Systematic review of risk prediction models for diabetes after bariatric surgery


Published in:
British Journal of Surgery

DOI:
10.1002/bjs.10255

2016

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Systematic review of risk prediction models for diabetes after bariatric surgery

R. Zhang¹, O. Borisenko¹, I. Telegina¹, J. Hargreaves³, A. R. Ahmed⁴, R. Sanchez Santos⁶, C. Pring⁵, P. Funch-Jensen⁷, B. Dillemans⁸ and J. L. Hedenbro²

¹Health Economy, Synergus AB, Stockholm, and ²Clinical Sciences Department of Surgery, Lund University, Lund, Sweden, ³Healthcare, Policy and Reimbursement, Covidien (UK) Commercial Ltd, now part of Medtronic, ??, ⁴Department of Surgery and Cancer, Imperial College London, London, and ⁵Department of Bariatric Surgery, St Richards Hospital, Chichester, UK, ⁶General Surgery, Hospital of Pontevedra, Pontevedra, Spain, ⁷Department of Clinical Medicine, Aarhus University, and Aleris Hamlet Hospital, Aarhus and Copenhagen, Denmark, and ⁸Department of General Surgery, St Jan’s Hospital, Bruges, Belgium

Correspondence to: Dr O. Borisenko, Health Economy, Synergus AB, Djursholmsvägen 20C, Stockholm, Sweden (email: oleg.borisenko@synergus.com)

Background: Diabetes remission is an important outcome after bariatric surgery. The purpose of this study was to identify risk prediction models of diabetes remission after bariatric surgery.

Methods: A systematic literature review was performed in MEDLINE, MEDLINE-In-Process, Embase and the Cochrane Central Register of Controlled Trials databases in April 2015. All English language full-text published derivation and validation studies for risk prediction models on diabetic outcomes after bariatric surgery were included. Data extraction included population, outcomes, variables, intervention, model discrimination and calibration.

Results: Of 2331 studies retrieved, eight met the inclusion criteria. Of these, six presented development of risk prediction models and two reported validation of existing models. All included models were developed to predict diabetes remission. Internal validation using tenfold validation was reported for one model. Two models (ABCD score and DiaRem score) had external validation using independent patient cohorts with diabetes remission at 12 and 14 months respectively. Of the 11 cohorts included in the eight studies, calibration was not reported in any cohort, and discrimination was reported in two.
**Conclusion:** A variety of models are available for predicting risk of diabetes following bariatric surgery, but only two have undergone external validation.
A: Introduction

The prevalence of obesity (BMI 30 kg/m² or more) has been increasing worldwide⁴⁵. Bariatric surgery is the most effective treatment for morbid obesity⁶, resulting in a significant decrease in weight, as well as amelioration of associated co-morbidities including type 2 diabetes (T2DM)⁴⁻⁶, cardiovascular diseases⁷, obstructive sleep apnoea⁸ and musculoskeletal disorders⁹. A decrease in the number of co-morbidities may lead to a reduction in the healthcare resources associated with managing severe and complex obesity¹⁰⁻¹².

Risk prediction models are based either on approximations of absolute probability, or the risk that a specific outcome can occur within a certain time period in a subject with an individual predictor profile (through the use of predictor variables (co-variables))¹³. Risk predictors include patient characteristics (such as age and sex), medical history, blood chemistry results and genetic markers. Predictors for diabetes resolution include the mode of diabetes control (diet, oral hypoglycaemic drugs, insulin), good glycaemic control, age at surgery, duration of diabetes and waist circumference¹⁴,¹⁵.

Development of a multivariable prediction model requires a number of steps: selecting a set of candidate predictors; identifying important predictors among them by regression analysis; specifying a model by assigning relative weights for each predictor in a combined risk calculator; estimating performance of the model by measuring model calibration and discrimination; and conducting internal validation to assess the potential for optimism and adjusting the model for overfitting when necessary¹³. Good models are usually derived from large observational studies.

