Prognostic Models for Predicting Remission of Diabetes Following Bariatric Surgery: A Systematic Review and Meta-analysis

Pushpa Singh,[1,2] Nicola J. Adderley,[3] Jonathan Hazlehurst,[1,2] Malcolm Price,[3] Abd A. Tahrani,[1,2,4] Krishnarajah Nirantharakumar,[2,3,4,5] and Srikanth Bellary[2,6]

BACKGROUND

Remission of type 2 diabetes following bariatric surgery is well established, but identifying patients who will go into remission is challenging.

PURPOSE

To perform a systematic review of currently available diabetes remission prediction models, compare their performance, and evaluate their applicability in clinical settings.

DATA SOURCES

A comprehensive systematic literature search of MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, Embase, and Cochrane Central Register of Controlled Trials (CENTRAL) was undertaken. The search was restricted to studies published in the last 15 years and in the English language.

STUDY SELECTION

All studies developing or validating a prediction model for diabetes remission in adults after bariatric surgery were included.

DATA EXTRACTION

The search identified 4,165 references, of which 38 were included for data extraction. We identified 16 model development and 22 validation studies.

DATA SYNTHESIS

Of the 16 model development studies, 11 developed scoring systems and 5 proposed logistic regression models. In model development studies, 10 models showed excellent discrimination with area under the receiver operating characteristic curve ≥0.800. Two of these prediction models, ABCD and DiaRem, were widely externally validated in different populations, in a variety of bariatric procedures, and for both short- and long-term diabetes remission. Newer prediction models showed excellent discrimination in test studies, but external validation was limited.

LIMITATIONS

While the key messages were consistent, a large proportion of the studies were conducted in small cohorts of patients with short duration of follow-up.

CONCLUSIONS

Among the prediction models identified, the ABCD and DiaRem models were the most widely validated and showed acceptable to excellent discrimination. More studies validating newer models and focusing on long-term diabetes remission are needed.

[1]Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, U.K.
[2]Department of Diabetes and Endocrinology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, U.K.
[3]Institute of Applied Health Research, University of Birmingham, Birmingham, U.K.
[4]Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, U.K.
[5]Midlands Health Data Research, Birmingham, U.K.
[6]School of Life and Health Sciences, Aston University, Birmingham, U.K.

Corresponding author: Krishnarajah Nirantharakumar, k.nirantharan@bham.ac.uk

Bariatric surgery is an established cost-effective treatment option in patients with type 2 diabetes. In addition to sustained weight loss, it is associated with significant improvements in glycemic control, including achieving type 2 diabetes remission (1–3), and reduction in the risk of micro- and macrovascular complications and mortality (4–6). The proportion of patients achieving diabetes remission following bariatric surgery varies between studies and is estimated to be between 30 and 70%. This proportion lessens with longer follow-up and with longer diabetes duration at the time of surgery (7–9). This observed variation in the remission prevalence may be attributed to differences in definitions of diabetes remission, the population studied, the type of bariatric surgery, and the duration of follow-up.

Type 2 diabetes is one of the main indications for bariatric surgery in people with obesity (10). Given the variation in the rates of diabetes remission following bariatric surgery, a number of studies aiming to identify predictors of diabetes remission following bariatric surgery have been published (11–14). Variables associated with better β-cell function such as younger age, shorter diabetes duration, high C-peptide, lack of insulin treatment, and lower preoperative HbA$_{1c}$ (15) and lower preoperative BMI have been identified as predictors of type 2 diabetes remission postsurgery.

In consideration of the importance of predicting diabetes remission for individualizing care and helping patients and health care professionals to make informed decisions, several scoring systems incorporating the above-mentioned variables to predict diabetes remission have been developed (16–18). With acknowledgment of the mounting literature in this area and the increasing use of bariatric surgery worldwide, there is a need to describe the available prediction models and assess their ability to predict diabetes remission in patients with type 2 diabetes undergoing bariatric surgery and their utility in clinical practice.

METHODS

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (19). The protocol was registered on International prospective register of systematic reviews (PROSPERO), reg. no. CRD42019124644.

Data Sources and Searches

We searched MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations, Embase (Ovid), and Cochrane Central Register of Controlled Trials (CENTRAL). An example of the search strategy used in Embase can be found in Supplementary Table 1. Key search terms were type 2 diabetes, remission, and bariatric surgery. The search terms for prognostic/predictive models included prediction, prognosis, sensitivity, specificity, ROC (receiver operating characteristics) curve, and AUC (area under the ROC curve) with wild cards as well as other search terms as per published guidance (20).

The search was limited to articles published in the English language and in the last 15 years, as the concept of diabetes remission was established and coined by the American Diabetes Association in 2009. The final search was performed on 26 January 2019 and updated on 8 August 2020 (Fig. 1). EPPI-Reviewer 4 software was used for compiling the references, first screening by title and abstract, second screening of full text, and collaborating among reviewers (21).

Study Selection

First screening by title and abstract was performed by two reviewers (P.S. and S.B.) independently. Discrepancies were discussed to reach a consensus. We included clinical studies (observational or interventional studies) (Setting, S) involving adults with type 2 diabetes (Participants, P) who subsequently had bariatric/metabolic surgery (Interventions, I) and those who developed or validated a prediction model to predict diabetes remission (Outcome, O). Multiple definitions of diabetes remission were used, but we included only studies with definitions of HbA$_{1c}$ of ≤6.5% (48 mmol/mol) and participants being off glucose-lowering medication, with follow-up of at least a year.

We excluded review articles, studies with participants including children/adolescents or those with gastric cancer or gastric ulcer, studies where the intervention was other than bariatric/metabolic surgery, studies with an outcome of diabetes remission defined as HbA$_{1c}$ >6.5% (48 mmol/mol), studies with a follow-up period <12 months, and studies where the analysis was limited to identifying predictors.

Data Extraction and Quality Assessment

The data extraction template was drafted on DistillerSR software (22). Data were collected by P.S. and independently collected by second reviewers (N.J.A., J.H., and S.B.). We adapted the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) toolkit (23) to design the data collection domains; we gathered data on the country, data source, type of study, demographics of participants, type of bariatric procedure, length of follow-up, definition of diabetes remission, statistical method used for model development, and performance measures (Supplementary Table 3). We calculated the discrimination scores for models when this was not reported by the authors and where data to calculate them were available in the publication.

Risk of Bias Assessment

We used a customized version of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) to assess risk of bias and applicability (24). The assessment was done under four domains for risk of bias—participants, predictors, outcomes, and analysis—and three domains for applicability and generalizability: participants, predictors, and outcomes. The participants domain covers bias in patient selection and study design, the predictors domain is related to definition of predictors included in the prediction model, the outcome domain covers definition and measurement of the outcome, and the analysis domain relates to statistical analysis, handling missing data and overfitting (24) (Supplementary Table 4).

