Review, performance comparison, and validation of models predicting type 2 diabetes remission after bariatric surgery in a Western European population

Review, performance comparison, and validation of models predicting type 2 diabetes remission after bariatric surgery in a Western European population.

Abstract

Purpose: The majority of patients with type 2 diabetes (T2DM) achieve remission after bariatric surgery. Several models are available to preoperatively predict T2DM remission. This study compares the performance of these models in a Western population one year after surgery and explores their predictive value in comparison to a model specifically designed for our study population.

Materials and Methods: Prediction models were retrieved using a literature search. Patients were retrospectively selected from a database of the Antwerp University Hospital. Performance of the models was assessed by determining the area under the receiver operating characteristic curve (AUROC), the accuracy, the goodness-of-fit, and by comparing them to a custom-made logistic model.

Results: The probability of T2DM remission was calculated using 11 predictive scoring models and 8 regression models in a cohort of 250 patients. Complete T2DM remission occurred in 64.0% of patients. The IMS score (AUROC=0.912; accuracy=83.6%), DiaBetter score (0.907; 82.0%), and Ad-DiaRem score (0.903; 82.8%) best predicted T2DM remission and closely approached the performance of the custom-constructed model (0.917; 84.0%). The model by Ioffe et al. (0.630; 69.2%), Umemura et al. (0.692; 71.4%), and the ABCD score (0.757; 72.8%) were the least accurate.

Conclusion: Most T2DM remission models reliably predicted one-year T2DM remission, with limited inter-model differences. The accuracy of most models approached that of the custom-constructed model, indicating a high predictive capability and performance in our patient cohort. To date, most models are only validated to estimate T2DM remission one year after surgery and they do not predict long-term remission.

Key Words

Diabetes remission, bariatric surgery, Type 2 diabetes mellitus, Predictive models

Abbreviations

95% CI	95% Confidence Interval
ADA	American Diabetes Association
AUROC	Area Under the Receiver Operating Characteristic curve
BMI	Body Mass Index
FPG	Fasting Plasma Glucose
HOMA	Homeostatic Model Assessment
IMS	Individualized Metabolic Score
MDR	Metabolic surgery Diabetes Remission score
OGTT	Oral Glucose Tolerance Test
OR	Odds Ratio
%EBMIL	Percentage Excess BMI Loss
%TWL	%Total Weight Loss
ROC	Receiver Operating Characteristics Curve
SD	Standard Deviation
T2DM	Type 2 Diabetes Mellitus

Introduction

Around 20% of patients with obesity have type 2 diabetes (T2DM), which is associated with a shortened life expectancy due to cardiovascular, renal, ophthalmic, and neurological complications [1-3]. Since the observation that T2DM can be brought into remission with obesity surgery, T2DM has become one of the main indications for performing bariatric procedures such as the Roux-en-Y Gastric Bypass (RYGB), sleeve gastrectomy, or biliopancreatic diversion.

On average, remission rates after bariatric surgery vary between 30% and 90% depending on the procedure and the patient [4-8]. Nevertheless, because of the growing number of bariatric procedures performed every year, a large number of patients still fails to achieve remission. Although many variables have been correlated with T2DM remission, none are specific enough to predict T2DM remission on their own [8, 9]. Therefore, to date, several models have been developed that estimate the probability of T2DM remission after bariatric surgery based on a combination of predicting variables. Unfortunately, most of these prediction models have only been validated in specific patient populations, so their usability in different patient cohorts is often unclear.

This study aimed to compare the accuracy of prediction models for T2DM remission after RYGB surgery in a Western European patient cohort. Additionally, to evaluate their population-specific statistical performance, existing models were compared to a custom-fitted model that was specifically designed to have maximal predictive accuracy in the studied patient cohort.

Materials and methods

Literature search and model selection

Predictive scoring systems and regression models were identified through a systematic literature search in PubMed using the search term “diabetes remission AND prediction AND metabolic surgery”. No filters were used. Articles published before May 1, 2020 were screened for the presence of a scoring system or regression model intended to preoperatively predict T2DM remission after bariatric or

metabolic surgery. If a model was applied or cited, the original publication was retrieved and the original study cohort was taken into consideration.

Study cohort

A retrospective study cohort was compiled from an actively maintained database of the obesity surgery unit of the Antwerp University Hospital, localized in Western Europe (Belgium). Inclusion was limited to patients with T2DM who underwent a RYGB between January 1, 2008 and January 1, 2019, had a BMI ≥35 kg/m², a complete preoperative metabolic assessment, and a full 12-month follow-up. T2DM was defined in accordance with the ADA criteria as a fasting plasma glucose level >126 mg/dl (7 mmol/L), a 2-hr plasma glucose level >200 mg/dl (11.1 mmol/L) during an oral glucose tolerance test (OGTT), or HbA1c above 6.5% (48 mmol/mol)[10]. Furthermore, patients were only included if their records contained all required data necessary to calculate the different prediction models. All RYGB procedures were conducted using an identical protocol, consisting of the construction of a small gastric pouch (30 cc) connected to an alimentary limb of 100-120 cm and a biliopancreatic limb of 50-70 cm. Patients undergoing other surgical procedures, without diabetes, with insufficient preoperative data or incomplete follow-up were excluded.

Data collection

Data collected for this study included all variables necessary to predict T2DM remission with the prediction models found in the literature together with additional variables that were reported as potential predictors for T2DM remission. A full list of collected variables is provided in supplementary material 1. Partial and complete T2DM remission were defined in accordance with the criteria of the ADA and summarized in Table 2 of Supplementary File 1 [11]. Both definitions required that remission was sustained for at least one year. Because the majority of T2DM remission prediction models determined T2DM remission one year after bariatric surgery, our study also determined the remission status 12 months after the RYGB. Insulin resistance, insulin sensitivity, and beta cell function were calculated using the updated HOMA model [12, 13].

Complete remission was the primary outcome of this study. However, since some studies combined complete and partial T2DM remission, analyses were repeated for this definition. Furthermore, metformin therapy is sometimes continued after achieving complete T2DM remission. Therefore, a third definition of remission ("complete remission with metformin") was added. Results for the last 2 definitions are reported in the supplementary material.

