinical decisions based on an (incorrect) prediction model may negatively influence patient outcomes. The aforementioned systematic review showed, for instance, that the ABCD (age, blood pressure, clinical features, and duration) score was validated in 9 studies and the Diabetes Remission (DiaRem) score in 6 studies, showing different results depending on the population.

The Michigan Bariatric Surgery Collaborative (MBSC) developed a prediction model for serious postoperative complications after bariatric surgery within 30 days [5]. The MBSC model reported a moderate ability to discriminate between patients with and without serious complications (C-statistic = .66) and good calibration, meaning that the model does not systematically over- or underestimate absolute complication risks [6]. When calibration is good, estimates from such prediction models can support shared decision making by informing patients of their individualized risks of a serious complication. When discrimination is good, the estimates also can be used to support treatment decision making by identifying patients at high risk for serious complications who may benefit from less invasive treatment options.

Most prediction models are used in the setting in which they were developed, and few are externally validated [7]. However, it is known that prediction models often perform less well outside the setting where they were developed, which can lead to systematically over- or underestimation

of risks [8], particularly if the patient population is rather different, which may occur between countries or over time. For the MBSC model, it is unknown whether it is generalizable to other populations. Hence external validation is needed to evaluate the performance of the model in a new setting before it is used in clinical practice [9,10]. The benefit of externally validating a model is that it allows prior predictive data knowledge to be retained rather than discarded, making the model more generalizable across new settings [11]. The ultimate predictive model would be a universal one that is highly accurate and widely applicable across all geographic settings.

In this context, a generalizable risk-prediction model would be an important addition for bariatric surgeons to support their shared decision making in daily practice. Therefore, this study aims to perform an external validation of the MBSC risk-prediction model in the Dutch population and assess its performance among bariatric patients treated in the Netherlands.

Methods

Study sample

Data were derived from the Dutch Audit for Treatment of Obesity (DATO). This audit has a nationwide coverage and has been mandatory since 2015, so it reflects real-world practice among patients undergoing bariatric surgery in the Netherlands. The DATO collects detailed information on patient, co-morbidity, treatment, follow-up, and short- and long-term outcome characteristics for patients undergoing bariatric surgery. Details of the DATO regarding data collection, quality, and validation have been described elsewhere [12].

The DATO’s scientific committee unanimously approved the use of the data for this study (reference number DATO-2022-142), and the study was carried out in accordance with the regulations of the Dutch Institute for Clinical Auditing. No informed consent was required because this is an opt-out quality registry and is performed in accordance with the ethical standards of Dutch law.

A population-based validation cohort was created within the DATO, including all patients undergoing primary bariatric surgery in the Netherlands from January 1, 2015, to December 31, 2020. Similar to the original MBSC model, patients undergoing revisional bariatric surgery were excluded because of the considerable heterogeneity in this group and associated increased risk for postoperative complications. Minimal data requirements for analysis were information on age, sex, type of procedure, and short-term (≤ 30 days) complications.

Outcome

The DATO registers the same serious complications as the MBSC, that is, complications categorized by the MBSC as grade 2 or grade 3 [5]. Grade 2 complications include abdominal abscess, bowel obstruction, leak, bleeding, wound infection or dehiscence, respiratory failure, renal failure, venous thromboembolism, and band-related problems requiring reoperation. Grade 3 complications include myocardial infarction or cardiac arrest, renal failure requiring long-term dialysis, respiratory failure requiring >7 days of mechanical ventilation or tracheostomy, and death.

External validation approach

To validate and update the risk-prediction model in a new setting, the following 3 steps were taken: First, we validated the original MBSC model (as published in 2011) for the DATO population. Second, the original MBSC model was updated using the same predictors but including patients undergoing bariatric surgery between 2015 and 2020. This was done because the original model was developed including patients undergoing bariatric surgery between 2006 and 2010, and outcomes may have improved over time given new scientific knowledge and improved surgical strategies. Subsequently, this updated MBSC model was validated for the DATO population. Finally, the current MBSC prediction model including potentially different predictors and patients undergoing bariatric surgery between 2015 and 2020 was validated for the DATO population.

Predictors

The predictors used in the original MBSC risk prediction model from 2011 remained the same for the updated MBSC model including patients undergoing bariatric surgery between 2015 and 2020 [5]. The current MBSC model includes 9 predictors, of which age, sex, ethnicity, and procedure type are forced into the model. The other predictors were added based on significantly improving the model: gastroesophageal reflux disease (GERD), cardiovascular disease (coronary artery disease, dysrhythmia, peripheral vascular disease, stroke, hypertension, and hyperlipidemia), prior venous thromboembolism (VTE), mobility limitation (requiring ambulation aids, nonambulatory, or bed bound),

and private insurance [13]. The predictors ethnicity, VTE, health insurance, and mobility limitation are not registered in the DATO and were therefore not included in the external validation. All the predictors registered in the DATO were coded according to the definitions and criteria stated in the original publication of the MBSC model [5].