Statistical Analysis

A prediction model has three main phases: model development (preferably with internal validation), external validation, and investigation of clinical impact (25). Model development and validation involve identifying predictors, selecting the important predictors by regression analysis/modeling, proposing a model by assigning relative weights to the individual predictors included, conducting internal validation, and validating in
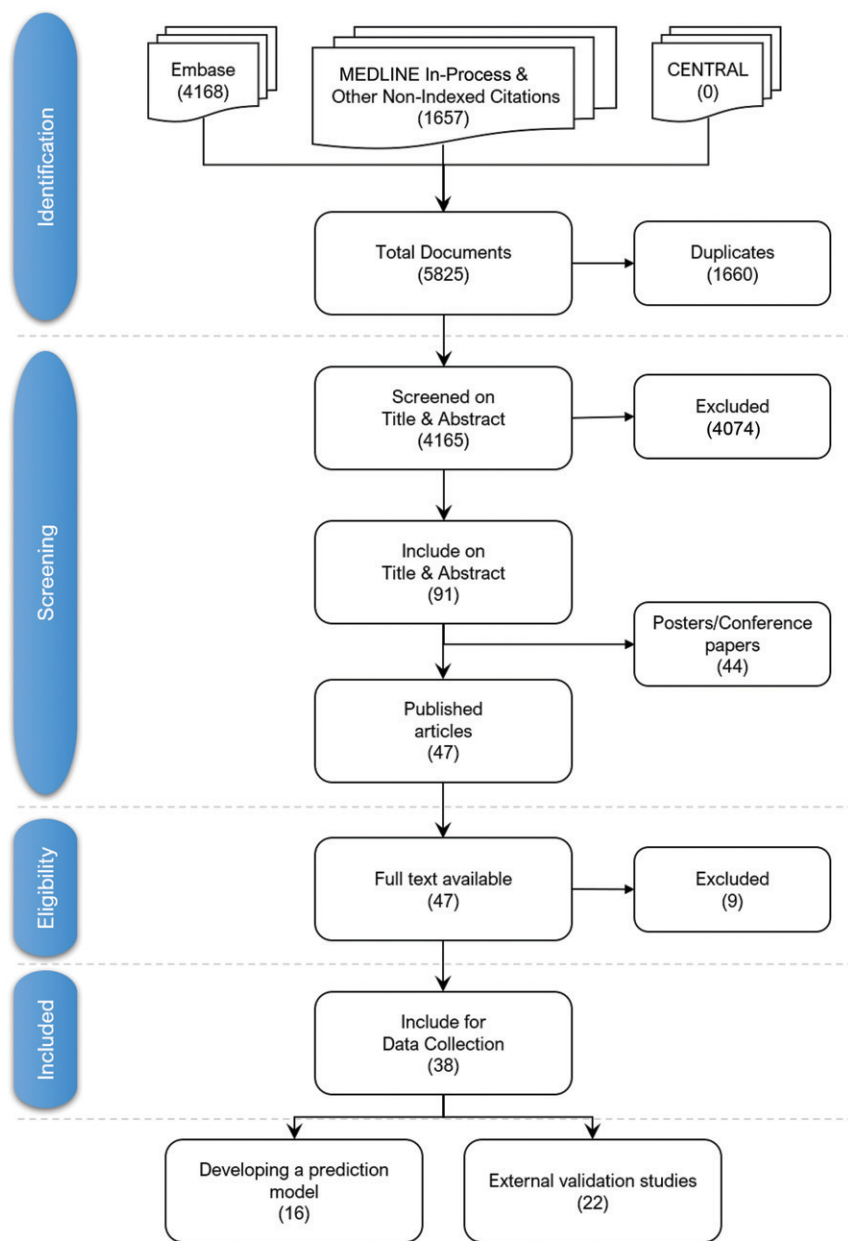
Figure 1—PRISMA flowchart.

poor discrimination, 0.700–0.799 as acceptable discrimination, 0.800–0.899 as excellent discrimination, and 0.900–1.000 as outstanding. For studies where AUC was not reported, where possible we calculated discrimination using published tables providing information on the participants, their score, and their outcome in terms of remission and nonremission. We used Stata for the analysis, generating AUC graphs and values (with 95% CI). Calibration was estimated by calculating the expected number (E) who should experience diabetes remission as reported in the model/score development paper and obtaining the observed number (O) from the tables providing information on score and outcome. This information was then used to calculate E-to-O ratio (30). E-to-O ratio of 1 represents perfect calibration, <1 represents underestimation of the events, and >1 represents overestimation (30).

As our search yielded studies with significant heterogeneity, we undertook three separate random-effects meta-analyses of studies based on *1*) their duration of follow-up, *2*) the HbA$_{1c}$ cutoffs used to define remission, and *3*) the type of bariatric surgery (Fig. 2*A–F*). We excluded studies from analysis where AUC was not known or could not be estimated with 95% CI.

Our first meta-analysis was based on follow-up duration; studies were grouped into those with follow-up of 1 year and those with follow-up of >1 year. In studies where diabetes remission was defined using two HbA$_{1c}$ cutoff values (e.g., 6.0% [42 mmol/mol] and 6.5% [48 mmol/mol]), we included the AUC for the higher cutoff only to avoid duplication of data sources.

The second meta-analysis was based on HbA$_{1c}$ cutoffs, and studies were grouped into those with HbA$_{1c}$ cutoffs of 6.5% (48 mmol/mol) and 6% (42 mmol/mol). In studies where diabetes remission was assessed at two follow-up points, we included the data with longer follow-up duration.

The third meta-analysis was based on type of surgery; studies were grouped into Roux-en-Y gastric bypass (RYGB) or sleeve gastrectomy (SG) groups. We excluded studies in which the discrimination score was not available for specific interventions separately. Similar to the previous two meta-analyses, we included the AUC for the higher HbA$_{1c}$ cutoff and

an external cohort to avoid overfitting (26). In this review, we classed the studies with development and internal validation of a prediction model as model development studies and studies with external validation of prediction models in a new cohort as validation studies.

We explored these phases for the identified prediction models and assessed the performance (26,27). We assessed the performance of the models based on discrimination, defined as the ability to distinguish between those who will and who will not achieve the outcome of interest, and calibration, defined as the ratio of those

expected to have a desired outcome to those observed to achieve the outcome.

Data Synthesis and Analysis

For prediction models presented as a scoring system, the sensitivity and specificity will vary depending on different cut points. Therefore, we chose to assess discrimination using area under the ROC curve (AUC), which covers all the sensitivity and specificity values at different cut points (28) and allows comparison of the prediction models. AUC of 0.5 signifies no ability to discriminate. For our study, we followed the categorizations used by Zhang et al. (29), defining 0.501–0.699 as