This study was conducted in accordance with the Helsinki declaration and retrospective data collection was approved by the ethical committee of the Antwerp University Hospital.

Statistical analysis

Baseline and postoperative patient characteristics were statistically analyzed using a Student's t, Mann-Whitney U, or Chi-Square test when appropriate. T2DM remission scores were calculated for each patient and each prediction model. Beta values of logistic regression models were recalibrated for this study.

The custom prediction model which indicated the maximal predictive potential of a T2DM remission model in the study cohort was constructed based on a two-stage selection process. In the first stage, the link between all preoperative variables and T2DM remission was univariately analyzed using a logistic regression model, only retaining variables with $p<0.010$. In the second stage of the model construction, a multivariate logistic regression analysis using a forward stepwise selection strategy based on the Wald coefficient was conducted to construct the final T2DM prediction model. In this analysis, predictors for T2DM were ranked based on their Wald coefficient, with higher coefficients indicating more importance of the predictor in the fit of the model. Subsequently, predictors were removed until the point when the goodness of fit of the model started to drop. Outcomes of the regression model were presented as the beta value together with the corresponding Wald coefficient, odds ratios (OR), and 95% confidence intervals (95% CI).

For each of the existing models, the sensitivity and specificity were plotted in receiver operating characteristic (ROC) curves. The area under the ROC curve (AUROC) was calculated using the method described by DeLong et al. [14]. The statistical performance of the different remission models was assessed by comparing the accuracy, predicted-to-observed ratio, and AUROC values between the existing models and with the custom model. Cut-off values were determined using the Youden index. The goodness of fit of the different logistic regression models was determined with the Hosmer-Lemeshow statistical test. A p-value below $p<0.05$ indicated a lack of fit. Statistical analyses were conducted using SPSS (Version 26, IBM Corporation Armonk, NY, USA) and R statistical software (Version 3.6.3, package pROC, R Core Team 2020).

Results

Literature search and model overview

The literature search identified 172 articles, describing 12 predictive scoring models and 8 predictive logistic regression models [8, 15-32]. The model by Ugale et al. was excluded since it required a mixed-meal test before surgery which is not routinely performed in our patients [31]. An overview of the 19 models and their predictor variables is provided in Table 1 and in more detail in Supplementary File 1.

Patient characteristics

Between January 1, 2008 and January 1, 2019 a total of 1525 patients underwent a RYGB of whom 359 (23.5%) had T2DM before surgery (Figure 1). Of these patients, 109 were excluded because of a missing C-peptide measurement (n=50), an unknown remission status (n=21), or an undocumented T2DM duration (n=12). Another 26 patients were lost to follow-up or were postoperatively followed elsewhere. As a result, the final study cohort consisted of 250 patients.

After one year, 160 (64.0%) patients achieved complete T2DM remission, 8 (3.2%) had partial remission, and 14 (5.6%) patients were biochemically in complete remission but were still treated with

metformin (Table 2, Supplementary File 2). Patients with complete remission had a significantly shorter preoperative diabetes duration, better metabolic control (e.g. lower HbA1c levels), more weight loss, and required less antidiabetic medication (Table 2). T2DM duration, insulin use, the number of antidiabetic drugs, fasting glucose levels, C-peptide levels, and HbA1c values strongly correlated with each other ($p<0.001$). Although weight loss after bariatric surgery cannot be used as a preoperative predictor, the amount of weight loss proved to be strongly associated with T2DM remission ($p<0.001$). Weight loss alone correctly predicted T2DM remission in 71.2% of patients (AUROC 72.7%), indicating that other variables were involved as well. Models which incorporate %TWL, together with conversions to other definitions of T2DM and units are provided in Supplementary File 4 and 5.

The mean remission scores of patients with and without diabetes remission differed significantly for each model ($p<0.001$, Table 3).

Construction of a reference model to predict T2DM remission in the study cohort.

To assist in the evaluation of the existing prediction models, a custom logistic regression model was constructed based on the characteristics of the patient cohort (Variables included and their impact on the model are displayed in Table 4). The most important predictors were (in descending order): diabetes duration, insulin use, the number of antidiabetics used and % HbA1c:

$$\begin{aligned} Log(Odds) = \; & 5.707 \\ & - 0.153 \times diabetes\ duration\ (years) \\ & - 0.276\ \times HbA1c\ (\%) \\ & - 1.434\ (if\ on\ insulin\ ) \\ & - \begin{vmatrix} 2.039\ (if\ treatment\ with\ 1\ antidiabetic\ drug) \\ 2.599\ (if\ treatment\ with\ \geq 2\ antidiabetic\ drugs) \end{vmatrix} \end{aligned}$$

This model correctly predicted T2DM remission in 83.6% of patients with a positive and negative predictive value of respectively 84.8% and 81.0%, a sensitivity of 90.6%, specificity of 71.1%, and AUROC of 0.913. After adding cut-off points to the model, the accuracy increased to 84.0% and the AUROC to 0.917 (Figure 3a). The model overestimated T2DM remission in only 5% of patients. A conversion of the model to a scoring tool can be found in Supplementary File 3.

$$
\begin{aligned}
Log(Odds) = \; & 5.165 \\
& - \begin{cases} 0.852\ (if\ T2DM\ duration\ 2-3\ years) \\ 2.123\ (if\ T2DM\ duration \geq 4\ years) \end{cases} \\
& - \begin{cases} 1.190\ (if\ HbA1c\ (\%)\ 6-7\% \\ 2.643\ (if\ HbA1c\ (\%)\ 7-8\% \\ 1.275\ (if\ HbA1c\ (\%)\ 8-9\% \\ 2.428\ (if\ HbA1c\ (\%) \geq 9\% \end{cases} \\
& - \mid 1.755\ (if\ on\ insulin\ ) \\
& - \begin{cases} 1.594\ (if\ treatment\ with\ 1\ antidiabetic\ drug) \\ 1.782\ (if\ treatment\ with\ \geq 2\ antidiabetic\ drugs) \end{cases}
\end{aligned}
$$

Assessment of the performance of existing models for T2DM remission by comparison with the custom model.