Statistical analysis

Missing values

Missing data for predictor variables were input with multiple imputation techniques using all other available information on the patients to prevent bias due to missing data and loss of statistical power. A total of 5 complete data sets with 5 iterations were derived and averaged using the Multi-variate Imputation via Chained Equations (MICE) package in RStudio (Revolution Analytics, Mountain View, CA, USA).

Predictive performance

Discrimination and calibration are assessed to study the performance of the model. *Discrimination* refers to the ability of the model to distinguish between patients with and without a serious postoperative complication (i.e., patients with a serious complication have higher predicted risks than those without complications), quantified by the C-statistic or the area under the curve (AUC). An AUC of $<.6$ was defined as rather poor discrimination, $>.6$ as moderate discrimination, and $>.7$ and $>.8$ as good and excellent discrimination, respectively [14]. The calibration is evaluated using a visual calibration plot to assess how well the absolute predicted risk corresponds with the observed risk within subgroups of patients in daily practice. A calibration slope of 1 denotes a good fit between the observed and predicted risks, and a calibration slope >1 or <1 denotes miscalibration (i.e., respectively, systematic over- or underestimation of the predicted risk within subgroups of patients). The Brier score was used as a measure of overall model fit, with a significant score indicating a good fit.

Validation and updating methods

Several updating methods have been described for redeveloping a prediction model in case the model performs poorly in a new setting. These updating methods consist of logistic calibration, re-estimating coefficients, and selectively adding predictors. These model revision and extension methods, as described by Steyerberg et al. [15], were applied to update the models in the external validation set in case of poor initial performance. All external validations and updating methods are graphically plotted using the `val.prob` function of the `rms` package using RStudio version 4.0.2

Sensitivity analysis

A new risk-prediction model within the DATO population was developed and internally validated, including all

Table 1
Patient characteristics from the DATO between 2015 and 2020

Patient characteristic	DATO 2015–2020 (n = 51,291)	MBSC 2006–2010 (n = 25,469)*
Sex, n (%)		
Male	10,801 (21.1)	5525 (21.7)
Female	40,490 (78.9)	19,944 (78.3)
Age (y), mean (SD)	44.0 (11.62)	45.7 (11.4)
BMI (kg/m ²), mean (SD)	42.85 (5.26)	48.0 (8.5)
Procedure type, n (%)		
Open RYGB	39 (.1)	1092 (4.3)
RYGB	37,955 (74.0)	13,758 (54)
Adjustable gastric band	111 (.2)	8015 (31.5)
Sleeve gastrectomy	13,180 (25.7)	2279 (8.9)
BPD/DS	6 (.0)	325 (1.3)
T2D, n (%)	9795 (19.1)	8540 (33.7)
Hypertension, n (%)	17,536 (34.2)	13,544 (54.9)
Dyslipidemia, n (%)	10,161 (19.8)	12,774 (50.2)
GERD, n (%)	8007 (15.6)	12,315 (48.4)
OSAS, n (%)	9139 (17.8)	11,374 (44.8)
Musculoskeletal pain, n (%)	22,302 (43.5)	19,714 (77.4)
Pulmonary disease, n (%)	10,249 (20.0)	6588 (25.9)
Coronary artery disease, n (%)	22,351 (43.6)	1369 (5.4)
Previous venous thromboembolism, n (%)	—	979 (3.8)
Any smoking history, n (%)	—	9919 (39.0)
Mobility limitations, n (%)	—	1360 (5.3)
Private insurance, n (%)	—	18,333 (72.0)

DATO = Dutch Audit for Treatment of Obesity; MBSC = Michigan Bariatric Surgery Collaborative; SD = standard deviation; BMI = body mass index; RYGB = Roux-en-Y gastric bypass; BPD = biliopancreatic diversion; DS = duodenal switch; T2D = type 2 diabetes; GERD = gastroesophageal reflux disease; OSAS = obstructive sleep apnea syndrome.

* Characteristics of patients undergoing bariatric surgery in Michigan between 2006 and 2010 from Finks et al. [5].

patients between 2015 and 2020, and its performance was compared with the results from the external validation. All baseline characteristics showing a significant association ($P < .1$) with the outcome serious complications in univariate logistic regression analysis were included in the multivariable model. Multivariable logistic regression modeling was used to identify significant predictors ($P < .157$) using a stepwise backward selection. Bootstrapping with 250 samples was conducted for internal validation of the model and to correct for optimism [6,15]. Diabetes was forced into the model as a clinically relevant predictor.