Table 1—Model development studies with their predictors

Prediction model	Predictors included	Discrimination (AUC) in model development studies	Discrimination (AUC) in external validation studies
ABCD; Lee et al. (16), 2013	Age, BMI, C-peptide, and diabetes duration	0.792 (0.728–0.856)*	Fig. 2A, C, and E
DiaRem; Still et al. (17), 2014	Age, HbA1c, diabetes medication other than metformin, and insulin use	0.840 (0.795–0.886)*	Fig. 2B, D, and F
Robert et al. (31), 2013	BMI, diabetes duration, HbA1c, fasting glucose, and diabetes medication	0.950 (0.838–0.992)	Shen et al.: 0.681 ± 0.056
DRS; Ugale et al. (33), 2014	Age, baseline BMI, diabetes duration, microvascular complications, macrovascular complication, insulin use, and stimulated C-peptides	NA	Ahuja et al.: 0.732 (0.633–0.83)*
Ad-DiaRem; Aron-Wisnewsky et al. (34), 2017	Age, HbA1c, insulin use, diabetes medication other than metformin, number of glucose-lowering agents, and diabetes duration	0.911	Shen et al. 0.849 ± 0.039, Dicker et al. 0.85 (0.76–0.93), Kam et al. at 1 year 0.752 (0.688–0.808), Kam et al. at 3 years 0.794 (0.715–0.860), 5y-DR 84%
DiaBetter; Pucci et al. (35), 2017	HbA1c, diabetes duration, and kind of diabetes medication	0.867 (0.817–0.916)	Shen et al. 0.826 ± 0.041, Kam et al. at 1 year 0.760 (0.697–0.815), Kam et al. at 3 years 0.804 (0.726–0.868),
IMS; Aminian et al. (18), 2017	Number of diabetes medication, insulin use, diabetes duration, and HbA1c	NA	Shen et al. 0.849 ± 0.040, Park et al. 0.76 (0.685–0.836),* Chen et al. 0.766 (0.716–0.817)* in GB, Chen et al. 0.599 (0.501–0.697)* in SG, Umemura et al. 0.516 (0.330–0.702)*
DiaRem2; Still et al. (36), 2018	Age, HbA1c, diabetes medication other than metformin and insulin use, diabetes duration	0.876	NA
5y-DR (37), 2018	Preoperative factors: diabetes duration, no. of medications, HbA1c. Postoperative factors: no. of medications, fasting CBG, weight loss, 1-year remission	90%	NA
MDR (38), 2020	Age, HOMA2-B, diabetes duration, and HbA1c	0.79 (0.71–0.88)	NA
Umemura et al. (39), 2020	Insulin, diabetes duration	0.865 (0.775–0.954)*	NA
Hayes et al. (40), 2011	Insulin use and HbA1c	NA	0.632 ± 0.059
Dixon et al. (13), 2013	BMI, diabetes duration, C-peptide	0.90 (0.84–0.95)	0.800 ± 0.047
Ramos-Levi et al. (41), 2014	Model 1, age, sex, FG, diabetes duration, insulin; Model 2: age, sex, FG, diabetes duration, insulin, C-peptide; Model 3: age, sex, FG, diabetes duration, insulin, % wt loss; Model 4: age, sex, FG, diabetes duration, insulin, % wt loss, C-peptide	0.838 (0.725–0.951); 0.923 (0.852–0.996); 0.923 (0.851–0.996); 0.981 (0.951–1.000)	0.811 ± 0.047
Cotillard et al. (42), 2015	Age, sex, BMI, fasting glycemia, HbA1c hypertension, diabetes duration, insulin therapy, number of antidiabetes drugs, C-peptide	NA	NA
Stallard et al. (43), 2016	Diabetes duration, FPG, use of noninsulin antidiabetes medications, and use of insulin	0.860 (0.763–0.957)	NA

CBG, capillary blood glucose; FG, fasting glucose; FPG, fasting plasma glucose; GB, gastric bypass; HOMA2-B, HOMA2 of β-cell function; NA, not available; SG, sleeve gastrectomy; wt, weight. *Calculated by authors of this systematic review.

Table 2—Study characteristics of model development studies

Publication reference	Source of data	Groups and numbers	Age (years)	BMI (kg/m²)	Diabetes duration (years)	HbA1c (%)	Outcomes	Types of surgery	Presentation	V Dev	Ext V
			Participant characteristics							Validation	
ABCD; Lee et al. (16), 2013	Retrospective, Taiwan, multicenter, 2005–2010	N = 63; 17 M, 56 F					n = 48 (76%), FU = 1 year	RYGB	Scoring system	Y	Y
		R	36.5 ± 10.7	40.9 ± 8.9	2.1 ± 3.7	8.2 ± 1.8					
		NR	44.5 ± 7.7	33.3 ± 7.4	4.1 ± 4.5	8.5 ± 1.8					
DiaRem; Still et al. (17), 2014	Retrospective, U.S., multicenter, 1 January 2004–February 2011	N = 690; 184 M, 506 F					n = 463 (67%), FU = 14 months	RYGB	Scoring system	Y	Y
		NI (n = 438)	48.8 ± 10.3	49.5 ± 8.0	6.8 ± 1.2	NA					
		I (n = 252)	53.6 ± 8.9	49.2 ± 8.8	8.2 ± 1.7	NA					
Robert et al. (31), 2013	Retrospective, observation, France, 2007–2010	N = 46; M:F = 1:3	45.3 ± 1.6	49.5 ± 1.22	3 (IQR 2.0–6.42)	7.44 ± 0.24	DR = 62.8% at 1 year of FU	RYGB (26), GB (11), SG (9)	Scoring system	N	Y
DRS; Ugale et al. (33), 2014	Retrospective, India, single, 1 February 2008–March 2010	N = 75; 49 M, 26 F					n = 42 (56%), FU = 1–2.5 years	SG	Scoring system	N	Y
		IISG	51.7 ± 13.3	23.4 ± 4.5	9.9 ± 4.8	8.1 ± 0.59					
		IIDSG	57.6 ± 11.5	25.6 ± 4.5	10.1 ± 5	9 ± 0.78					
Ad-DiaRem; Aron-Wisnewsky et al. (34), 2017	Retrospective, France, 1999–2014	N = 213; M 30%					n = 97 (45.5%), FU = 1 year	RYGB	Scoring system	Y	Y
		R	46 ± 10	48.1 ± 7.4	3.5 ± 3.8	7.0 ± 1.1					
		NR	53 ± 9	45.4 ± 7	11.1 ± 7.6	8.4 ± 1.6					
DiaBetter; Pucci et al. (35), 2017	Retrospective, U.K., single, 1 January 2008–December 2015	N = 210					n = 144 (68.6%), FU = 2 years	RYGB, SG	Scoring system	Y	Y
		RYGB (107)	51.6 ± 8	43.1 ± 6.3	5.6 ± 5.1	4.7 ± 5.4					
		SG (103)	49.7 ± 8.8	48.2 ± 7.8	7.8 ± 1.5	7.3 ± 1.4					
IMS; Aminian et al. (18), 2017	Retrospective, U.S., single, 2004–2011	N = 659; F = 451 (68%)	51 ± 10	46.4 ± 9.0	6 (3–11)	7.4 (6.4–8.6)	n = 291 (44.2%), FU = 5 years	RYGB, SG	Scoring system	N	Y
DiaRem2; Still et al. (36), 2018	Retrospective, U.S., single, 2009–2015	N = 307; F = 69%	51.2 ± 10.1	49.2 ± 10.3	6	NA	n = 135 (44.0%), FU = 1 year	RYGB	Scoring system	N	N

Table 2—Continued

Publication reference	Source of data	Groups and numbers	Age (years)	BMI (kg/m²)	Diabetes duration (years)	HbA$_{1c}$ (%)	Outcomes	Types of surgery	Presentation	V Dev	Ext V
			Participant characteristics							Validation	
5y-DR (37), 2018	Retrospective, France	N = 175; F = 136 (77.71%)	48.3 ± 10.3	47.37 ± 7.43	6.75 ± 6.53	7.5 ± 1.6	66 (37.7) at 1 year, 94 (53.7) at 5 years, FU = 5.1 ± 0.7 years	RYGB	Scoring system	Y	N
MDR; Moh et al. (38), 2020	Retrospective, Singapore, 2007–2018	N = 114	46 ± 9	40.1 ± 6.6	6 (2–10)	8.8 ± 1.9	54 (47.4%), FU = 1 year	RYGB, SG	Scoring system	N	N
Umemura et al. (39), 2020	Retrospective, Japan, single, 2008–2018	N = 49; F = 22 (44.9%)	46.2 ± 12.6	42.5 ± 6.4	5.6 ± 5.7	8.0 ± 1.9	n = 38 (77.6%), FU = 1 year	SG	Scoring system	N	N
Hayes et al. (40), 2011	New Zealand, single, 1 November 1997–May 2007	N = 127; 45 M, 82 F	48.5 ± 10.1	46.8 ± 9.4	4.5 ± 5	7.7 ± 1.7	n = 107 (84.3%), FU = 1 year	RYGB	Logistic regression	Y	Y
Dixon el al. (13), 2013	Retrospective, Taiwan, single	N = 154; 49 M	39.5 ± 10.7	37.2 ± 8.8	2 (0.5–5.0)	9.1 ± 1.7	n = 107 (69.5%), FU = 1 year	RYGB	Logistic regression	N	Y
Ramos-Levi et al. (41), 2014	Retrospective, Spain, single, 2006–2011	N = 141; 30 M, 81 F	53	43.7 ± 5.6	5 (2.0–10.0)	7.3 (6.5–8.4)	n = 74 (52.5%), FU = 1 year	RYGB, SG, DS	Logistic regression	N	Y
Cotillard et al. (42), 2015	France, single	N = 84; 15 M, 45 F; DR (n = 50); DNR (n = 34)	46.96 ± 9.14; 54.47 ± 11.02	46.93 ± 5.82; 46.1 ± 6.62	3.86 ± 4.64; 14.21 ± 7.63	7.01 ± 1.03; 8.21 ± 1.32	n = 50 (59.5%), FU = 1 year	RYGB	Logistic regression	N	N
Stallard et al. (43), 2016	Retrospective, Canada, single, 1 January 2011–June 2014	N = 98; 22 M, 76 F	49.7 ± 8.5	49.7 (48.1–51.1)	6.7 ± 6.6	7.6 (7.3–7.9)	n = 52 of 77 (67.5%), FU = 1 year	RYGB, SG	Logistic regression	N	N