Existing prediction scoring models had an average AUROC between 75.7% and 91.2% (Figure 2). Similar AUROC values were seen for the logistic regression models, ranging between 63.0% and 90.2% (Figure 3). Models always had a higher sensitivity than specificity and tended to overestimate T2DM remission rates as shown by the predicted-to-observed ratio (Table 5). The general accuracy of the models was high, with most models correctly predicting diabetes remission in more than 80% of cases.

Compared to the constructed model (Figure 3a), the IMS score, DiaBetter score, Ad-DiaRem score, and DiaRem score best predicted T2DM remission with a difference in AUROC of less than 2% (Table 5)[15, 16, 24, 30]. Similarly, the logistic regression models of Stallard et al., Ramos Levi et al., and Hayes et al. predicted T2DM remission with the highest accuracy (Table 5)[18, 25, 28]. The predictive power of most models was not significantly altered if other definitions of T2DM remission were used (Supplementary File 6).

Discussion

Numerous models to predict T2DM remission have been reported in the literature and, due to the large availability of models, it is often unclear for clinicians which model is best suited to predict diabetes remission in their specific population.

Our results indicate that most of the existing prediction models accurately forecast T2DM remission in a European cohort of bariatric surgery candidates. Although most prediction models include different predictor variables, the predictive performance remains largely comparable. This phenomenon can be

explained by the presence of collinearity. Primary variables such as diabetes duration, HbA1c levels, C-peptide, or the need for diabetes medication (e.g. insulin) reflect the same impairment of the beta cell, the adipose tissue, and the obesity-related insulin resistance. Therefore, these predicting variables are strongly correlated and thus largely interchangeable without significantly affecting the accuracy of the model. Moreover, because they are strong predictors, the inclusion of only some of them will quickly increase the model's predictive potential. The best performance is seen in models that include 3 (or more) diabetes-related predictor variables. From this point, further gains in performance are mostly achieved by adding secondary, non-T2DM-related variables to the model. These secondary variables mostly reflect differences in basic patient characteristics (e.g. age, gender, BMI). These variables do not raise the predictive power of the model as much as the primary variables, but they help to maximize the accuracy.

As part of the secondary aim of this study, we evaluated how well the existing models were able to predict T2DM in the studied patients' cohort. We expected it to be impossible for any model to predict T2DM remission with an accuracy of 100% due to a large number of unpredictable and missing preoperative and postoperative variables. Therefore, starting from over 60 variables, we devised a custom-fitted model with the intent of having the model with the highest achievable predictive performance in the studied cohort. By comparing the existing models with this custom model, we were able to evaluate how well each model performed.

Especially the IMS score, DiaBetter score, Ad-DiaRem score and DiaRem score closely matched the performance of the custom-designed model. Other models (e.g. score by Ioffe, score by Umemura) predicted T2DM remission up to 18% less accurate and overestimating remission by up to 50% prediction errors (custom model only 5%). While this indicates that most models performed generally well in our study, this also suggests that the validation of T2DM models is advised before models can be translated to other patient cohorts with potentially different compositions. It is also for this reason that we do not recommend that our statistical model is clinically used as a T2DM remission prediction model.

Another observation of this study is the strong correlation of weight loss with T2DM remission. This can most likely be explained by the reduction of insulin resistance due to the loss of visceral adiposity, reduced glucose production and better beta cell function. Recently, identical metabolic improvements were seen after weight loss caused by a RYGB or a diet alone [33]. Hence, long-term weight loss should be pursued in every patient with T2DM either by weight-loss surgery or by dietary interventions.

To date, mostly studies performed in Asia have compared different prediction models for T2DM remission [34-38]. These studies are generally difficult to translate to patient cohorts elsewhere in the world due to differences in patient characteristics (e.g. BMI, age), genetic backgrounds, metabolic factors, environmental effects, and treatment modalities. In a recent study in Taiwan, Shen et al. studied different predictive models in patients undergoing a sleeve gastrectomy [35]. In contrast to our results, they observed a higher discriminative power of the ABCD score, while the score by Robert et al. performed much worse. Similarly, Chen et al. reported better predictions with the ABCD score compared to the IMS score [38].

The use of models to predict T2DM remission is subject to some limitations. Because the majority of models predict the T2DM remission status one year after surgery, we also limited our focus to the first postoperative year. While this was methodologically necessary to correctly compare the different models, it prevented us from drawing any conclusion about the long-term performance of these models. In the long term, we expect prediction models to be less accurate because of the increasing effects of external factors (e.g. treatment, lifestyle, genetics, dietary habits) over time. This was recently demonstrated in a comparative study by Dicker et al. [37].

Secondly, these models are not explanatory. They predict an outcome but do not explain the underlying mechanisms which lead to T2DM remission. T2DM remission is likely linked to a large number of variables, of which most are not even included in the current models. Factors such as physical activity, genetic background, biochemical changes after surgery (e.g. inflammation), and altered metabolism of the adipose tissue all affect the glucose homeostasis before and after bariatric

surgery [39-41]. Conversely, because these models do not indicate how remission is achieved, they also fail to explain why T2DM persists in some patients. A long duration of T2DM or insufficient weight loss are undoubtedly associated with T2DM persevering after bariatric surgery. Yet, some patients in our study seemingly having an ideal profile still failed to achieve remission without a clear cause. To address this issue, future models should determine T2DM according to the profile of patients, which is based on functional, biochemical, metabolic and genetic factors. Such models require the inclusion of large patient cohorts, sampled internationally, and with the collection of a large number of variables.

Finally, most of the current models are static. Continuing advancements in the management of T2DM and new treatment options may necessitate adaptations to the models over time.

Conclusion

This study demonstrated that most of the existing models to predict T2DM remission have a high accuracy in a population of patients undergoing bariatric surgery in Europe. Nevertheless, since the accuracy of these models can differ between patient cohorts, validation before clinically use is recommended. To date, most models focus on short-term remission which limits their clinical use. Newer models should focus on predicting long-term T2DM remission based on more specific metabolic profiles and variables.