Results

A total of 51,219 patients underwent primary bariatric surgery in the Netherlands between 2015 and 2020 and were included in the validation cohort. Overall, 986 patients (1.92%) experienced serious complications. Patient characteristics are shown in Table 1 and compared with the patients included in the original MBSC model, showing a rather different case mix. Patients from the DATO, on average, were younger, had a lower body mass index, and more often had coronary artery diseases. Furthermore, patients from the DATO less often had type 2 diabetes, hypertension, dyslipidemia, GERD, obstructive sleep apnea syndrome (OSAS), and musculoskeletal pain. In addition, patients in the DATO population more often underwent a Roux-en-Y gastric

bypass or sleeve gastrectomy rather than adjustable gastric band compared with the MBSC population.

Fig. 1 shows the calibration plot for the external validation of the original MBSC model for the DATO population. The coefficient for the variable *age* was significantly different in the DATO population and was updated as recommended [15]. The model was extended by predictors that are significantly associated with serious complications in the DATO population: *GERD* (no GERD/GERD with or without medication) and *OSAS* (no OSAS/OSAS with or without medication). This model (Fig. 1) shows good calibration, as indicated by the slope of 1.00, which suggests that predicted risk aligns well with the observed risk. However, the model had rather poor discrimination, as shown by the AUC of .574. The Brier score was significant, which indicates a good overall model fit.

The updated MBSC model including patients undergoing surgery between 2015 and 2020 again shows good calibration with a slope of 1.00 and moderate discrimination shown by the AUC of .602 (Fig. 2). The coefficients for *age* and *procedure type* were significantly different in the DATO validation cohort and thus were updated for the new geographic setting. Additionally, the model was extended with the predictors *GERD*, *OSAS*, *hypertension* (no hypertension/hypertension with or without medication), and *renal disease* (no renal disease/chronic renal

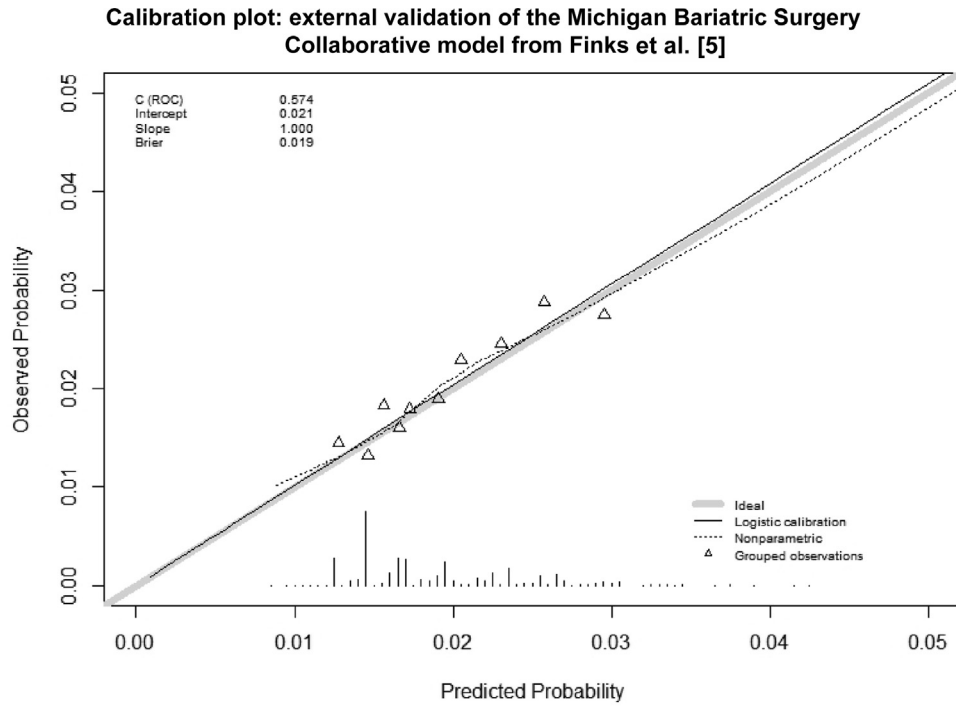


Fig. 1. C-statistic (receiver operating curve [ROC]) is the area under the curve of the model, which is .574. The calibration slope is 1.0. The Brier score is significant and denotes a good model fit.