Risk prediction models are used to guide clinicians and patients in a joint decision-making process for selection of appropriate treatments¹⁶. Accurate prognostic assessment may safeguard against putting patients in a high-risk situation and prevent an
unnecessary economic burden on a healthcare system. To achieve this, prognostic models must be accurate and generalizable. Internal validation is not sufficient to confirm that a model which successfully predicts the outcome of interest is valuable or applicable to new individuals\textsuperscript{17}. Thus, external validation of derivation cohorts in new individuals is very important.

Calibration and discrimination are major evaluation methods for prediction models\textsuperscript{13}. Calibration refers to the agreement between observed and predicted outcomes. It can be assessed graphically by plotting, or statistically by testing for goodness of fit\textsuperscript{13}. Discrimination refers to the ability to discriminate individuals with the outcome from those without it. Statistics commonly used to evaluate discrimination performance of prediction models include the concordance (or c) statistic, or area under the receiver operating characteristic (ROC) curve\textsuperscript{18}.

Within an obese population, in the subgroup of diabetic patients, more patients indicated a cure of diabetes (58 per cent) as the most important outcome rather than weight loss (33 per cent). Understanding the potential benefits of surgery in relation to remission of diabetes may impact on the decision-making processes of patients and physicians.

The objective of this systematic review was to identify studies that have developed or validated risk prediction models for remission of T2DM after bariatric surgery and describe their performance.

+A: Methods

+B: Literature search and citation screening

A systematic literature search was performed in MEDLINE, MEDLINE-In-Process, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL). A detailed description of the search strategy used in each database, and the selection process as adapted from the PRISMA framework\textsuperscript{19}, is presented in Appendix S1 (supporting information). Searches were
conducted on 28 April 2015, and were restricted to full-text articles. There was no restriction on the timespan of the search.

Abstract screening was carried out by two reviewers. The evaluation of full-text publications was performed by a single reviewer using the inclusion and exclusion criteria provided below. A second reviewer checked the appropriateness of inclusion of studies. Disagreements were resolved by consensus.

+B: Study selection

Studies were considered for inclusion based on the following criteria: intervention (bariatric surgery); type of study (observational studies, RCTs); predictive model (at least 2 risk factors or validation studies); outcomes reported (diabetes outcomes); language (English). Validation studies were included when the study validated the model in relation to the same outcome as reported in the derivation study.

+B: Data extraction and analysis

The following data from each included publication were extracted by one reviewer: population characteristics; intervention; selection of variables; number of subjects in the derivation or/and validation cohorts; source of the study population; utilization outcome; internal validation; model calibration; and discrimination.

+B: Assessment of model performance

Data related to discrimination (the ability of a model to recognize individuals who experience the outcome from those who do not) and calibration (agreement between the model estimated outcome and the observed outcome) were abstracted. Discrimination was identified from the c-statistic, or area under the ROC curve (AUC)\(^{13}\); an AUC of 0.500 suggested no discriminatory power, 0.501–0.699 poor discriminatory power, 0.700 to 0.799 acceptable discriminatory power, 0.800–0.899 excellent discriminatory power, and 0.900 indicated outstanding
discriminatory power\textsuperscript{20,21}. Model calibration was identified from Hosmer–Lemeshow tests or correlation coefficients for each study\textsuperscript{13}.

\textbf{+A: Results}

The search strategy yielded 2330 citations. Of these, 102 studies were eligible for full-text review and eight studies, evaluating six risk prediction models, were selected (\textit{Fig. 1}). Articles excluded with reasons for exclusion are shown in \textit{Table S1} (supporting information). All models focused on prediction of remission of T2DM. Among the six published risk prediction models, two were validated in one or more independent cohorts\textsuperscript{22,23}. Among the remaining four models, one was internally validated (by 10-fold validation method\textsuperscript{24}), whereas the remaining models were not validated\textsuperscript{25–27}. \textit{Table 1} summarizes the characteristics of the included studies. Models were developed and validated in cohorts varying widely in patient sample size, with a median of 103 (range 46–690) patients.