Data are mean ± SD or median (interquartile range) unless otherwise indicated. N = total number of participants. n = number of participants achieving diabetes remission. DS, duodenal switch; Ext V, external validation; F, female; FU, follow-up; GB, gastric band; I, insulin; IIDSG, ileal interposition with diverted sleeve gastrectomy; M, male; NI, noninsulin; NR, nonremitters; R, remitters; single, single center; V Dev, validated in internal/external cohort in model development stage; Y, yes.

Table 3—Study characteristics of validation studies

Publication reference	Source of data	Participant characteristics					Outcomes/events	Types of surgery	Study validated
		Groups and numbers	Age (years)	BMI (kg/m²)	Diabetes duration (years)	HbA$_{1c}$ (%)			
Lee et al. (44), 2015, modified	Retrospective, single, Taiwan, 2006–2013	N = 85; 44 M, 41 F	41.9 ± 10.9	39.0 ± 7.4	2.7 ± 3.1	8.1 ± 1.7	n (CR+PR) = 63 of 85 (74.1%), FU = 1 year	SG	ABCD
Lee et al. (45), 2015	Retrospective, single, Taiwan, 2006–2009	N = 157; 52 M, 105 F	35.7	39.8 ± 8.0	NA	8.3 ± 1.9	n = 111 (77.1%) at 1 year, n = 97 (71.3%) at 5 years, FU = 6 years (5–8)	RYGB, SG	ABCD
Lee et al. (46), 2015	Single, Taiwan, 2007–2013	N = 80 of 512 had BMI <30 kg/m²; F = 50 (62.5%)	47.7 ± 9.1	26.9 ± 2.2	6.5 ± 5.1	9.1 ± 1.8	n (CR) = 20 (25%), FU = 1 year	RYGB, SG	ABCD
Sampaio-Neto et al. (47), 2015	Retrospective, single, Brazil, 2012–2013	N = 70; 6 M, 64 F	47.9 ± 9.9	NA	NA	7.6 ± 1.8	n (CR+PR) = 42 (35 + 7) (60%), FU = 1 year	RYGB	DiaRem
Lee et al. (48), 2016	Retrospective, single, Taiwan, 2007–2013	N = 245; 95 M, 150 F	44.2 ± 10.4	35.7 ± 7.8	5.8 ± 5.0	8.8 ± 1.6	n = 130 (53.1%), FU = 1 year	RYGB	DiaRem, ABCD
Lee et al. (49), 2017	Prospective, single, Taiwan, 2007–2014	N = 579; 230 M, 349 F; SG 48 M, 61 F; RYGB 182 M, 288 F SG (N = 109) RYGB (N = 470)	43.2 ± 11.0 41.8 ± 10.9	35.7 ± 7.2 36.9 ± 7.2	3.3 ± 3.5 4.5 ± 4.8	8.8 ± 1.5 8.6 ± 1.7	n = 361 (62.3%)/579 at 1 year, n = 71 (49.7%)/143 at 5 years	RYGB, SG	ABCD
Mehaffey et al. (50), 2017	Prospective, U.S., 2004–2013	N = 57, 75% F, 2 years FU; N = 31, 55% F, 10 years FU 2-year FU 10-year FU	49.2 45.8	52.1 ± 1.5 48.0 ± 6.6	NA NA	8.3 7.7	n at 2 years = 37 (65%), n at 10 years = 18 (58%)	RYGB	DiaRem
Honarmand et al. (51), 2017	Retrospective, MC, U.S., 2010–2015	N = 900; 667 F	51.0 ± 9.1	49 ± 8.07	NA	7.6 ± 1.5	n = 333 (37%), FU = 1 year	RYGB	DiaRem

Table 3—Continued

Publication reference	Source of data	Participant characteristics					Outcomes/events	Types of surgery	Study validated
		Groups and numbers	Age (years)	BMI (kg/m²)	Diabetes duration (years)	HbA$_{1c}$ (%)			
Tharakan et al. (52), 2017	Retrospective, single, U.K., 2007–2014	N = 262; 105 M, 157 F	51 ± 9.5	45.3 ± 7.1	NA	8.2 ± 1.8	n = 85 (32.5%), FU = 1 year	RYGB	DiaRem
Raj et al. (53), 2017	Prospective, single, India, 2014–2015	N = 53; 26 M, 27 F	45.86 ± 11.69	43.25 ± 7.4	3 (range 0–40)	8.07 ± 1.98	n = 43 (81.1%), FU = 1 year	RYGB, SG	ABCD
Seki et al. (54), 2018	Retrospective, single, Japan, 2007–2015	N = 72; 37 M, 35 F	46.8 ± 9.0	31.7 ± 2.0	9.6 ± 6.9	8.9 ± 1.5	n (CR) = 22 (31%), n (PR) = 32 (49%), FU = 1 year	SG	ABCD
Shen et al. (55), 2018	Retrospective, Taiwan, 2011–2016	N = 128; 58 M, 70 F	42.4 ± 10.6	39.2 ± 5.8	3.2 ± 3.8	8.0 ± 1.7	n (CR) = 92 (71.9%), n (PR) = 103 (80.5%), FU = 1 year	SG	ABCD, IMS, DiaRem, Ad-DiaRem, DiaBetter
Naitoh et al. (56), 2018	Retrospective, MC, Japan, 2005–2015	N = 298; 140 M, 158 F					n (CR+PR) = 247 (82.9%), FU = 1 year	LSG, LSG/DJB	ABCD
		LSG (N = 177)	45.2	45.2	5.9 ± 6.1	7.3 ± 6.0			
		LSG/DJB (N = 131)	45.2	43.5	7.7 ± 1.7	8.3 ± 1.7			
Ahuja et al. (57), 2018	Retrospective, single, India, 2010–2015	N = 102	45.63 ± 11.12	44.85 ± 9.24	5.3 ± 5.08	8.26 ± 1.8	n = 72 (70.6%), FU = 1 year	RYGB	DiaRem, DRS, ABCD
Almalki et al. (58), 2018	Retrospective, single, Taiwan, 2007–2015	N = 406; F = 64%	42.6	36.4	3.1	8.6	n = 291 (71.7%), FU = 5 years	RYGB	ABCD
Chen et al. (59), 2018	Single, Taiwan, 2004–2012	N = 310; 114 M, 196 F	40.1 ± 11.1	37.8 ± 7.6	3.6 ± 4.4	8.6 ± 1.8	n = 224 (72.3%), FU = 5 years	RYGB, SG	IMS, ABCD
Wood et al. (60), 2018	Retrospective, single, U.S., 2002–2014	N = 520; 124 M, 396 F	46.7–53.0	NA	NA	7.2 to 7.5	n = 249; RYGB n = 173, SG n = 63, GB n = 13	RYGB, SG, GB	DiaRem