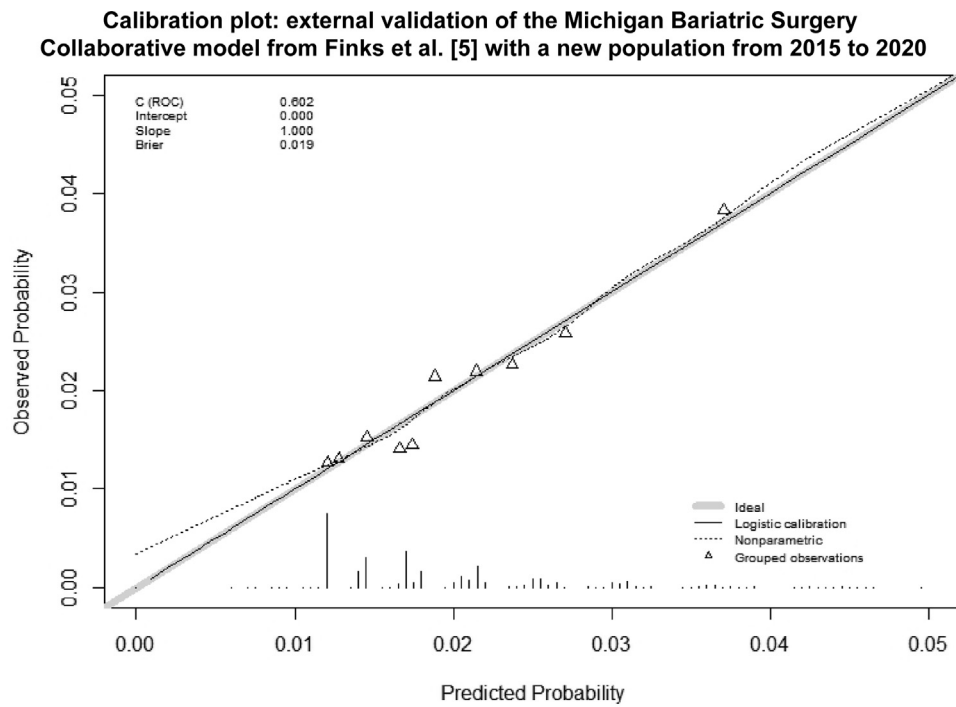


Fig. 2. C-statistic (receiver operating curve [ROC]) is the area under the curve of the model, which is .602. The calibration slope is 1.0. The Brier score is significant and denotes a good model fit.

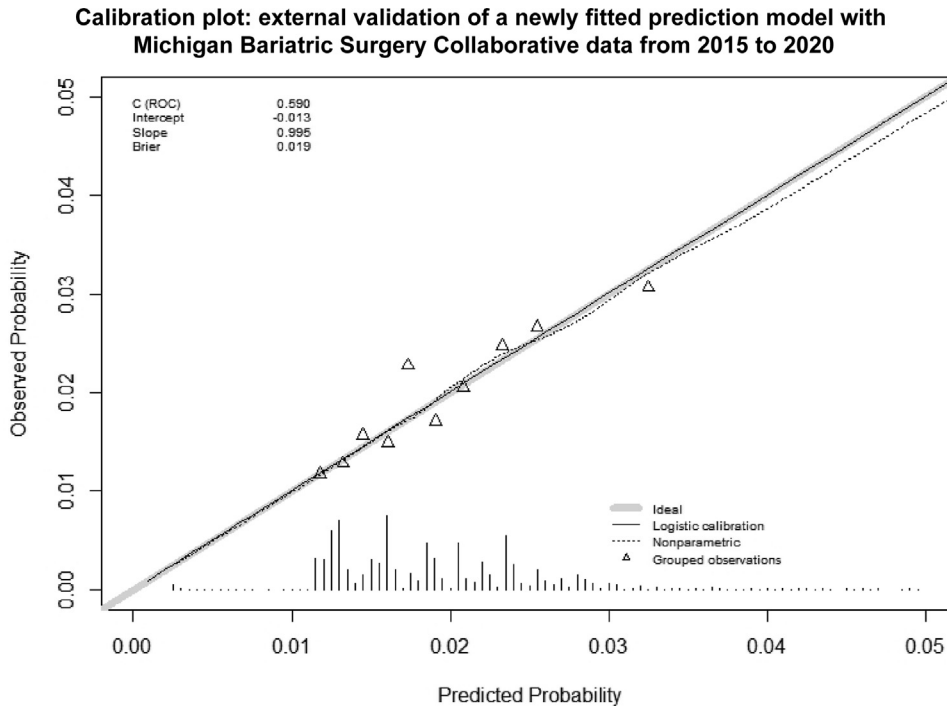


Fig. 3. C-statistic (receiver operating curve [ROC]) is the area under the curve of the model, which is .590. The calibration slope is .995. The Brier score is significant and denotes a good model fit.

insufficiency, renal failure requiring dialysis, nephrotic syndrome, and other renal diseases). These predictors were significantly associated with the occurrence of serious complications in the DATO cohort. The Brier score showed good overall model fit.