\textbf{+B: Risk prediction models validated in at least one independent cohort}

The Diabetes Surgery score (ABCD score)\textsuperscript{22} and diabetes remission (DiaRem) score\textsuperscript{23} were both validated in one or more independent cohorts.

\textbf{+C: ABCD score}

The ABCD score includes four categorical variables to predict remission of T2DM: BMI, C-peptide, T2DM duration and age). The ABCD score ranges from 0 to 10 points by summing the points for each variable, with high scores indicating a greater chance of remission\textsuperscript{22}. The ABCD score was derived from a multicentre cohort including 63 patients who had a BMI of at least 35 kg/m\textsuperscript{2}, or a BMI below 35 kg/m\textsuperscript{2} but with poorly controlled T2DM, and who had undergone laparoscopic gastric bypass for uncontrolled T2DM. Patient follow-up in the derivation cohort was at least 3 years. Internal validation was not performed in the derivation cohort.
The model has been validated in three independent cohorts, including a total of 341 patients who underwent bariatric surgery\textsuperscript{22,28,29}, using the outcome T2DM remission at 1 year after surgery. The validation cohorts were mainly from the same institutions as the derivation cohort, but at a later time. Three cohorts consisted of patients with a mean age ranging from 42 to 48 years, 48–64 per cent females, a BMI between 26.9 and 39.0 kg/m\textsuperscript{2}, and a T2DM duration ranging from 2.4 to 6.5 years before the surgery. Patients underwent either laparoscopic gastric bypass (Roux-en-Y gastric bypass, LYGB) or laparoscopic sleeve gastrectomy.

Model calibration or discrimination was not reported in either the derivation or validation cohorts.

\textbf{+B: DiaRem score}

The DiaRem score was developed to predict remission of T2DM after RYGB\textsuperscript{23}. It includes four variables. Three of these are categorical (age, glycosylated haemoglobin A1c (HbA1c) level and other diabetes drug groups) and a single binomial variable: treatment with insulin. The DiaRem score ranges from 0 to 22, with low scores predicting a high probability of remission and high scores the converse. The DiaRem score was derived from a retrospective cohort of 690 patients with T2DM with at least 14 months of follow-up. Internal validation was not performed in the derivation cohort.

The model has been validated in two independent single-centre cohorts of 359 patients undergoing RYGB at 14 months after the surgery\textsuperscript{23}. These cohorts involved patients with a mean BMI of 48.4 and 49.5 kg/m\textsuperscript{2}, a female prevalence of 68 and 73 per cent, and a mean insulin use of 28 and 36 per cent in each cohort. All patients underwent RYGB surgery.

Model calibration or discrimination was not reported in either the derivation or validation cohorts.
+B: Risk prediction models without validation in independent cohorts

There were four models without external cohort validation\textsuperscript{24–27}.

Dixon and colleagues\textsuperscript{25} developed a risk calculator that estimates the likelihood of an individual achieving remission of T2DM\textsuperscript{25}. The model includes two continuous variables: BMI and diabetes duration. The derivation cohort evaluated in a prospective longitudinal study included 103 patients who underwent laparoscopic gastric bypass with a follow-up of 12 months. Internal validation was not performed. Discrimination was assessed with a ROC curve. Optimal cut-off points of BMI exceeding 27 kg/m\textsuperscript{2} and a duration of diabetes shorter than 7 years provided sensitivities, specificities and AUC values of 68 per cent, 71 per cent and 0.69, and 69 per cent, 63 per cent 0.66, respectively.

Hayes and co-workers\textsuperscript{24} published a risk calculator for remission of T2DM with two variables. The model included one binominal variable (diabetes status) and one continuous variable (preoperative HbA1c). The derivation cohort was evaluated a prospective single-centre study of 127 patients with T2DM who underwent gastric bypass. Follow-up of the derivation cohort was 12 months. Internal validation was conducted using the tenfold cross-validation method. Calibration or discrimination statistics were not evaluated.

