Systematic review

Systematic review of risk prediction models for diabetes after bariatric surgery

R. Zhang[1], O. Borisenko[1], I. Telegina[1], J. Hargreaves[3], A. R. Ahmed[4], R. Sanchez Santos[6], C. Pring[5], P. Funch-Jensen[7], B. Dillemans[8] and J. L. Hedenbro[2]

[1]Health Economy, Synergus AB, Stockholm, and [2]Clinical Sciences Department of Surgery, Lund University, Lund, Sweden, [3]Healthcare, Policy and Reimbursement, Covidien (UK) Commercial Ltd, now part of Medtronic, Fareham, [4]Department of Surgery and Cancer, Imperial College London, London, and [5]Department of Bariatric Surgery, St Richard's Hospital, Chichester, UK, [6]General Surgery, Hospital of Pontevedra, Pontevedra, Spain, [7]Department of Clinical Medicine, Aarhus University, Aarhus, and Aleris Hamlet Hospital, Copenhagen, Denmark, and [8]Department 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 2330 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 assessed 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.**

Paper accepted 27 May 2016

Published online in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.10255

Introduction

The prevalence of obesity (BMI 30 kg/m^2 or more) has been increasing worldwide[1,2]. Bariatric surgery is the most effective treatment for morbid obesity[3], resulting in a significant decrease in weight, as well as amelioration of associated co-morbidities including type 2 diabetes mellitus (T2DM)[4–6], cardiovascular diseases[7], obstructive sleep apnoea[8] and musculoskeletal disorders[9]. A decrease in the number of co-morbidities may lead to a reduction in the healthcare resources associated with managing severe and complex obesity[10–12].

Risk prediction models are based either on approximations of absolute probability or on 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 circumference[14,15].

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 necessary[13]. 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 treatments[16]. Accurate prognostic assessment may safeguard against putting patients in a high-risk situation, which in turn could prevent an unnecessary economic burden on healthcare systems. To achieve this, prognostic models must be accurate and generalizable. Internal validation is not sufficient to confirm that a model that successfully predicts the outcome of interest is valuable or applicable to new individuals[17]. Thus, external validation of any such model in new individuals is very important.

Calibration and discrimination are major evaluation methods for prediction models[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[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[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)[19]. 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.

Methods

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[20], 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.

Study selection

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

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 and/or validation cohorts; source of the study population; internal validation; model calibration; and discrimination.

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 is typically 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 to 1 indicated outstanding discriminatory power[21,22]. Model calibration is commonly identified with Hosmer–Lemeshow tests or correlation coefficients[13].

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 (*Fig. 1*). Articles excluded with reasons for exclusion are shown in *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[23,24]. Among the remaining four models, one was internally validated (by 10-fold validation method[25]), whereas the remaining models were not validated[26–28]. *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.

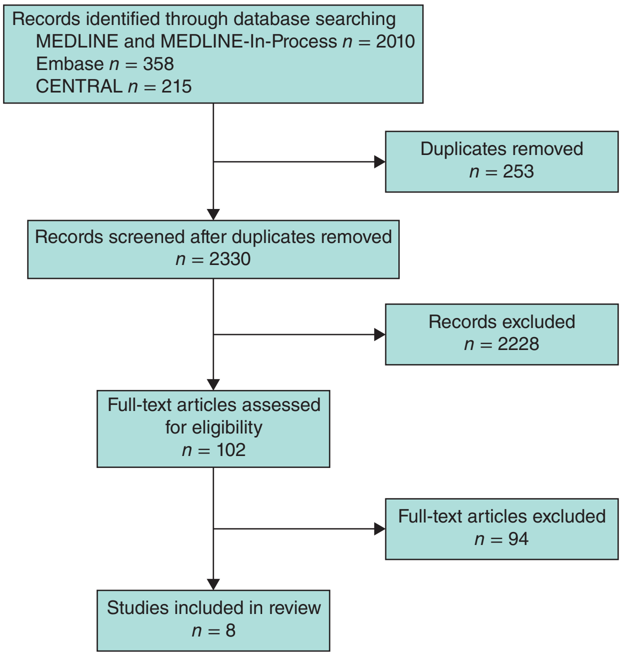

Fig. 1 PRISMA flow diagram of article selection

Risk prediction models validated in at least one independent cohort

The Diabetes Surgery score (ABCD score)[23] and diabetes remission (DiaRem) score[24] were both validated in one or more independent cohorts.