The current MBSC model has a calibration slope of .99 (Fig. 3) with a rather poor discrimination, shown by the AUC of .590. The coefficient for *age* was significantly different in the DATO population and thus updated. This model also was extended with the predictors *GERD*, *OSAS*, *hypertension*, and *renal disease*. The Brier score showed a good overall model fit. Table 2 shows the regression coefficients for all predictors included in the 3 MBSC models, as well as the coefficients for the best-performing DATO model (i.e., the external validation of the updated MBSC model).

Sensitivity analysis

The newly developed model on the DATO population of patients undergoing bariatric surgery between 2015 and 2020 included the variables *procedure type*, *age*, *sex*, *GERD*, *hypertension*, and *renal disease* based on statistical significance in the stepwise backward selection. The variable *diabetes* (no diabetes/diabetes with or without medication) was forced into the model based on clinical relevance. This (internally validated) model had an AUC of .606 (i.e., moderate discrimination with a calibration slope of 1.074; Supplemental Fig. 1), and the plot shows that the prediction

model systematically underestimates the actual risks, particularly for those at higher predicted risk.

Discussion

This study provided an external validation of the MBSC risk-prediction model for the Dutch population using the nationwide DATO as the validation cohort, which includes all patients receiving bariatric surgery in the Netherlands. The best performance was shown for the updated MBSC model (Fig. 2), which showed a moderate discrimination slightly lower than the original model (.60 versus .66) and a good calibration. Some predictors (*age* and *procedure type*) had significantly different effects in the validation cohort and were therefore updated. In addition, the model was extended with significant predictors of serious complications in the DATO: *GERD*, *OSAS*, *hypertension*, and *renal disease*. Although the ideal model would have higher discriminative ability (preferably AUC >.8), this is likely not feasible given the moderate discrimination in the development cohort (AUC = .66). For meaningful use in clinical practice, the model needs good calibration, meaning that predicted risks are similar to the actual observed risks and therefore can be communicated to patients and physicians as part of shared decision making.

The best-performing DATO model includes several predictors that are also reported in the literature to be significantly associated with serious complications. These predictors include *age*, *male sex*, *procedure type*, *GERD*,

Table 2

Regression coefficients for predictors of the original MBSC prediction model, the updated MBSC prediction model, the current MBSC model, and the externally validated DATO risk-prediction model

Predictor	Original MBSC*	Updated MBSC†	Current MBSC‡	Best-performing DATO model§
Intercept	−5.12	−2.0677	−2.8827	−3.8707
Age (yr)				
>50	.3225	.1288	—	.4060
Age (yr)				
<20	—	—	Ref.	—
20–29	—	—	.5948	—
30–39	—	—	.808	—
40–49	—	—	.9579	—
50–59	—	—	.9173	—
≥60	—	—	.9806	—
Male sex	.2321	.2335	.2841	.1669
Procedure type				
Adjustable gastric band	Ref.	−2.7925	−2.3007	−1.5805
Laparoscopic RYGB	1.2759	−1.6749	−1.6331	−.9366
Sleeve gastrectomy	.8988	−2.5589	−2.4741	−.5744
Open RYGB	1.2556	Ref.	Ref.	Ref.
BPD/duodenal switch	2.2702	−1.0542	−1.4858	−11.9429
SADI-S/gastric balloon	—	−11.3539	−11.1019	−11.6160
Ethnicity (White)	—	—	Ref.	—
Black	—	—	.2192	—
Hispanic	—	—	.0892	—
Other	—	—	−.1149	—
Coronary artery disease	.4260	.2498	.184	.0043
VTE	.6410	.7172	.7677	—
Pulmonary disease	.3150	.1353	—	.0023
Smoking history	.1797	.0347	—	—
GERD	—	—	.2075	.1978
OSAS	—	—	—	.1521
Hypertension	—	—	—	.1845
Renal disease	—	—	—	.4041
Mobility limitations	.4784	.474	.4276	—
Private insurance	—	—	−.2962	—
AUC	.66	.65	.655	.602

MBSC = Michigan Bariatric Surgery Collaborative; DATO = Dutch Audit for Treatment of Obesity; Ref. = reference; RYGB = Roux-en-y gastric bypass; BPD = biliopancreatic diversion; SADI-S = single-anastomosis duodenoileal bypass with sleeve gastrectomy; VTE = venous thromboembolism; GERD = gastroesophageal reflux disease; OSAS = obstructive sleep apnea syndrome; AUC = area under the curve.

* Original MBSC model predicting serious postoperative complications between 2006 and 2010.

† Updated MBSC model including patients undergoing bariatric surgery between 2015 and 2020.

‡ Current MBSC model including patients undergoing bariatric surgery between 2015 and 2020.

§ External validation of the DATO model (Fig. 2) including patients undergoing bariatric surgery between 2015 and 2020.