Robert \textit{et al.}\textsuperscript{27} published a risk prediction score for remission of T2DM with five variables. The model was based on a retrospective cohort of 46 patients with T2DM, who had a BMI of 35 kg/m\textsuperscript{2} or more with follow-up of 12 months. The risk prediction score ranges from 0 to 5 with five binominal variables (\textit{Table 1}). Discrimination was assessed by ROC analysis; the AUC was 0.950 (95 per cent 0.838 to 0.992; $P < 0.001$) indicating outstanding discriminatory power. A cut-off value of more than 2 provided 97 per cent sensitivity in predicting diabetes remission, with 86 per cent specificity.
Ugale and colleagues\textsuperscript{26} proposed a scoring system with seven variables for postoperative diabetes remission. The scoring system was derived from a retrospective cohort of 75 patients with poorly controlled T2DM who underwent the experimental method of ileal interposition in combination with two varieties of sleeve gastrectomy. The mean follow-up was 30.2 and 12.7 months for two groups with different types of sleeve gastrectomy. Internal validation, calibration and discrimination were not reported.

\textbf{+A: Discussion}

This systematic review identified and evaluated six risk prediction models for diabetes outcomes after bariatric surgery. Only two models (ABCD score\textsuperscript{22} and DiaRem score\textsuperscript{23}) have been validated in external cohorts and both have been validated in more than two independent cohorts. Data regarding the quality of the models (model calibration and discrimination) were not, however, reported for either instrument. Model discrimination demonstrates how well a model can discriminate future events of remission from non-events. In a hypothetical example, a \textit{c}-statistic of 0.70 indicates that in 70 per cent of cases a randomly selected patient with remission of diabetes will have a higher model score than a patient with no remission. However, \textit{c}-statistics do not indicate how similar predicted chances of remission are to observed values. This can be tested by means of calibration analysis, which can be assessed either visually (how close the predicted and observed values are) or with specific tests. Miscalibrated models may lead to the situation where a patient with a high chance of remission is actually assigned to a low chance of remission, thereby leading to a biased interpretation of the benefits of surgery. Reporting both calibration and discrimination is a standard step in evaluating the performance of risk prediction models\textsuperscript{30}.

In the model of Dixon and colleagues\textsuperscript{25}, the AUC was 0.69 (inadequate), meaning that only in 69 per cent of random cases would a score indicating higher chances of remission actually be higher in patients who experienced remission than in patients with no remission.
Outstanding discriminative ability was demonstrated (AUC 0.950) for the model of Robert et al. The five variables in this are BMI, duration of diabetes, HbA1c level, concentration of fasting glucose and oral antidiabetic drugs.

Models were developed for different patient groups. The diabetes remission score proposed by Ugale and colleagues, and the model by Dixon et al. were developed from analysis of patients with a BMI below 30 kg/m², and a long history of diabetes (8–10 years). All other models originally included patients with a much higher BMI (39–50 kg/m²) and a shorter duration of diabetes (3–5 years).

Five of the six models were developed to predict remission 12–14 months after surgery and one model did not specify any time horizon. Although this might still be of relevance for patients and physicians, predictions should be interpreted with caution in relation to mid- and long-term effects of surgery. For example, in the Swedish Obesity Subjects Study, where most patients had undergone vertical banded gastroplasty, the percentage of patients for whom recovery of diabetes was reported was reduced from 72 per cent at 2 years to 36 per cent at 10 years.