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[23]. The ABCD score was derived from a multicentre cohort including 63 patients who had a BMI of at least 35 kg/m^2, or a BMI below 35 kg/m^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[23,29,30], 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, 48–63 per cent women, a BMI between 26·9 and 39·0 kg/m^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.

DiaRem score

The DiaRem score was developed to predict remission of T2DM after RYGB[24]. 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 389 patients in total undergoing RYGB at 14 months after the surgery[24]. These cohorts involved patients with a mean BMI of 48·4 and 49·5 kg/m^2, a female prevalence of 68 and 74 per cent, and a mean insulin use of 28 and 38 per cent in each cohort. All patients underwent RYGB surgery.

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

Risk prediction models without validation in independent cohorts

There were four models without external cohort validation[25–28].

Dixon and colleagues[26] developed a risk calculator that estimates the likelihood of an individual achieving remission of T2DM[26]. 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^2 or 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 and 0·66, respectively.

Hayes and co-workers[25] published a risk calculator for remission of T2DM with two variables. The model includes one binominal variable (diabetes status) and one continuous variable (preoperative HbA1c). The derivation cohort was evaluated in a prospective single-centre study

Table 1 Characteristics and performance of included risk prediction models

Reference	Derivation/ validation	Data collection	Model description	Population*	Model evaluation: Internal validation	Model evaluation: Calibration	Model evaluation: Discrimination
Lee *et al.*[23]	Derivation	n.r.	ABCD score: age, BMI, C-peptide, and T2DM duration (Remission of T2DM at 12 months)	63 patients undergoing LGB Age 38·4 years, 76% female, BMI 39·1 kg/m^2, T2DM duration 2·6 years Follow-up ≥ 3 years	n.r.	n.r.	n.r.
	Validation	Prospective	ABCD score	176 patients undergoing LGB BMI 36·7 kg/m^2, T2DM duration 2·4 years Follow-up 12 months	n.a.	n.r.	n.r.
Lee *et al.*[29]	Validation	n.r.	ABCD score	80 patients undergoing LGB or LSG Age 47·7 years, 63% female, BMI 26·9 kg/m^2, T2DM duration 6·5 years Follow-up 12 months	n.a.	n.r.	n.r.
Lee *et al.*[30]	Validation	Retrospective	ABCD score	85 patients undergoing LSG Age 41·9 years, 48% female, BMI 39·0 kg/m^2, T2DM duration 2·7 years Follow-up 12 months	n.a.	n.r.	n.r.
Still *et al.*[24]	Derivation	Retrospective	DiaRem score: age, HbA1c, other diabetes drugs, and treatment with insulin (Remission of T2DM at 14 months)	690 patients undergoing RYGB Age 51·2 years, 73% female, BMI 49·4 kg/m^2, insulin use 36% Follow-up ≥ 14 months	n.r.	n.r.	n.r.
	Validation (in Scottsdale cohort)	n.r.	DiaRem score	276 patients undergoing RYGB 68% female, BMI 48·4 kg/m^2, insulin use 28% Follow-up 14 months	n.a.	n.r.	n.r.
	Validation (in Danville cohort)	n.r.	DiaRem score	113 patients undergoing RYGB 74% female, BMI 49·5 kg/m^2, insulin use 38% Follow-up 14 months	n.a.	n.r.	n.r.
Dixon *et al.*[26]	Derivation	Prospective	BMI, and diabetes duration (Remission of T2DM at 12 months)	103 patients undergoing LGB Age 47·5 years, 60% female, BMI ≤ 30 (mean 26) kg/m^2, T2DM duration 8·2 years Follow-up 12 months	n.r.	n.r.	AUC 0·69 (cut-off point BMI >27 kg/m^2); AUC 0·66 (cut-off point T2DM duration < 7 years)

Table 1 Continued

Reference	Derivation/ validation	Data collection	Model description	Population*	Model evaluation		
					Internal validation	Calibration	Discrimination
Hayes *et al.*[25]	Derivation	Prospective	Use of insulin *versus* other drugs and preoperative HbA1c (Remission of T2DM at 12 months)	127 patients undergoing gastric bypass Age 48·5 years, 65% female, mean BMI 46·8 kg/m^2, diabetes duration 4·5 years Follow-up 12 months	Yes (tenfold)	n.r.	n.r.
Robert *et al.*[28]	Derivation	Retrospective	BMI ≤ 50 kg/m^2, Diabetes duration ≤ 4 years, HbA1c ≤ 7·1%, Fasting glucose ≤ 114 mg/dl, Oral antidiabetic agent treatment without insulin (Remission of T2DM at 12 months)	46 patients undergoing LAGB, LRYGB or LSG Age 45·3 years, 68% female, BMI 49·5 kg/m^2, diabetes duration 3·0 years Follow-up 12 months	n.r.	n.r.	AUC 0·950 (95 per cent c.i. 0·838 to 0·992; $P < 0·001$)
Ugale *et al.*[27]	Derivation	Retrospective	Diabetes remission score Age, BMI, T2DM duration, Microvascular complications, Macrovascular complications, Preoperative insulin use, and Stimulated C-peptide	75 patients undergoing IISG or IIDSG Age 54·0 years, 35% female, BMI 24·3 kg/m^2, T2DM duration 10·0 years Mean follow-up 30·2 months (IISG), 12·7 months (IIDSG)	n.r.	n.r.	n.r.