OSAS, hypertension, renal disease, and pulmonary disease [3,5,16–18]. Age, male sex, procedure type, and pulmonary disease also were included in the original MBSC model and GERD in the current MBSC model, whereas OSAS, hypertension, and renal disease were not included in any of the MBSC models. The predictor BMI, which has been identified as a risk factor in previous studies [19,20], did not have an independent significant association with serious complications in the current study, nor was it included in any of the MBSC models. Part of the explanation could be that patients with BMIs >50 kg/m² are known to be at increased risk for 30-day morbidity, whereas the DATO and the MBSC cohort had lower average BMIs of 42.85 (standard deviation [SD] = 5.26) and 48 (SD = 8.5) kg/m², respectively [21].

Further differences in predictors between the DATO and the MBSC model are that the variables ethnicity, mobility limitations, VTE, and private insurance are not recorded in the DATO. It has to be noted that all patients in the Netherlands have health insurance by law, including coverage for bariatric surgery. Although national registries have the common purpose to assess and improve the quality of care, registries often differ in defining and collecting variables [22]. In addition, the need for mobility aids or being bed bound is rarely the case in Dutch patients with morbid obesity, likely explained by the considerably lower average BMI compared with the MBSC population, making it redundant to record this predictor. Notably, mobility limitations seem to occur in 5% of the MBSC cohort (Table 1). The mobility limitations variable therefore most likely acts as a

proxy to capture the risk of patients with extremely high BMIs in the MBSC model (i.e., $>50 \text{ kg/m}^2$), who have increased risk for 30-day morbidity [21].

Furthermore, the DATO does not register ethnicity. However, co-morbidities such as diabetes or hypertension may act as a proxy because they occur more frequently in some ethnic groups and may have different associations with the outcome [23,24]. This would explain why hypertension was needed to extend the DATO model based on its significant association with serious complications and thereby may have captured part of what was covered by the *ethnicity* variable in the MBSC model. Diabetes, in contrast, did not significantly add to the DATO model, most likely because it was already captured by the predictors *cardiovascular disease* and *renal disease*, both long-term consequences of diabetes.

The best-performing DATO model shows good calibration. This means, for instance, that the individual risk prediction for a woman aged 55 years with hypertension undergoing a sleeve gastrectomy who has a predicted risk of 2.1% will accurately match the observed risk for patients with these characteristics. This is essential for using the prediction model in clinical practice because making clinical decisions based on a miscalibrated prediction model that systematically under- or overestimates the risks in some subgroups could be harmful if, for example, a procedure carries a much higher serious complication risk for particular patients. The discriminative ability, in contrast, is .602, which means that the model is consistent 60% of the time in predicting higher risks for patients who will experience serious complications. The serious complication rates after bariatric surgery in the DATO are consistent with the MBSC and current literature [25,26], which is important because differences in outcome incidence will affect model performance and particularly induce miscalibration [10,15]. Nonetheless, the DATO population is relatively homogeneous in its patient characteristics, which makes it harder for the model to discriminate between patients with and without serious complications [27]. In addition, the relatively low complications rates, likely due to centralization and high annual volumes [28,29], reflect the high quality of bariatric care in both cohorts, making the occurrence of serious complications a difficult-to-predict clinical problem. Finally, some predictors may discriminate better for specific complications such as leak rather than all complications combined [18,30].

Other risk-prediction models include the bariatric surgical risk calculator, recently developed by the Metabolic and Bariatric Surgery Accreditation and Quality Improvement Program (MBSAQIP) [31]. This model also has moderate discrimination for overall serious complications but good discrimination for specific complications such as leak. To our knowledge, this model is merely used in the setting in which it was developed and provides information for patients of the MBSAQIP cohort, but no external validation has been reported, which could lead to inaccurate risk predictions when the model is used in a new patient population. Continued international

collaborations with multiple national cohorts and external validations in diverse patient populations are likely needed to further enhance the generalizability and optimize existing prediction models to ensure meaningful use in clinical practice.

It is imperative to conduct an external validation of a prediction model before it is implemented in clinical practice because the model generally performs not as well in a new setting. The results of this study show good calibration for the best-performing DATO model, which can be used to inform patients and physicians about the absolute risks during shared decision making. This study also highlights the importance of external validation of prediction models to retain prior information and add information that is significantly important for the new setting, which, in turn, improves generalizability [32]. Also important is that the current model contains common noninvasive predictors that are relatively easy to retrieve during patient consultations [33]. Future studies are needed to show whether implementation of the current risk-prediction model affects clinical decision making and is accepted by surgeons in daily practice. Overall, this study calls attention to online-accessible bariatric surgery risk calculators, which are being used sporadically. It is a reminder that without external validation, a risk calculator may not always be accurate in a patient population different from the setting where it was developed, potentially compromising patient outcomes.