Statements:

- **Funding:** No funding was obtained for this study.
- **Conflict of interest:** All authors declare that they have no conflict of interest.
- **Ethical approval:** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This is a retrospective study. For this type of study formal consent is not required.
- **Informed consent:** Informed Consent does not apply.

References

1. Bhupathiraju SN, Hu FB. Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. Circ Res. 2016;118(11):1723-35.

2. Clarke PM, Gray AM, Briggs A, Farmer AJ, Fenn P, Stevens RJ, et al. A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). Diabetologia. 2004;47(10):1747-59.

3. Gray N, Picone G, Sloan F, Yashkin A. Relation between BMI and diabetes mellitus and its complications among US older adults. South Med J. 2015;108(1):29-36.

4. Kodama S, Fujihara K, Horikawa C, Harada M, Ishiguro H, Kaneko M, et al. Network meta-analysis of the relative efficacy of bariatric surgeries for diabetes remission. Obesity Reviews. 2018;19(12):1621-9.

5. Madadi F, Jawad R, Mousati I, Plaeke P, Hubens G. Remission of Type 2 Diabetes and Sleeve Gastrectomy in Morbid Obesity: a Comparative Systematic Review and Meta-analysis. Obesity Surgery. 2019;29(12):4066-76.

6. Purnell JQ, Selzer F, Wahed AS, Pender J, Pories W, Pomp A, et al. Type 2 Diabetes Remission Rates After Laparoscopic Gastric Bypass and Gastric Banding: Results of the Longitudinal Assessment of Bariatric Surgery Study. Diabetes Care. 2016;39(7):1101.

7. Sjöholm K, Pajunen P, Jacobson P, Karason K, Sjöström CD, Torgerson J, et al. Incidence and remission of type 2 diabetes in relation to degree of obesity at baseline and 2 year weight change: the Swedish Obese Subjects (SOS) study. Diabetologia. 2015;58(7):1448-53.

8. Panunzi S, Carlsson L, De Gaetano A, Peltonen M, Rice T, Sjostrom L, et al. Determinants of Diabetes Remission and Glycemic Control After Bariatric Surgery. Diabetes Care. 2016;39(1):166-74.

9. Park JY. Prediction of Type 2 Diabetes Remission after Bariatric or Metabolic Surgery. J Obes Metab Syndr. 2018;27(4):213-22.

10. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. Diabetes Care. 2020;43(Supplement 1):S14.

11. Buse JB, Caprio S, Cefalu WT, Ceriello A, Del Prato S, Inzucchi SE, et al. How Do We Define Cure of Diabetes? Diabetes Care. 2009;32(11):2133.

12. Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care. 1998;21(12):2191-2.

13. Wallace TM, Levy JC, Matthews DR. Use and Abuse of HOMA Modeling. Diabetes Care. 2004;27(6):1487.

14. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.

15. Aminian A, Brethauer SA, Andalib A, Nowacki AS, Jimenez A, Corcelles R, et al. Individualized Metabolic Surgery Score: Procedure Selection Based on Diabetes Severity. Ann Surg. 2017;266(4):650-7.

16. Aron-Wisnewsky J, Sokolovska N, Liu Y, Comaneshter DS, Vinker S, Pecht T, et al. The advanced-DiaRem score improves prediction of diabetes remission 1 year post-Roux-en-Y gastric bypass. Diabetologia. 2017;60(10):1892-902.

17. Dixon JB, Chuang L-M, Chong K, Chen S-C, Lambert GW, Straznicky NE, et al. Predicting the Glycemic Response to Gastric Bypass Surgery in Patients With Type 2 Diabetes. Diabetes Care. 2013;36(1):20.

18. Hayes MT, Hunt LA, Foo J, Tychinskaya Y, Stubbs RS. A model for predicting the resolution of type 2 diabetes in severely obese subjects following Roux-en Y gastric bypass surgery. Obes Surg. 2011;21(7):910-6.

19. Ioffe OY, Kryvopustov MS, Dibrova YA, Tsiura YP. Type 2 diabetes mellitus remission and its prediction after two-stage surgical treatment of patients with morbid obesity. Wiad Lek. 2019;72(5 cz 1):739-43.

20. Lee WJ, Almulaifi A, Tsou JJ, Ser KH, Lee YC, Chen SC. Laparoscopic sleeve gastrectomy for type 2 diabetes mellitus: predicting the success by ABCD score. Surg Obes Relat Dis. 2015;11(5):991-6.

21. Lee WJ, Hur KY, Lakadawala M, Kasama K, Wong SK, Chen SC, et al. Predicting success of metabolic surgery: age, body mass index, C-peptide, and duration score. Surg Obes Relat Dis. 2013;9(3):379-84.

22. Moh MC, Cheng A, Tan CH, Lim BK, Tan BC, Ng D, et al. Metabolic Surgery Diabetes Remission (MDR) Score: a New Preoperative Scoring System for Predicting Type 2 Diabetes Remission at 1 Year After Metabolic Surgery in the Singapore Multi-ethnic Asian Setting. Obesity Surgery. 2020.

23. Park JY, Kim YJ. Prediction of Diabetes Remission in Morbidly Obese Patients After Roux-en-Y Gastric Bypass. Obes Surg. 2016;26(4):749-56.

24. Pucci A, Tymoszuk U, Cheung WH, Makaronidis JM, Scholes S, Tharakan G, et al. Type 2 diabetes remission 2 years post Roux-en-Y gastric bypass and sleeve gastrectomy: the role of the weight loss and comparison of DiaRem and DiaBetter scores. Diabet Med. 2018;35(3):360-7.

25. Ramos-Levi AM, Matia P, Cabrerizo L, Barabash A, Sanchez-Pernaute A, Calle-Pascual AL, et al. Statistical models to predict type 2 diabetes remission after bariatric surgery. J Diabetes. 2014;6(5):472-7.

26. Robert M, Ferrand-Gaillard C, Disse E, Espalieu P, Simon C, Laville M, et al. Predictive factors of type 2 diabetes remission 1 year after bariatric surgery: impact of surgical techniques. Obes Surg. 2013;23(6):770-5.