*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 gastroctomy; 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.

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 *et al.*[28] 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^2 or more with follow-up of 12 months. The risk prediction score ranges from 0 to 5 with five binominal variables (*Table 1*). Discrimination was assessed by ROC analysis; the AUC was 0·950 (95 per cent c.i. 0·838 to 0·992; $P < 0·001$), indicating outstanding discriminatory power. A cut-off value in the risk prediction score of 2 or more provided 97 per cent sensitivity in predicting diabetes remission, with 86 per cent specificity.

Ugale and colleagues[27] 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.

Discussion

This systematic review identified and evaluated six risk prediction models for diabetes outcomes after bariatric surgery. Only two models (ABCD score[23] and DiaRem score[24]) have been validated in external cohorts and both have been validated in at least 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 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, 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[31].

In the model of Dixon and colleagues[26], the AUC was 0·69. This means that 69 per cent of randomly selected patients who experience the outcome (remission) score higher in the model relative to a randomly selected patient who did not experience remission.

Outstanding discriminative ability was demonstrated (AUC 0·950) for the model of Robert *et al.*[28]. The five variables in this model were 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[27], and the model by Dixon *et al.*[26] were developed from analysis of patients with a BMI below 30 kg/m^2, and a long history of diabetes (8–10 years). All other models[23–25,28–30] originally included patients with a much higher BMI (39–50 kg/m^2) and a shorter duration of diabetes (2·4–6·5 years).

Five of the six models were developed to predict remission at 12–14 months after surgery[23–26,28–30] and one model[27] 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 Obese Subjects Study[32], 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[23]. It was validated in two cohorts: one with gastric bypass and the other including both bypass and sleeve gastrectomy[29,30]. The DiaRem score was developed and validated in patients undergoing RYGB[24]. A systematic review[33] 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[28] included a mixture of all current treatment options. The diabetes remission model proposed by Ugale *et al.*[27] 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[23], the diabetes remission score[27] and the model of Dixon *et al.*[26] were developed in Asian populations. The DiaRem score was based on a cohort in the USA[24], and the model reported by Hayes and colleagues[25] was developed in New Zealand. The only model developed in a European cohort is that by Robert and co-workers[28].

Output format can also be important for ease of use of models in clinical practice. The ABCD score[23], the DiaRem score[24], the diabetes remission score proposed by Ugale and colleagues[27] and the algorithm proposed by Robert *et al.*[28] are risk scores. Therefore, the various model estimates are translated to a particular number of points along a scale of increasing risk, which in turn correspond to a certain probability of remission. Scores might be less intuitive than risks 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[25] 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[26] 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.

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. B.D. 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.

References

1 Buchwald H, Avidor Y, Braunwald E, Jensen MD, Pories W, Fahrbach K *et al*. Bariatric surgery: a systematic review and meta-analysis. *JAMA* 2004; **292**: 1724–1737.

2 Berghöfer A, Pischon T, Reinhold T, Apovian CM, Sharma AM, Willich SN. Obesity prevalence from a European perspective: a systematic review. *BMC Public Health* 2008; **8**: 200.

3 Baptista V, Wassef W. Bariatric procedures: an update on techniques, outcomes and complications. *Curr Opin Gastroenterol* 2013; **29**: 684–693.

4 Yu J, Zhou X, Li L, Li S, Tan J, Li Y *et al.* The long-term effects of bariatric surgery for type 2 diabetes: systematic review and meta-analysis of randomized and non-randomized evidence. *Obes Surg* 2015; **25**: 143–158.

5 Yip S, Plank LD, Murphy R. Gastric bypass and sleeve gastrectomy for type 2 diabetes: a systematic review and meta-analysis of outcomes. *Obes Surg* 2013; **23**: 1994–2003.