The strength of this study is that the DATO and the MBSC are both population-based registries that capture the whole population rather than a selection of patients. Furthermore, because the MBSC model was updated including patients treated between 2015 and 2020, possible differences over time in treatment, for example, were taken into account. Moreover, an external validation enhances the model's generalizability and retains prior information, and given the good calibration, the risk-prediction model can be implemented in clinical practice.

Some limitations should be noted. First, this study only looked at serious complications occurring within 30 days and did not investigate the risk of any long-term outcomes after bariatric surgery, whereas these long-term outcomes also may influence decision making, for example, regarding type of bariatric procedure. Furthermore, we did not capture all the variables included in the MBSC model. However, as explained previously for the risk factor *ethnicity* (which was forced in the original MBSC model), it seems likely that this may have been captured by extending the model with the variables *hypertension*, *cardiovascular disease*, and *renal disease*. Moreover, the model's discriminative ability is only moderate, so the model seems less useful to identify high-risk patients, for example, for inclusion in a trial.

Conclusion

The external validation of the MBSC model for the Dutch bariatric population has good calibration, meaning that it

adequately predicts individual risks in a real-world setting, but it has only moderate discrimination. This model could provide useful information for bariatric surgeons in daily practice to enable communicating individualized complication risks to patients as part of shared decision making.

Acknowledgments

The authors thank all surgeons, registrars, physician assistants, and administrative nurses who registered patients in the Dutch Audit for Treatment of Obesity (DATO). This article was written on behalf of the DATO Research Group: L. M. de Brauw, M.D., Ph.D. (Spaarne Gasthuis, Haarlem); S. M. M. de Castro, M.D., Ph.D. (OLVG Hospital, Amsterdam); S. L. Damen, M.D. (Medical Centre Leeuwarden, Leeuwarden); A. Demirkiran, M.D., Ph.D. (Red Cross Hospital, Beverwijk); M. Dunkelgrün, M.D., Ph.D. (Franciscus Gasthuis and Vlietland, Rotterdam); I. F. Faneyte, M.D., Ph.D. (ZGT Hospital, Almelo and Hengelo); J. W. M. Greve, M.D., Ph.D. (Zuyderland Medical Centre, Heerlen); G. van't Hof, M.D. (Dutch Bariatric Centre South-West, Bergen op Zoom); I. M. C. Janssen, M.D., Ph.D. (Dutch Obesity Clinics, Zeist); E. H. Jutte, M.D. (Medical Centre Leeuwarden, Leeuwarden); R. A. Klaassen, M.D. (Maasstad Hospital, Rotterdam); E. A. G. L. Lagae, M.D., Ph.D. (ZorgSaam Zorggroep Zeeuws-Vlaanderen, Terneuzen); B. S. Langenhoff, M.D., Ph.D. (ETZ Hospital, Tilburg); R. S. L. Liem, M.D. (Groene Hart Hospital and Dutch Obesity Clinic, Gouda and The Hague); A. A. P. M. Luijten, M.D., Ph.D. (Máxima Medical Centre, Eindhoven); S. W. Nienhuijs, M.D., Ph.D. (Catharina Hospital, Eindhoven); R. Schouten, M.D., Ph.D. (Flevo Hospital, Almere); R. M. Smeenk, M.D., Ph.D. (Albert Schweitzer Hospital, Dordrecht); D. J. Swank, M.D., Ph.D. (Dutch Obesity Clinic West, Den Haag); M. J. Wiezer, M.D., Ph.D. (St. Antonius Hospital, Utrecht); and W. Vening, M.D., Ph.D. (Rijnstate Hospital, Arnhem).

Disclosures

A. A. Ghaferi receives salary support from Blue Cross Blue Shield of Michigan as the director of the Michigan Bariatric Surgery Collaborative.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.soard.2022.09.008>