Table 3—Continued

Publication reference	Source of data	Participant characteristics					Outcomes/events	Types of surgery	Study validated
		Groups and numbers	Age (years)	BMI (kg/m²)	Diabetes duration (years)	HbA$_{1c}$ (%)			
Dicker et al. (61), 2019	Retrospective, Israel, 1999–2011	N = 2,190; 64.8% F	47.1 ± 10.9	43.5 ± 6.3	NA	7.7 ± 1.6	n = 897 (59.7%)/1,502 at 2 years, n = 782 (53.6%)/1,459 at 5 years	RYGB, SG, GB	Ad-DiaRem, DiaRem
Kam et al. (62), 2020	Retrospective, China	N = 214 at 1 year, 117 (54.7%) F; 131 at 3 years	48 (37–57)	30.6 (28.7–32.9)	6.0 (3–10)	8.0 (7.1–9.8)	113 of 214 (52.8%) at 1 year, 59 of 131 (45.0%) at 3 years	RYGB	ABCD, DiaRem, Ad-DiaRem, DiaBetter
Park et al. (63), 2020	Retrospective, Korea	N = 135, 103 (76%) F	40 ± 11	39.0 ± 6.3	1 (0–5)	7.5 (6.8–8.9)	n = 88 (65.2%), FU = 1 year	RYGB, SG	IMS
Lee et al. (64), 2020	Retrospective, 2006–2014	N = 59, 38 F	47.7 ± 12.4	37.6 ± 5.1	2.7 ± 3.1	8.3 ± 2.2	37 (62.7%) at 1 year, 24 (42.4%) at 5 years	SG	ABCD
Guerron et al. (65), 2020	Retrospective, North Carolina, 2000–2007	N = 602, 441 (73.3%) F	50.6 ± 10.2	47.1 ± 7.8	NA	7.5 ± 1.4	n (CR) = 215 (35.7%), n (only PR) = 134 (22.3%)	RYGB, SG, BPD/DS, LAGB	DiaRem

Data are mean ± SD or median (interquartile range) unless otherwise indicated. N = total number of participants. n = number of participants achieving diabetes remission. BPD, biliopancreatic diversion; CR, complete remission; DS, duodenal switch; F, female; FU, follow-up; GB, gastric band; LAGB, laparoscopic adjustable gastric band; LSG, laparoscopic sleeve gastrectomy; LSG/DJB, laparoscopic sleeve gastrectomy with duodenal-jejunal bypass; M, male; MC, multicenter; NA, not available; PR, partial remission; single, single center.

longer follow-up duration where AUC was available for more than one HbA$_{1c}$ cutoff or length of follow-up.

RESULTS

Following the initial search, we retrieved 5,825 articles. After removal of 1,660 duplicates, 4,165 publications were identified for title and abstract screening: 91 publications were identified as eligible for full text screening, and 44 were excluded, as these were conference papers or posters with limited information, especially on methods and risk of bias. The remaining 47 published articles were screened by full text; 9 were excluded after screening of the full text (reasons outlined in Supplementary Table 2). The remaining 38 published articles were retained for data extraction.

Study Characteristics

Of the 38 studies included in this review, 16 focused on model development (Tables 1 and 2) and 22 focused on external validation (Table 3). External validation studies were defined as studies validating a predefined prediction model in a different population or time period from the population/time period used to develop the model.

Model Development Studies

Of the 16 model development studies, 11 produced scoring systems, while the other 5 were logistic regression prediction models. The scoring systems were as follows: ABCD, for age, BMI, C-peptide, and duration (Lee et al. [16]), from Taiwan in 2013; Robert et al. (31) from France in 2013; Diabetes Remission score (DiaRem) (Still et al. [17,32]) from the U.S. in 2014; diabetes remission score (DRS) (Ugale et al. [33]) from India in 2014; Individualized Metabolic Surgery (IMS) score (Aminian et al. [18]) from the U.S. in 2017; Advanced-DiaRem (Ad-DiaRem) (Aron-Wisnewsky et al. [34]) from France in 2017; DiaBetter (Pucci et al. [35]) from the U.K. in 2017; Dia-Rem2 (Still et al. [36]) in 2018, an updated version of the preexisting Dia-Rem model developed by the same group; 5-year diabetes remission (5y-DR) (Debédat et al. [37]) from France in 2018; Metabolic Surgery Diabetes Remission (MDR) score (Moh et al. [38]) from Japan in 2020; and Umemura et al. (39) from Singapore in 2020. The five logistic
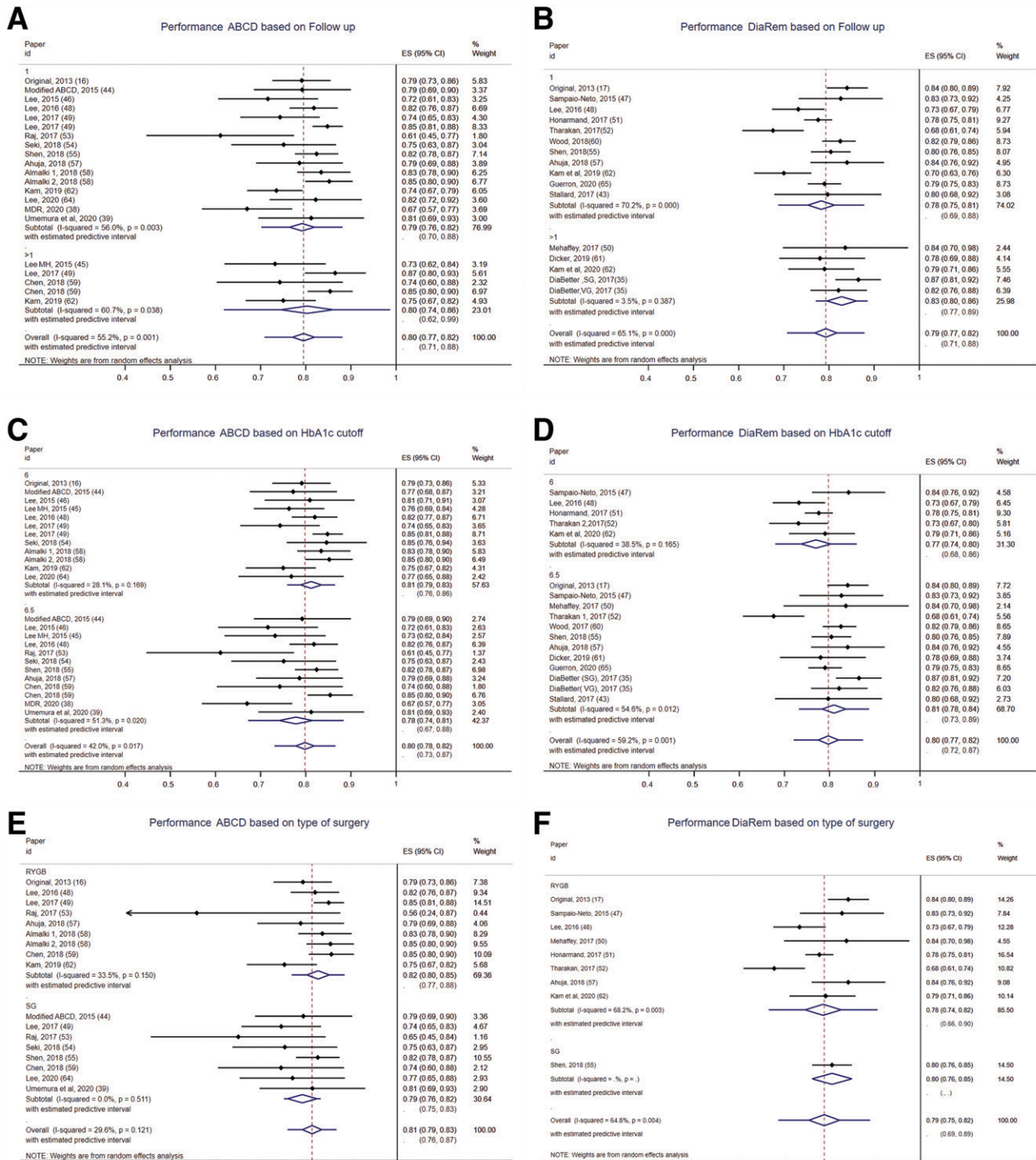
Figure 2—Meta-analysis. *A*: Performance of ABCD based on follow-up. *B*: Performance of DiaRem based on follow-up. *C*: Performance of ABCD based on HbA$_{1c}$ cutoff defining diabetes duration. *D*: Performance of DiaRem based on HbA$_{1c}$ cutoff defining diabetes duration. *E*: Performance of ABCD based on type of bariatric surgery. *F*: Performance of DiaRem based on type of bariatric surgery. ES, effect size.