6 Ribaric G, Buchwald JN, McGlennon TW. Diabetes and weight in comparative studies of bariatric surgery *vs* conventional medical therapy: a systematic review and meta-analysis. *Obes Surg* 2014; **24**: 437–455.

7 Kwok CS, Pradhan A, Khan MA, Anderson SG, Keavney BD, Myint PK. Bariatric surgery and its impact on cardiovascular disease and mortality: a systematic review and meta-analysis. *Int J Cardiol* 2014; **173**: 20–28.

8 Ashrafian H, Toma T, Rowland SP, Harling L, Tan A, Efthimiou E *et al*. Bariatric surgery or non-surgical weight loss for obstructive sleep apnoea? A systematic review and comparison of meta-analyses. *Obes Surg* 2015; **25**: 1239–1250.

9 Groen VA, van de Graaf VA, Scholtes VA, Sprague S, van Wagensveld BA, Poolman RW. Effects of bariatric surgery for knee complaints in (morbidly) obese adult patients: a systematic review. *Obes Rev* 2015; **16**: 161–170.

10 Lopes EC, Heineck I, Athaydes G, Meinhardt NG, Souto KE, Stein AT. Is bariatric surgery effective in reducing comorbidities and drug costs? A systematic review and meta-analysis. *Obes Surg* 2015; **25**: 1741–1749.

11 Morgan DJ, Ho KM, Armstrong J, Litton E. Long-term clinical outcomes and health care utilization after bariatric surgery: a population-based study. *Ann Surg* 2015; **262**: 86–92.

12 Gesquiere I, Aron-Wisnewsky J, Foulon V, Haggege S, Van der Schueren B, Augustijns P *et al*. Medication cost is significantly reduced after Roux-en-Y gastric bypass in obese patients. *Obes Surg* 2014; **24**: 1896–1903.

13 Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG *et al*. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012; **98**: 683–690.

14 Adams ST, Salhab M, Hussain ZI, Miller GV, Leveson SH. Preoperatively determinable factors predictive of diabetes mellitus remission following Roux-en-Y gastric bypass: a review of the literature. *Acta Diabetol* 2013; **50**: 475–478.

15 Panunzi S, De Gaetano A, Carnicelli A, Mingrone G. Predictors of remission of diabetes mellitus in severely obese individuals undergoing bariatric surgery: do BMI or procedure choice matter? A meta-analysis. *Ann Surg* 2015; **261**: 459–467.

16 Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009; **338**: b375.

17 Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98**: 691–698.

18 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128–138.

19 Weinstein AL, Marascalchi BJ, Spiegel MA, Saunders JK, Fagerlin A, Parikh M. Patient preferences and bariatric surgery procedure selection; the need for shared decision-making. *Obes Surg* 2014; **24**: 1933–1939.

20 Moher D, Liberati A, Tetzlaff J, Altman D; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med* 2009; **6**: e1000097.

21 Hanley J, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.

22 Hosmer D, Lemeshow S. *Applied Logistic Regression*. Wiley-Interscience: New York, 2000.

23 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**: 379–384.

24 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**: 38–45.
25 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**: 910–916.
26 Dixon JB, Hur KY, Lee WJ, Kim MJ, Chong K, Chen SC *et al.* Gastric bypass in type 2 diabetes with BMI < 30: weight and weight loss have a major influence on outcomes. *Diabet Med* 2013; **30**: e127–e134.
27 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**: 89.
28 Robert M, Ferrand-Gaillard C, Disse E, Espalieu P, Simon C, Laville M. Predictive factors of type 2 diabetes remission 1 year after bariatric surgery: impact of surgical techniques. *Obes Surg* 2013; **23**: 770–775.
29 Lee WJ, Almulaifi A, Chong K, Chen SC, Tsou JJ, Ser KH *et al*. The effect and predictive score of gastric bypass and sleeve gastrectomy on type 2 diabetes mellitus patients with BMI < 30 kg/m^2. *Obes Surg* 2015; **25**: 1772–1778.
30 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 Related Dis* 2015; **11**: 991–996.
31 Vickers AJ, Cronin AM. Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework. *Semin Oncol* 2010; **37**: 31–38.
32 Sjöström L, Lindroos AK, Peltonen M, Torgerson J, Bouchard C, Carlsson B *et al*.; Swedish Obese Subjects Study Scientific Group. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N Engl J Med* 2004; **351**: 2683–2693.
33 Li J, Lai D, Wu D. Laparoscopic Roux-en-Y gastric bypass *versus* laparoscopic sleeve gastrectomy to treat morbid obesity-related comorbidities: a systematic review and meta-analysis. *Obes Surg* 2016; **26**: 429–442.

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)