References

- [1] Kappen TH, van Klei WA, van Wolfswinkel L, et al. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res* 2018;2:1–11.
- [2] O'Neill C, Hutchings HA, Abbott Z, et al. Prognostic prediction tools and clinician communication: a qualitative study of the effect of the STUMBL tool on clinical practice. *BMC Emerg Med* 2020;20:1–7.
- [3] DeMaria EJ, Murr M, Byrne TK, et al. Validation of the obesity surgery mortality risk score in a multicenter study proves it stratifies mortality risk in patients undergoing gastric bypass for morbid obesity. *Ann Surg* 2007;246:578–82.
- [4] Singh P, Adderley NJ, Hazlehurst J, et al. Prognostic models for predicting remission of diabetes following bariatric surgery: a systematic review and meta-analysis. *Diabetes Care* 2021;44:2626–41.
- [5] Finks JF, Kole KL, Yenumula PR, et al. Predicting risk for serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. *Ann Surg* 2011;254:633–40.
- [6] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* 2010;21:128.
- [7] Siontis GCM, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;68:25–34.
- [8] Kappen TH, Vergouwe Y, Van Klei WA, et al. Adaptation of clinical prediction models for application in local settings. *Med Decis Making* 2012;32(3):E1–10.
- [9] Collins GS, De Groot JA, Dutton S, et al. External validation of multi-variable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40.
- [10] Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14:49–58.
- [11] Janssen KJM, Vergouwe Y, Kalkman CJ, et al. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth* 2009;56:194–201.
- [12] Poellemeijer YQM, Liem RSL, Nienhuijs SW. A Dutch nationwide bariatric quality registry: DATO. *Obes Surg* 2018;28:1602–10.
- [13] Michigan Bariatric Surgery Collaborative (MBSC) [Internet]. Arbor-Metrix, Inc.; c2022 [cited 2022 Aug 11]. MBSC outcomes calculator [about 3 screens]. Available from: <https://mbsc.arbormetrix.com/Registry/public/calculator/uiCalculator/?menuId=1013>.
- [14] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010;5:1315–6.
- [15] Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer; 2009.
- [16] Morino M, Toppino M, Forestieri P, et al. Mortality after bariatric surgery: analysis of 13,871 morbidly obese patients from a national registry. *Ann Surg* 2007;246:1002–7.
- [17] Flum DR, Belle SH, King WC, et al. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med* 2009;361:445–54.
- [18] Chang SH, Freeman NLB, Lee JA, et al. Early major complications after bariatric surgery in the USA, 2003–2014: a systematic review and meta-analysis. *Obes Rev* 2018;19:529–37.
- [19] Villamere J, Gebhart A, Vu S, et al. Body mass index is predictive of higher in-hospital mortality in patients undergoing laparoscopic gastric bypass but not laparoscopic sleeve gastrectomy or gastric banding. *Am Surg* 2014;80(10):1039–43.
- [20] Parikh J, Yermilov I, Mcgory M, et al. Is high BMI associated with specific complications after laparoscopic Roux-en-Y gastric bypass? *Am Surg* 2007;73(10):959–62.
- [21] Nasser H, Ivanics T, Leonard-Murali S, et al. Perioperative outcomes of laparoscopic Roux-en-Y gastric bypass and sleeve gastrectomy in super-obese and super-super-obese patients: a national database analysis. *Surg Obes Relat Dis* 2019;15:1696–703.
- [22] Akpınar EO, Marang-van de Mheen PJ, Nienhuijs SW, et al. National bariatric surgery registries: an international comparison. *Obes Surg* 2021;31:3031–9.
- [23] Cheng YJ, Kanaya AM, Araneta MRG, et al. Prevalence of diabetes by race and ethnicity in the United States, 2011–2016. *JAMA* 2019;322:2389–98.
- [24] Fei K, Rodriguez-Lopez JS, Ramos M, et al. Racial and ethnic subgroup disparities in hypertension prevalence. *New York City Health*

- and Nutrition Examination Survey, 2013–2014. *Prev Chronic Dis* 2017;14:E33.
- [25] Poelmeijer YQM, Liem RSL, Våge V, et al. Perioperative outcomes of primary bariatric surgery in north-western Europe: a pooled multi-national registry analysis. *Obes Surg* 2018;28:3916.
- [26] Birkmeyer NJO, Dimick JB, Share D, et al. Hospital complication rates with bariatric surgery in Michigan. *JAMA* 2010;304:435–42.
- [27] Merkow RP, Hall BL, Cohen ME, et al. Relevance of the C-statistic when evaluating risk-adjustment models in surgery. *J Am Coll Surg* 2012;214:822–30.
- [28] Bouchard P, Demyttenaere S, Court O, et al. Surgeon and hospital volume outcomes in bariatric surgery: a population-level study. *Surg Obes Relat Dis* 2020;16:674–81.
- [29] Zevin B, Aggarwal R, Grantcharov TP. Volume-outcome association in bariatric surgery: a systematic review. *Ann Surg* 2012;256:60–71.
- [30] Husain F, Jeong IH, Spight D, et al. Risk factors for early postoperative complications after bariatric surgery. *Ann Surg Treat Res* 2018;95:100.
- [31] Grieco A, Huffman KM, Cohen ME, et al. The Metabolic and Bariatric Surgery Accreditation and Quality Improvement Program bariatric surgical risk/benefit calculator: 30-day risk. *Surg Obes Relat Dis* 2021;17:1117–24.
- [32] Arterburn D, Johnson ES, Butler MG, et al. Predicting 90-day mortality after bariatric surgery: an independent, external validation of the OS-MRS prognostic risk score. *Surg Obes Relat Dis* 2014;10:774–9.
- [33] Baker T, Gerdin M. The clinical usefulness of prognostic prediction models in critical illness. *Eur J Intern Med* 2017;45:37–40.