regression models were those of Hayes et al. (40) from New Zealand in 2011, Dixon et al. (13) from Taiwan in 2013, Ramos-Levi et al. (41) from Spain in 2014, Cotillard et al. (42) from France in 2015, and Stallard et al. (43) from Canada in 2016 (Table 2).

Participants
Of 16 studies, 14 used retrospective data and 2 used prospective data (13,40). Eight studies included participants who had undergone RYGB (13,16,17,34,36,37,40,42), two studies included participants who had SG (33,39), and the remaining

studies included more than one type of bariatric surgical procedure (18,31,35,38,41,43). Study sample size ranged from 46 to 690 participants, with a female preponderance except in two studies, DRS (Ugale et al. [33]) and the study of Umemura et al. (39), which had higher male

representation. The mean age of participants ranged between 36.5 and 57.6 years and mean BMI from 23.4 (33) to 49.7 kg/m^2. Diabetes duration was available for all studies except that of Umemura et al. (39). Mean diabetes duration ranged from 2.1 years in the study by Lee et al. (16) (ABCD model) to 9.9 years in that of Ugale et al. (33) (DRS model). Preoperative HbA$_{1c}$ ranged from 6.8% (51 mmol/mol) in the study by Still et al. (17) (DiaRem score) to 9.1% (76 mmol/mol) in the study of Dixon et al. (13).

Follow-up Duration
The median follow-up range was 1–5 years. The majority of the studies reported remission rates at 1 year. Three studies reported longer follow-up: 2 years in DiaBetter (Pucci et al. [35]) and 5 years in IMS (Aminian et al. [18]) and 5y-DR (Debédat et al. [37]).

Outcome Definition
Different definitions for diabetes remission were noted with some focusing on complete diabetes remission (defined as HbA$_{1c}$ <6.0% [42 mmol/mol] and no antidiabetes medication for at least 12 months) (13,16,31,37) and others combining complete and partial diabetes remission (defined as <6.5% [48 mmol/mol] and off medications for 12 months) (17,18,34,38,39). DiaRem2 (36) and Stallard et al. (43) defined diabetes remission as an HbA$_{1c}$ of <5.7% (39 mmol/mol) and ≤5.9% (41 mmol/mol), respectively, after patients were off antidiabetes medications at 12 months.

Method/Analysis and Presentation
Predictors in the models varied and included age, baseline BMI, C-peptide, diabetes duration, HbA$_{1c}$, insulin use, glucose-lowering medications, sex, and micro- and macrovascular complications. 5y-DR (37) included postoperative variables as predictors in the prediction model. The number of predictors ranged from 2 (40) to 10 (42). For five prediction models a logistic regression model was proposed; for the models of Dixon et al. (13) and Hayes et al. (40) a logarithmic equation was given, while Ramos-Levi et al. (41), Cotillard et al. (42), and Stallard et al. (43) defined the predictors to be included in the prediction model but gave no equation in their publication.

The method for deriving the scoring system varied among the 11 models. For DiaRem (17) and DiaRem2 (36) investigators reported hazard ratios using Cox regression and odds ratios of the final logistic models, respectively, to create a scoring system. Umemura et al. (39) used a weighing algorithm and gave an odds ratio. In IMS (18) a nomogram and benchmarks selected by an expert panel were used. Ad-DiaRem (34) and 5y-DR (37) used machine learning; Ad-DiaRem used a sparse support vector machine and formulated a linear integer programming task, and 5y-DR used a fully corrective binning approach to assign intervals and weight for each variable. MDR (38) used quartile and tertile cutoffs to obtain the weighting of each of the predictors in the scoring system. ABCD (16), Robert et al. (31), DRS (33), and DiaBetter (35) offered no information on how the weighting for individual predictors was decided.

Performance
For representing the model performance, AUC was presented in the publications of Ad-DiaRem (34), Dixon et al. (13), Robert et al. (31), Ramos-Levi et al. (41), Stallard et al. (43), DiaBetter (35), DiaRem2 (36), 5y-DR (37), and MDR (38). We calculated the AUC for ABCD (16), DiaRem (17) and the study by Umemura et al. (39) (Supplementary Material). No AUC or performance was reported for DRS (33), IMS (18) or by Hayes et al. (40) or Cotillard et al. (42), and data in the publications were insufficient to calculate these.

Of 12 prediction models for which AUC was available, 10 (Dixon et al. [13], Ramos-Levi et al. [41], Stallard et al. [43], DiaRem [17], Robert et al. [31], Ad-DiaRem [34], DiaBetter [35], DiaRem2 [36], 5y-DR [37], and Umemura et al. [39]) had excellent discrimination (0.80–0.89) and two (ABCD [16] and MDR [38]) had acceptable discrimination (0.70–0.79), irrespective of diabetes remission definition (Table 1).

Risk of Bias Assessment
The studies developing DiaRem (17) and Ad-DiaRem (34) were found to have low risk of bias, and that of Dixon et al. (13) was of unclear risk. The remaining model development studies had high risk of bias, mainly due to deficiencies in the

analysis domain. However, the applicability in practice was of low risk in all model development studies (Supplementary Table 5).

Validation Studies
We identified 22 studies externally validating the prediction models (44–65). Study characteristics are summarized in Table 3. Of the 22 external validation studies, 9 validated the ABCD score, 6 validated DiaRem, 1 validated IMS, and the remaining 6 compared two or more models.

Participants
In 19 studies retrospective data were used (44–48,51,52,54–65), and in 3 data were collected prospectively (49,50,53). The sample size ranged from 53 (53) to 2,190 (61), and mean age ranged from 35.7 years (45) to 51.0 years (51,52,65). All studies had a female predominance except one (54). Mean BMI ranged from 26.9 kg/m^2 (46) to 52.1 kg/m^2 (50). Median diabetes duration ranged from 1 year (63) to 9.6 years (54), and mean presurgery HbA$_{1c}$ ranged from 7.2% (55 mmol/mol) (60) to 9.1% (76 mmol/mol) (46).

Eight studies included participants who underwent RYGB (47,48,50–52,57,58,62), five included participants who underwent SG (44,54–56,64), and the remaining nine included a range of surgery types (45,46, 49,53,59–61,63,65) (Table 3).