27. Scopinaro N, Adami GF, Bruzzi P, Cordera R. Prediction of Diabetes Remission at Long Term Following Biliopancreatic Diversion. Obes Surg. 2017;27(7):1705-8.

28. Stallard R, Sahai V, Drover JW, Chun S, Keresztes C. Defining and Using Preoperative Predictors of Diabetic Remission Following Bariatric Surgery. JPEN J Parenter Enteral Nutr. 2018;42(3):573-80.

29. Still CD, Benotti P, Mirshahi T, Cook A, Wood GC. DiaRem2: Incorporating duration of diabetes to improve prediction of diabetes remission after metabolic surgery. Surgery for Obesity and Related Diseases. 2019;15(5):717-24.

30. Still CD, Wood GC, Benotti P, Petrick AT, Gabrielsen J, Strodel WE, et al. Preoperative prediction of type 2 diabetes remission after Roux-en-Y gastric bypass surgery: a retrospective cohort study. Lancet Diabetes Endocrinol. 2014;2(1):38-45.

31. Ugale S, Gupta N, Modi KD, Kota SK, Satwalekar V, Naik V, et al. Prediction of remission after metabolic surgery using a novel scoring system in type 2 diabetes - a retrospective cohort study. J Diabetes Metab Disord. 2014;13(1):89-.

32. Umemura A, Sasaki A, Nitta H, Nikai H, Baba S, Takahara T, et al. Prognostic factors and a new preliminary scoring system for remission of type 2 diabetes mellitus after laparoscopic sleeve gastrectomy. Surgery Today. 2020.

33. Yoshino M, Kayser BD, Yoshino J, Stein RI, Reeds D, Eagon JC, et al. Effects of Diet versus Gastric Bypass on Metabolic Function in Diabetes. New England Journal of Medicine. 2020;383(8):721-32.

34. Kam H, Tu Y, Pan J, Han J, Zhang P, Bao Y, et al. Comparison of Four Risk Prediction Models for Diabetes Remission after Roux-en-Y Gastric Bypass Surgery in Obese Chinese Patients with Type 2 Diabetes Mellitus. Obesity Surgery. 2020;30(6):2147-57.

35. Shen SC, Wang W, Tam KW, Chen HA, Lin YK, Wang SY, et al. Validating Risk Prediction Models of Diabetes Remission After Sleeve Gastrectomy. Obes Surg. 2019;29(1):221-9.

36. Ha J, Kwon Y, Kim NH, Park S, Menzo EL, Rosenthal RJ. Discordance in prediction for prognosis of type 2 diabetes after metabolic surgery: comparison of the ABCD, DiaRem, and individualized metabolic surgery models. Ann Surg Treat Res. 2019;97(6):309-18.

37. Dicker D, Golan R, Aron-Wisnewsky J, Zucker JD, Sokolowska N, Comaneshter DS, et al. Prediction of Long-Term Diabetes Remission After RYGB, Sleeve Gastrectomy, and Adjustable Gastric Banding Using DiaRem and Advanced-DiaRem Scores. Obes Surg. 2019;29(3):796-804.

38. Chen JC, Hsu NY, Lee WJ, Chen SC, Ser KH, Lee YC. Prediction of type 2 diabetes remission after metabolic surgery: a comparison of the individualized metabolic surgery score and the ABCD score. Surg Obes Relat Dis. 2018;14(5):640-5.

39. Blundell JE, Dulloo AG, Salvador J, Frühbeck G, on behalf of the Easo Sab Working Group on Bmi G. Beyond BMI - Phenotyping the Obesities. Obesity Facts. 2014;7(5):322-8.

40. Catalán V, Gómez-Ambrosi J, Rotellar F, Silva C, Rodríguez A, Salvador J, et al. Validation of endogenous control genes in human adipose tissue: relevance to obesity and obesity-associated type 2 diabetes mellitus. Horm Metab Res. 2007;39(7):495-500.

41. Unamuno X, Gómez-Ambrosi J, Rodríguez A, Becerril S, Frühbeck G, Catalán V. Adipokine dysregulation and adipose tissue inflammation in human obesity. Eur J Clin Invest. 2018;48(9):e12997.

Tables

Table 1 – Overview of the different diabetes remission models

Remission prediction model	Year	Variables
Predictive Scoring models		
ABCD score v1 [21]	2013	Age, BMI, C-peptide, diabetes duration
Score by Robert et al. [26]	2013	BMI, diabetes duration, fasting glucose, HbA1c, oral anti-diabetic agents (not insulin)
Score by Dixon et al. [17]	2013	C-peptide, diabetes duration, weight loss*
DiaRem score [30]	2014	Age, HbA1c, insulin use, other antidiabetic drugs (not metformin)
ABCD score v2 [20]	2015	Age, BMI, C-peptide, diabetes duration
IMS score [15]	2017	Diabetes duration, HbA1c, insulin use, number of diabetes drugs
Ad-DiaRem score [16]	2017	Age, diabetes duration, HbA1c, insulin use, other glucose-lowering agents, number of glucose-lowering agents
DiaBetter score [24]	2018	Diabetes duration, HbA1c, type of diabetes medication
DiaRem2 score [29]	2019	Age, HbA1c, insulin use, diabetes duration
Score by Umemura et al. [32]	2020	Fasting insulin, diabetes duration
MDR score [22]	2020	Age, HOMA2 beta cell function, diabetes duration, HbA1c
Predictive regression models		
Hayes et al. [18]	2011	HbA1c, insulin use
Dixon et al. [17]	2013	BMI, C-peptide, diabetes duration
Ramos-Levi et al. [25]	2014	Age, C-peptide, diabetes duration, fasting glucose, gender, insulin use
Panunzi et al. [8]	2016	BMI, diabetes duration, fasting glucose
Park et al. [23]	2016	Age, BMI, C-peptide, HbA1c, insulin use
Scopinaro et al. [27]	2017	BMI, insulin use
Stallard et al. [28]	2017	Diabetes duration, fasting glucose, insulin use, number of diabetes medication other than insulin
Ioffe et al. [19]	2019	C-peptide

Overview of the variables included in the different scoring systems and predictive logistic regression models. Two versions of the ABCD score with revised score assignments have been published to date. *postoperative variable. Additional details for each model are available in the supplementary material (Supplementary Material 1). IMS, individualized metabolic score; MDR score, metabolic surgery diabetes remission score.