The surgery types were mainly gastric bypass and sleeve gastrectomy. The ABCD score was developed from a cohort of patients who underwent gastric bypass. It was validated in two cohorts: one with gastric bypass and the other including both bypass and sleeve gastrectomy. The DiaRem score was developed and validated in patients undergoing RYGB. A systematic review that compared the co-morbidity outcomes after laparoscopic RYGB and sleeve gastrectomy showed that RYGB and sleeve gastrectomy had similar effects on T2DM. Only the model of Robert and colleagues included a mixture of all current treatment options. The diabetes remission model proposed by Ugale et al. utilized modifications of sleeve gastrectomy.
The geographical origin of the derivation cohorts might also be important for understanding the value of the developed models. The ABCD score, the diabetes remission score and the model of Dixon et al. were developed in Asian populations. The DiaRem score was based on a cohort in the USA and the model reported by Hayes and colleagues was developed in New Zealand. The only model developed in a European cohort is that by Robert and co-workers.

Output format can also be important for ease of use of models in clinical practice. The ABCD score, DiaRem score, the diabetes remission score proposed by Ugale and colleagues and the algorithm proposed by Robert et al. are risk calculator scores. Therefore, the output of each model is a value at a point along a predefined scale. Scores might be less intuitive than risk or chances of remission, and may require extensive use to allow easy day-to-day interpretation of the results in relation to an individual patient’s prognosis. Hayes and colleagues proposed two formulas to determine whether a patient is likely to recover from diabetes if the value in one formula (‘class resolved’) is higher than that in the other (‘class not resolved’). Dixon and co-workers proposed a simple formula to calculate the likelihood of remission.

All published risk prediction models have limitations in quality and further validation is required. They still might be of relevance for clinical practice. The type of surgery, patient population, output format, and availability of inputs to physician and patients can all influence the choice of model. Limitations of each model need to be evaluated, acknowledged and considered before implementation into clinical practice. The optimal management of bariatric surgery requires accurate assessment of prognosis, and this is still challenging.
**+A: Acknowledgements**

The present study was sponsored principally by Covidien Inc., and co-sponsored by grants from Johnson & Johnson, the Aleris Foundation, Lund University and the Crafoord foundation.

R.Z., O.B. and I.T. are employees of Synergus AB, which was commissioned by Covidien Inc. to perform the study. J.H. is an employee of Covidien Inc. BD reports consulting fees from Covidien Inc. and Johnson & Johnson, all outside the submitted work. P.F.J. reports consulting fees from Covidien Inc., Johnson & Johnson and AstraZeneca, all outside the submitted work. J.L.H. reports lecturing fees from Johnson & Johnson and Covidien, and research funding from the Aleris foundation, Lund University and the Crafoord foundation, all outside the submitted work.

**Disclosure:** The authors declare no other conflict of interest.

**+A: References**


Supporting information
Additional supporting information may be found in the online version of this article:
Appendix S1 Search strategies (Word document)
Table S1 Publications excluded after review of full text (Word document)

Fig. 1 PRISMA flow diagram of article selection
Table 1 Characteristics and performance of included risk prediction models