Follow-up Duration
One study had a follow-up period of 10 years (50), six studies had a follow-up period of 5 years (45,49,58,59,61,64), one had 3 years' follow-up (62), and the remaining studies had follow-up of 1 year.

Outcome Definition
For defining diabetes remission, the HbA$_{1c}$ cutoff was taken as 5.7% (39 mmol/mol) in two studies (51,65) and 6.0% (42 mmol/mol) in 13 studies (44–49,52,54, 55,58,62,64), and in the remaining studies diabetes remission was defined as HbA$_{1c}$ ≤6.5% (48 mmol/mol).

Performance
Although 16 models were identified in model development studies, few of these were externally validated in more than one external cohort; models that were externally validated in more than one

cohort were predominantly scoring systems. Direct comparison of the models was seen in only six studies (48,55, 57,59,61,62).

Here we present the assessment of validation studies based on the prediction models validated. As ABCD and DiaRem scores were validated most frequently, we present these studies first followed by the remainder of the prediction models externally validated.

ABCD Score

In the original model development paper, the authors also reported an external validation in a new cohort (16). We calculated the AUC to be 0.79 (95% CI 0.73–0.86) (acceptable discrimination) (Table 1 and Fig. 2) and calibration (E-to-O ratio) as 1.01 in the external cohort. In a subsequent study, ABCD score cutoff values for each variable were modified (44). In this cohort, we calculated AUC as 0.77 (0.68–0.87) and 0.79 (0.69–0.90) for complete and partial diabetes remission, respectively (44). Calibration was not available.

The ABCD score with the new cutoffs (44) has been externally validated in 13 studies (45,46,48,49,53–59,62,64). Of these 13 validation studies, 5 looked at long-term diabetes remission at 3–5 years (45,49,59,62,64) and the remaining 8 at 1 year. In one study poor discrimination was found (53), while in others the discrimination was found to be acceptable to excellent depending on the type of surgery and follow-up duration. Model development studies for MDR (38) and Umemura et al. (39) also validated ABCD in their cohort and found the performance to be poor and excellent, respectively.

It was difficult to ascertain calibration score, as it was not widely available, and when available the results were inconsistent. Calibration was only mentioned in two studies (55,62) and found to be overestimating by 13% (55) and 12% (62) for diabetes remission at 1 year and underestimating by 15% (62) at 3 years.

ABCD Meta-analysis. For ABCD, meta-analysis of the results from multiple studies showed acceptable discrimination with AUC of 0.79 (95% CI 0.76–0.82) for 1-year follow-up and 0.80 (0.74–0.86) for longer-term follow-up (Fig. 2A). At the different HbA$_{1c}$ cutoffs, discrimination was excellent with an AUC of 0.81 (95% CI 0.79–0.83)

for HbA$_{1c}$ cutoff of 6.0% (42 mmol/mol) and acceptable at 0.78 (0.74–0.81) for HbA$_{1c}$ cutoff of 6.5% (48 mmol/mol) (Fig. 2B). For RYGB, meta-analysis showed excellent discrimination for ABCD with an AUC of 0.82 (95% CI 0.80–0.85), while for SG, discrimination was acceptable with AUC of 0.79 (0.76–0.82) (Fig. 2C).

DiaRem Score

The DiaRem score has been externally validated in 11 studies (47,48,50–52,55, 57,60–62,65) (Table 1 and Fig. 2). Three studies looked at long-term (>1 year) (50,61,62) diabetes remission, and the remaining focused on remission at 1 year.

Excellent discrimination was found for five external validation studies (47,50,55, 57,60), acceptable for five (48,51,61,62, 65), and poor for one (52).

Calibration was presented in two studies (55,62). We were able to calculate the E-to-O ratio for a further six studies: DiaRem underestimated the probability of diabetes remission in the studies by Ahuja et al. (57) (E-to-O ratio 0.67) and Mehaffey et al. (50) (E-to-O ratio 0.63 at 2 years and 0.71 at 10 years). It overestimated the probability of diabetes remission in the other four studies (47,48,51,52), with E-to-O ratios of 1.31, 1.71, 1.14, and 1.25 in the studies by Honarmand et al. (51), Lee et al. (48), Sampaio-Neto et al. (47), and Tharakan et al. (52), respectively. Calibration was inconsistent across the studies.

In the model development studies of Ad-DiaRem (34), DiaBetter (35), DiaRem2 (36), and 5y-DR (37), DiaRem was validated in their cohorts and excellent discrimination was found, while Stallard et al. (43) found good discrimination.

DiaRem Meta-analysis. In meta-analysis, discrimination for DiaRem was as follows: AUC 0.78 (95% CI 0.75–0.81) for short-term and 0.83 (95% CI 0.80–0.86) for longer-term follow-up (Fig. 2B). At HbA$_{1c}$ cutoffs of 6.0% (42 mmol/mol) and 6.5% (48 mmol/mol), the AUCs were 0.77 (95% CI 0.74–0.80) and 0.81 (95% CI 0.78–0.84), respectively (Fig. 2D).

For RYGB, meta-analysis showed acceptable discrimination for DiaRem with AUC of 0.78 (95% CI 0.74–0.82). No meta-analysis was performed for SG, as there was only one study identified with validation of DiaRem in an SG cohort (Fig. 2F).

Performance of Other Prediction Models

The discrimination scores for other prediction model are summarized in Table 1. The IMS score was externally validated by three validation studies (55,59,63) and one model development study (Umemura et al. [39]). Discrimination was found to be excellent in the study by Shen et al. (55) and acceptable in the RYGB cohort of Chen et al. (59) and in the study by Park et al. (63) but poor in the SG cohort of Chen et al. (59) and Umemura et al. (39).

Ad-DiaRem (34) was externally validated in three validation studies (55,61, 62) and 5y-DR model development study (37). Kam et al. (62) in their study found acceptable performance, while the other three found excellent performance.

DiaBetter (35), Dixon et al. (13), and Ramos-Levi et al. (41) were noted to have excellent discrimination (55), and DRS (33) had good performance in one external validation study (57). The Robert et al. (31) and Hayes et al. (40) prediction models performed poorly in one external validation with an AUC <0.70 (55). No external validation studies are available for DiaRem2 (36), Stallard et al. (43), Cotillard et al. (42), 5y-DR (37), MDR (38), or Umemura et al. (39).

Calibration in an external validation study for Ad-DiaRem, DiaBetter, Dixon et al., and Ramos-Levi et al. found the models to be overestimating with predicted (or expected)-to-observed ratios of 1.06, 1.05, 1.13, and 1.12, respectively (55). Hayes et al. and Robert et al. were noted to be overestimating by 23–30% with predicted-to-observed ratios of 1.23 and 1.30, respectively (55).

Risk of Bias Assessment

Guerron et al. (65) had low risk of bias, three studies (45,50,53) were classified as high risk of bias as a result of the analysis domain, and the remaining external validation studies had unclear risk of bias (Supplementary Table 6). We rated risk of bias in the analysis domain as unclear either if information on missing data was not reported or was not included in the analysis or if model performance was not reported; if neither was reported, we rated the domain as high risk of bias. However, no concerns were raised in terms of applicability, with all studies rated as low risk.

Ideally, a sensitivity analysis restricted to low risk of bias studies should be performed. In our review, only one external

validation study—validating the DiaRem model—was rated as low risk of bias (65); results of the meta-analyzed studies for DiaRem were consistent with the findings of this study.