Table 2 – Patient characteristics

Parameters ***Mean ± SD or N (%)***	**Total cohort (n=250)**	**Complete remission (n=160)**	**No remission (n=90)**	**p-value**
Age *(years)*	50.8 ± 9.7	49.9 ± 10.7	52.3 ± 7.5	0.035
Female gender	142 (56.8)	92 (57.5)	50 (55.6)	0.766
Smoking	58 (23.2)	40 (25.0)	18 (20.0)	0.369
Anthropometric data				
Weight *(kg)*	118.6 ± 21.0	119.0 ± 20.7	117.9 ± 21.5	0.681
Length *(cm)*	168.9 ± 10.1	168.6 ± 9.5	169.5 ± 11.1	0.501
BMI *(kg/m²)*	41.5 ± 5.9	41.8 ± 5.7	41.0 ± 6.2	0.302
Diabetes characteristics				
Diabetes duration *(years)*	4.1 ± 5.8	1.6 ± 3.2	8.6 ± 6.8	<0.001
HbA1c *(%)*	7.2 ± 1.6	6.6 ± 1.2	8.2 ± 1.6	<0.001
Fasting C-peptide *(ng/mL)*	4.4 ± 2.0	4.7 ± 1.8	4.0 ± 2.3	0.008
Fasting plasma glucose *(mg/dL)*	143.4 ± 59.9	125.8 ± 47.1	174.7 ± 67.2	<0.001
OGTT glucose after 2 hrs. *(mg/dL)*	229.5 ± 67.5	226.3 ± 60.2	236.1 ± 80.6	0.362
Fasting blood insulin *(mU/L)*	30.4 ± 28.4	32.7 ± 39.6	29.6 ± 23.0	0.585
OGTT insulin after 2 hrs. *(mU/L)*	159.4 ± 127.0	180.9 ± 127.2	86.9 ± 96.9	<0.001
HOMA2 Insulin resistance	3.9 ± 2.9	3.9 ± 1.8	4.1 ± 4.2	0.683
HOMA2 % Beta cell function *(%)*	121.8 ± 64.8	146.1 ± 60.3	78.4 ± 47.6	<0.001
HOMA2 % Insulin sensitivity *(%)*	36.7 ± 53.6	30.0 ± 11.6	48.8 ± 87.1	0.044
Microvascular complications	20 (8.0)	2 (1.3)	18 (20.0)	<0.001
Macrovascular complications	26 (10.4)	9 (5.6)	17 (18.9)	0.001
Diabetes treatment				
Metformin	187 (74.8)	107 (66.9)	80 (88.9)	<0.001
Sulphonyl urea	43 (17.2)	17 (10.6)	26 (28.9)	<0.001
GLP-1 agonists	14 (5.6)	7 (4.4)	7 (7.8)	0.261
Insulin	64 (25.6)	9 (5.6)	55 (61.1)	<0.001
Lipid metabolism				
Triglycerides *(mg/dL)*	203.9 ± 151.2	199.6 ± 161.9	211.5 ± 130.7	0.551
Total cholesterol *(mg/dL)*	184.9 ± 42.1	191.8 ± 41.3	172.7 ± 41.0	0.001
High-density lipoprotein *(mg/dL)*	43.4 ± 12.2	44.2 ± 12.2	42.1 ± 12.1	0.191
Statin treatment	84 (33.6)	45 (28.1)	39 (43.3)	0.015
Postoperative outcomes (12 months)				
BMI *(kg/m²)*	30.8 ± 5.3	30.1 ± 5.1	32.2 ± 5.5	0.003
%TWL *(%)*	25.5 ± 8.5	27.9 ± 7.9	21.4 ± 7.9	<0.001
%EBMIL *(%)*	67.7 ± 24.6	73.1 ± 23.6	58.1 ± 23.3	<0.001
Triglycerides *(mg/dL)*	126.2 ± 64.5	111.7 ± 48.3	149.0 ± 78.9	<0.001
Total cholesterol *(mg/dL)*	162.9 ± 32.4	165.8 ± 32.4	158.4 ± 32.1	0.122
High-density lipoprotein *(mg/dL)*	51.2 ± 15.1	53.1 ± 16.3	48.4 ± 12.9	0.038
Complete remission	160 (64.0)			
Partial remission	8 (3.2)			
Complete remission with continued preventive metformin therapy	14 (5.6)			

Patient characteristics are displayed for the total cohort and based on the diabetes remission status. Results are displayed as the mean ± standard deviation (SD) or the number with percentage. BMI, body mass index; OGTT, oral glucose tolerance test; %TWL, percentage total weight loss; %EBMIL, percentage excess BMI loss. Statistical analysis with a student's t-test or Chi-Square test. Details on the patient characteristics for the combined complete and partial remission group and group preventively treated with metformin after surgery is provided in Supplementary Material 2.

Table 3 – Implementation of the different diabetes remission prediction models in the study cohort.