<table>
<thead>
<tr>
<th>Reference</th>
<th>Derivation/validation</th>
<th>Data collection</th>
<th>Model description (utilization outcome)</th>
<th>Population*</th>
<th>Internal validation</th>
<th>Calibration</th>
<th>Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al.22</td>
<td>Derivation</td>
<td>n.r.</td>
<td>ABCD score Age BMI C-peptide T2DM duration</td>
<td>63 patients undergoing LGB Age 38.4 years, 76% female, BMI 39.1 kg/m², T2DM duration 2.6 years Follow-up ≥ 3 years</td>
<td>n.r.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Validation</td>
<td>Prospective</td>
<td>ABCD score</td>
<td>176 patients undergoing LGB BMI 36.7 kg/m² T2DM duration 2.4 years Follow-up 12 months</td>
<td>n.a.</td>
<td>n.r.</td>
<td>n.r.</td>
<td></td>
</tr>
<tr>
<td>Lee et al.28</td>
<td>Validation</td>
<td>n.r.</td>
<td>ABCD score Age BMI C-peptide T2DM duration</td>
<td>80 patients undergoing LGB or LSG Age 47.7 years, 63% female, BMI 26.9 kg/m², T2DM duration 6.5 years Follow-up 12 months</td>
<td>n.a.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Lee et al.29</td>
<td>Validation</td>
<td>Retrospective</td>
<td>ABCD score Age BMI C-peptide T2DM duration</td>
<td>176 patients undergoing LGB BMI 36.7 kg/m² T2DM duration 2.4 years Follow-up 12 months</td>
<td>n.a.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Still et al.23</td>
<td>Validation (in AZ cohort)</td>
<td>n.r.</td>
<td>DiaRem Score Age HbA1c Other diabetes drugs Treatment with insulin (Remission of T2DM at 14 months)</td>
<td>690 patients undergoing RYGB Age 51.2 years, 73% female, BMI 49.4 kg/m², insulin use 36% Follow-up ≥ 14 years</td>
<td>n.r.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Validation (in PA cohort)</td>
<td>n.r.</td>
<td>DiaRem Score Age HbA1c Other diabetes drugs Treatment with insulin (Remission of T2DM at 14 months)</td>
<td>276 patients undergoing RYGB 68% female, BMI 48.4 kg/m², insulin use 28% Follow-up 14 months</td>
<td>n.a.</td>
<td>n.r.</td>
<td>n.r.</td>
<td></td>
</tr>
<tr>
<td>Dixon et al.25</td>
<td>Derivation</td>
<td>Prospective</td>
<td>BMI Diabetes duration</td>
<td>103 patients undergoing LGB Age 47.5 years, 60% female, BMI ≤ 30 (mean 26) kg/m², T2DM duration 8.2 years Follow-up 12 months</td>
<td>n.r.</td>
<td>n.r.</td>
<td>AUC 0.69 (cut-off point BMI &gt;27 kg/m²); AUC 0.66 (cut-off point T2DM duration &lt; 7 years)</td>
</tr>
<tr>
<td>Hayes et al.24</td>
<td>Derivation</td>
<td>Prospective</td>
<td>Use of insulin versus other drugs Preoperative HbA1c</td>
<td>127 patients undergoing gastric bypass Age 48.5 years, 65% female, mean BMI 46.8 kg/m², diabetes duration 4.5 years Follow-up 12 months</td>
<td>Yes (tenfold)</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>Robert et al.27</td>
<td>Derivation</td>
<td>Retrospective</td>
<td>BMI ≤ 50 kg/m² Diabetes duration ≤ 4 years HbA1c ≤ 7.1% Fasting glucose ≤ 114 mg/dl Oral antidiabetic agent Treatment without insulin</td>
<td>46 patients undergoing LAGB, LRYGB or LSG Age 45.3 years, 68% female, BMI 49.5 kg/m², diabetes duration 3.0 years Follow-up 12 months</td>
<td>n.r.</td>
<td>n.r.</td>
<td>AUC 0.950 (95 per cent c.i. 0.838 to 0.992; P &lt; 0.001)</td>
</tr>
<tr>
<td>Ugale et al.26</td>
<td>Derivation</td>
<td>Retrospective</td>
<td>Diabetes remission score Age BMI T2DM duration Microvascular complications Macrovascular complications Preoperative insulin use Stimulated C-peptide</td>
<td>75 patients undergoing IISG or IIDSG Age 54.0 years, 35% female, BMI 24.3 kg/m², T2DM duration 10.0 years Mean follow-up 30.2 months (IISG), 12.7 months (IIDSG)</td>
<td>n.r.</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
</tbody>
</table>

*Age, BMI, type 2 diabetes mellitus (T2DM) duration reported as mean values unless otherwise specified. n.r., not reported; LGB, laparoscopic gastric bypass; n.a., not applicable; LSG, laparoscopic sleeve gastrectomy; HbA1c, glycated haemoglobin; (L)RYGB, (laparoscopic) Roux-en-Y gastric bypass; AUC, area under receiver operating characteristic (ROC) curve; LAGB, laparoscopic adjustable gastric banding; IISG, ileal interposition coupled sleeve gastrectomy; IIDSG, ileal interposition diverted sleeve gastrectomy.