A. Predictive scoring models				
	Complete remission			
	Yes (n= 160)	No (n= 90)		Wald Coëfficient
Predictive scoring models	Median [Min-Max; IQR]	Median [Min-Max; IQR]	p-value	
ABCD score v1 [21]	7 [3-10; 2]	5 [2-9; 2]	**<0.001**	39.5
Score by Robert et al. [26]	4 [1-5; 2]	2 [0-5; 2]	**<0.001**	71.1
Score by Dixon et al. [17]	3 [1-3; 1]	1 [0-3; 1]	**<0.001**	59.9
DiaRem score [30]	2.5 [0-20; 4]	15 [1-21; 9]	**<0.001**	72.8
ABCD score v2 [20]	7 [3-10; 2]	5 [2-9; 2]	**<0.001**	49.8
IMS score [15]	16 [0-130; 26]	93 [13-149; 57]	**<0.001**	74.2
Ad-DiaRem score [16]	5 [0-21; 2]	13 [4-21; 8]	**<0.001**	71.2
Diabetter score [24]	1 [0-8; 1]	6.5 [1-9; 4]	**<0.001**	78.4
DiaRem2 score [29]	2 [0-25; 3]	15 [1-25; 11.7]	**<0.001**	68.0
Score by Umemura et al. [32]	2 [0-2; 0]	1 [0-2; 1]	**<0.001**	19.9
MDR score [22]	8 [2-10; 2]	4.5 [0-9; 3]	**<0.001**	68.7

B. Predictive logistic regression models	
	Equation
Hayes et al. [18]	Log(odds) = 4.902 – (0.497 x HbA1c %) – (2.679 x insulin use)
Dixon et al. [17]	Log(odds) = 1.265 + (0.016 x BMI) – (0.028 x C-peptide) – (0.301 x diabetes duration)
Ramos-Levi et al. [25]	Log(odds) = 4.291 – (0.009 x age) – (0.089 x C-peptide) – (0.205 x diabetes duration) – (0.472 x female gender) – (0.009 x fasting glucose) – (1.972 x insulin use)
Panunzi et al. [8]	Log(odds) = 2.919 + (0.013 x BMI) – (0.273 x diabetes duration) – (0.012 x fasting glucose)
Park et al. [23]	Log(odds) = 5.857 – (0.025 x age) + (0.012 x BMI) – (0.046 x C-peptide) – (0.489 x HbA1c %) – (2.723 x insulin use)
Scopinaro et al. [27]	Log(odds) = 0.837 + (0.015 x BMI) – (3.264 x insulin use)
Stallard et al. [28]	Log(odds) = 3.276 – (1.509 x diabetes duration >4 years) – (0.005 x fasting glucose) – (2.466 x insulin use) – (0.639 x number of antidiabetic drugs other than insulin)
Ioffe et al. [19]	Log(odds) = -0.353 + (0.216 x C-peptide)

The median score for patients with and without T2DM remission after RYGB surgery for each score is displayed in part A. All scores significantly differed between both groups. The Wald statistic is provided for each remission score, indicating how well the scoring model explained T2DM remission. Logistic regression model equations, calibrated for the studied patient cohort, are displayed in part B of the table. Units used in the equations: C-peptide (ng/mL), glucose (mg/dL), BMI (kg/m²), HbA1c (%). For conversions to other units and to other remission definitions see supplementary file 5.

Table 4 – Construction of a custom logistic regression model.

	Univariate Analysis				Multivariate Analysis			
			Exp(*B*)				Exp(*B*)	
	B	**Wald**	**OR (95% CI)**	**p-value**	***B***	**Wald**	**OR (95% CI)**	**p-value**
Patient characteristics								
Age	-0.270	3.6	0.97 (0.95-1.00)	0.056				
Diabetes characteristics								
Diabetes duration	-0.299	52.7	0.74 (0.68-0.80)	<0.001	-0.153	9.5	0.86 (0.78-0.95)	0.002
Fasting C-peptide	0.216	7.7	1.24 (1.07-1.45)	0.005				
Fasting plasma glucose	-0.016	29.4	0.984 (0.979-0.990)	<0.001				
OGTT insulin after 2 hrs.	0.010	14.0	1.010 (1.005-1.016)	<0.001				
HOMA2 % Beta cell function	0.024	51.4	1.024 (1.018-1.031)	<0.001				
HOMA2 % Insulin sensitivity	-0.024	11.5	0.976 (0.962-0.990)	0.001				
HbA1c %	-0.809	42.9	0.45 (0.35-0.57)	<0.001	-0.276	4.4	0.76 (0.59-0.98)	0.036
Microvascular complications (Ref= no)	-2.983	15.5	0.051 (0.011-0.224	<0.001				
Macrovascular complications (Ref= no)	-1.363	9.8	0.26 (0.11-0.60)	0.002				
Diabetes treatment								
Diabetes treatment status (Ref= no)	-3.641	12.7	0.026 (0.004-0.194)	<0.001				
Number of antidiabetic drugs (Ref= 0)								
1	-2.586	6.3	0.075 (0.010-0.572)	0.012	-2.039	3.8	0.13 (0.02-1.02)	0.052
≥2	-4.949	22.6	0.007 (0.001-0.055)	<0.001	-2.599	5.4	0.074 (0.008-0.671)	0.021
Insulin use (Ref= no insulin)	-3.272	65.1	0.038 (0.017-0.084)	<0.001	-1.434	8.3	0.238 (0.090-0.633)	0.004
Metformin use (Ref= no metformin)	-1.377	13.5	0.25 (0.12-0.53)	<0.001				
Sulphonyl urea use (Ref= no SU)	-1.229	12.6	0.29 (0.15-0.58)	<0.001				
Lipid metabolism								
Total cholesterol	0.012	11.2	1.012 (1.005-1.019)	0.001				
Statin use (Ref=no)	-0.670	5.9	0.51 (0.30-0.88)	0.015				

Influence of the studied variables on complete remission in a univariate and multivariate logistic regression model. No significant effects of weight, length, preoperative BMI, GLP-1 analog use, fasting insulin levels, OGTT glucose levels after 12 minutes, triglycerides, or high-density lipoprotein levels were observed in the univariate analysis (not displayed). Only variables with a significance level below 0.10 were entered in a multivariate logistic regression analysis and selected using a forward stepwise selection strategy based on the probability of the likelihood-ratio statistical parameter. Following the analysis, only diabetes duration, percentage HbA1c, number of antidiabetic drugs, and insulin were retained in the final model. OGTT, oral glucose tolerance test; OR, odds ratio; 95% CI, 95% confidence interval.

Table 5 – Statistical comparison of the predictive performance of scoring models and logistic regression models.

Model	Hosmer-Lemeshow Chi-square	Hosmer-Lemeshow p-value	Model Sensitivity (%, 95% CI)	Model Specificity (%, 95% CI)	Model Accuracy (%, 95% CI)	Predicted-to-observed ratio	AUROC ratio compared to constructed model (p-value)
Logistic model constructed for the study cohort (reference)	1.509	0.993	90.0 (84.3-94.2)	73.3 (63.0-82.1)	84.0 (78.9-88.3)	1.05	1.00 (Reference)
Predictive scoring models							
ABCD score v1 [21]	13.576	0.059	81.9 (75.0-87.5)	56.7 (45.8-67.1)	72.8 (66.8-78.2)	1.06	**0.83 (p<0.001)**
Score by Robert et al. [26]	1.274	0.973	90.6 (85.0-94.7)	68.9 (58.3-78.2)	82.8 (77.5-87.3)	1.08	**0.95 (p=0.004)**
Score by Dixon et al. [17]	6.047	0.418	86.5 (80.0-91.4)	62.5 (51.5-72.6)	77.8 (72.0-82.8)	1.08	**0.91 (p<0.001)**
DiaRem score [30]	10.404	0.238	93.1 (88.0-96.5)	65.6 (54.8-75.3)	83.2 (78.0-87.6)	1.13	0.98 (p=0.142)
ABCD score v2 [20]	4.280	0.831	86.3 (79.9-91.2)	57.8 (46.9-68.1)	76.0 (70.2-81.2)	1.10	**0.88 (p<0.001)**
IMS score [15]	7.698	0.360	91.3 (85.8-95.1)	70.0 (59.4-79.2)	83.6 (78.4-88.0)	1.08	0.99 (p=0.558)
Ad-DiaRem score [16]	18.545	0.017	89.4 (83.5-93.7)	71.1 (60.6-80.2)	82.8 (77.5-87.3)	1.06	0.98 (p=0.237)
Diabetter score [24]	9.256	0.235	87.5 (81.4-92.2)	72.2 (61.8-81.2)	82.0 (76.7-86.6)	1.03	0.99 (p=0.273)
DiaRem2 score [29]	14.798	0.063	89.4 (83.5-93.7)	71.1 (60.6-80.2)	82.8 (77.5-87.3)	1.06	**0.96 (p=0.016)**
Score by Umemura et al. [32]	18.589	<0.001	90.5 (84.5-94.7)	21.4 (11.6-34.4)	71.4 (64.7-77.5)	1.20	**0.75 (p<0.001)**
MDR score [22]	14.399	0.045	88.8 (82.8-93.2)	70.0 (59.4-79.2)	82.0 (76.7-86.6)	1.06	**0.96 (p=0.003)**
Predictive regression models							
Hayes et al. [18]	18.167	0.020	93.1 (88.0-96.5)	65.6 (54.8-75.3)	83.2 (78.0-87.6)	1.13	0.97 (p=0.101)
Dixon et al. [17]	19.354	0.013	92.5 (87.3-96.1)	64.4 (53.7-74.3)	82.4 (77.1-86.9)	1.13	**0.94 (p=0.010)**
Ramos-Levi et al. [25]	10.375	0.240	92.5 (87.3-96.1)	64.4 (53.7-74.3)	82.4 (77.1-86.9)	1.13	0.98 (p=0.276)
Panunzi et al. [8]	29.668	<0.001	91.3 (85.8-95.1)	68.9 (58.3-78.2)	83.2 (78.0-87.6)	1.09	**0.96 (p=0.030)**
Park et al. [23]	21.094	0.007	93.1 (88.0-96.5)	65.6 (54.8-75.3)	83.2 (78.0-87.6)	1.13	0.97 (p=0.061)
Scopinaro et al. [27]	4.476	0.812	94.4 (89.6-97.4)	61.1 (50.3-71.2)	82.4 (77.1-86.9)	1.16	**0.86 (p<0.001)**
Stallard et al. [28]	9.005	0.342	90.0 (84.3-94.2)	73.3 (63.0-82.1)	84.0 (78.9-88.3)	1.05	0.98 (p=0.117)
Ioffe et al. [19]	30.819	<0.001	100.0 (97.7-100.0)	14.4 (7.9-23.4)	69.2 (63.1-74.9)	1.48	**0.69 (p<0.001)**

The performance of each model was compared to the logistic regression model that was constructed for our study cohort. The goodness-of-fit, which assesses whether the observed T2DM remission rates matches the expected remission rates, was determined by the Hosmer-Lemeshow statistical test. Lower Chi-square values and high p-values indicate a good fit. Sensitivity, specificity and accuracy of the models was determined for the complete patient cohort. ROC-curves and the area under the ROC curve (AUROC) are available in Figure 2 and 3. A p-value <0.05 indicates a significant difference in AUROC from the reference model. 95% CI, 95% confidence interval.

Figure legends

Figure 1 – Schematic overview of the patient selection process. Of the 359 patients with T2DM, 109 were excluded due to missing C-peptide levels, missing data or because of an incomplete follow-up.

Figure 2 – Receiver operating characteristic (ROC) curves of the different predictive scoring models for complete T2DM remission. Depending on the scoring range, a single (point with highest specificity and sensitivity) or multiple cut-off (if limited scoring options) points are displayed on the ROC curve with the specificity and specificity between parentheses. The area under the ROC curve (AUROC) is displayed as the percentage in green. Values between parentheses indicate the 95% confidence interval of the AUROC. The prediction model by Umemura et al. (not displayed) had the lowest AUROC with 69.2%. ROC-curves for other definitions of T2DM remission and a comparison with the constructed model are available in Supplementary Material 6.

Figure 3 - Receiver operating characteristic (ROC) curves of the different predictive logistic regression models for complete T2DM remission. The custom model that was designed to predict T2DM remission in the study cohort is displayed in figure 3a. The cut-off point with the highest sensitivity and specificity is displayed on the ROC curve with the specificity and specificity between parentheses. The area under the ROC curve (AUROC) is displayed as a percentage in green. Values between parentheses indicate the 95% confidence interval of the AUROC. ROC-curves for other definitions of T2DM remission and a comparison with the constructed model are available in Supplementary Material 6.