CONCLUSIONS

In this systematic review, we have identified currently available models for predicting diabetes remission following bariatric surgery. We assessed and compared the performance of these models and evaluated their applicability in clinical settings. The most externally validated models in our review were ABCD and DiaRem. Although the ABCD and DiaRem models were primarily developed for predicting diabetes remission at 1-year follow-up, they have been validated in studies predicting long-term diabetes remission. The AUC estimate for DiaRem for long-term diabetes remission and diabetes remission defined with an HbA_{1c} cutoff of 6.5% (48 mmol/mol) was higher than for ABCD. The AUC for ABCD for predicting short-term remission and diabetes remission defined by an HbA_{1c} cutoff of 6.0% (42 mmol/mol), was higher than that for DiaRem. Specifically for patients who underwent RYGB, AUC was higher in ABCD than for DiaRem. However, in all instances, CIs overlapped.

Due to the lack of discrimination (AUC) score with 95% CI, we were not able to include in our meta-analysis three studies (34,37,43) conducted on patients who underwent RYGB and validating DiaRem that otherwise showed excellent performance. Furthermore, many studies validating ABCD were conducted by the same authors who developed the ABCD model and included patient cohorts similar to the derivation population, raising the possibility of bias based on population selection. It was therefore not possible to determine whether one model was better than the other.

Remission of diabetes is an important outcome for patients considering bariatric surgery. A project by Diabetes UK, led by patients with type 2 diabetes and their carers, identified diabetes cure or reversal as a top research priority (66). With the increasing number of patients with obesity and type 2 diabetes now being offered bariatric surgery, it is important to identify those who are more likely to achieve remission. This

will enable patients and health care professionals to make informed choices when considering different treatment options. However, given the wide choice of prediction models currently available, it is difficult to identify the ones that best predict remission and are easy to use in routine clinical practice.

The models identified in our review had certain common characteristics in relation to the predictors included and the duration of follow-up, which for most studies was 12 months. On the other hand, there was considerable heterogeneity in the definition of diabetes remission, cohort size, and populations studied, including types of bariatric surgery, thereby adding to the difficulty in comparing these models. We found significant variation in the threshold for HbA_{1c} used to define diabetes remission, with cutoffs ranging from 5.7 to 6.5% (39 to 48 mmol/mol), and with some studies using a combination of partial and complete remission. However, in this review, we found that the definition did not affect the performance of the prediction models significantly.

Duration of diabetes remission is an important consideration in assessment of the benefits of bariatric surgery in patients with type 2 diabetes. In our review, we observed that 13 of 16 model development studies were designed with the aim of predicting diabetes remission at 1 year, thus underscoring the need for longer follow-up of cohorts (18). The rate of diabetes remission has been inversely associated with diabetes duration and has been noted to be greatest in patients with shorter diabetes duration (12,67). Moreover, diabetes remission is highest during the first year following the intervention and declines over subsequent years and with longer follow-up (7–9). In the prospective Swedish Obese Subjects (SOS) study, with follow-up of >18 years, the incidence of diabetes remission was 72.3% at 2 years, 38.1% at 10 years, and 30.4% at 15 years (68). In a randomized controlled trial with 5 years' follow-up, findings indicated diabetes relapse in 53% of patients in the RYGB group and 37% in the biliopancreatic diversion group among patients who achieved diabetes remission at 2 years' follow-up (9). Similar results were reported in a retrospective multisite study from the U.S. with 5 years' follow-up (69). These findings suggest that diabetes may relapse over time and that in a

high proportion of patients, remission of diabetes may only be achieved for a short term. Despite this, short-term diabetes remission may offer huge clinical and financial benefits to patients and health care systems. Besides the benefit of reduction in the incidence of micro- and macrovascular diabetes–related complications, short-term diabetes remission, through freedom from diabetes medications and reduced need for monitoring, may motivate patients to maintain weight loss and enhance their quality of life.

Future studies should therefore include a uniform and agreed definition of diabetes remission and a longer follow-up period to determine the effects of bariatric surgery on long-term diabetes remission. This is particularly important when considering the cost-effectiveness of bariatric surgery.

The outcomes of bariatric surgery such as weight loss and long-term metabolic benefit vary with the type of bariatric procedure (5,70,71). A network meta-analysis showed that the probability of achieving diabetes remission was greatest in mini–gastric bypass (91.2%), followed by biliopancreatic diversion without duodenal switch (87.3%), laparoscopic SG (61.4%), RYGB (59.3%), gastric banding (29.6%), and then great curvature plication (18.6%) (71). Despite this, none of the prediction models included the type of surgery as a predictor. However, when we analyzed the performance of prediction models in RYGB and SG separately, we found no major differences between the two procedures. With many new bariatric procedures becoming available, there is a need to develop and validate the models across the various bariatric procedures.

The indication for bariatric surgery in patients with BMI <35 kg/m^2 is contentious, and currently none of the guidelines recommend bariatric surgery in nonobese individuals. In a recent systematic review and meta-analysis Ji et al. (72) evaluated 12 studies examining the impact of bariatric surgery in patients with type 2 diabetes and BMI <30 kg/m^2 over a follow-up period ranging from 6 months to 3 years. They found a 1.58% (∼16 mmol/mol) reduction in HbA_{1c} at 2 years using a random-effects model (72). However, investigators of other studies comparing the impact of bariatric surgery in populations with and without obesity observed that surgery in a population

without obesity is a less effective tool for diabetes management (73). We found two studies—a model development study by Ugale et al. (33) and the validation study of Lee et al. (46)—with a focus on cohorts with mean BMI ≤30 kg/m^2; Ugale et al. did not provide model discrimination, and Lee et al. found acceptable discrimination in this normal weight population. Based on available data, it is difficult to assess the performance of prediction models in those with low BMI. The impact of BMI on the performance of the prediction models is important and requires further study.

Susceptibility to type 2 diabetes is known to vary among people of different ethnicities, and it is likely that these differences may extend to remission of diabetes following bariatric surgery. In the studies included in our review, the test cohort for DiaRem was 98% Caucasian, while for ABCD, the participants were from five Asian clinics. We identified one validation study where Wood et al. (60), validating the DiaRem score in a White and Hispanic population, noted an AUC of 0.84 (0.80–0.88) and 0.79 (0.71–0.86) in White and Hispanic patients, respectively. A meta-analysis including 14 studies showed greater weight loss in Caucasians compared with African Americans, but no difference was noted in the outcome of diabetes remission between these two ethnic groups (74,75). None of the prediction models identified ethnicity as a predictor, and data on direct comparisons between ethnic groups were limited to the above-mentioned studies. We were therefore unable to explore any possible differences between ethnicity and incidence of diabetes remission post–bariatric surgery.

Five of the models, including two scoring systems, ABCD (16) and DRS (33), and three logistic regression models, of Dixon et al. (13), Ramos-Levi et al. (41), and Cotillard et al. (42), included C-peptide levels (13,16,33,41,42) as one of the predictors. C-peptide can be measured as urinary C-peptide, urinary C-peptide–to–creatinine ratio, or venous blood C-peptide levels measured as random, fasting, or in a stimulated state (glucagon stimulation test, mixed-meal tolerance test) (76). These factors may pose difficulties in standardization and can present as a limitation in using certain scoring systems that have C-peptide as one of the predictors. Moreover, C-peptide is not measured routinely in the diagnosis or management of type 2 diabetes in most clinical settings. Prediction models using C-peptide, therefore, cannot be widely used by primary care physicians or in the early stages of weight management consultation. Models such as DiaRem and Ad-DiaRem that focus predominantly on routinely measured clinical parameters may therefore have greater applicability across a wider range of clinical settings. If C-peptide is available, however, ABCD (16) is a reliable prediction model with a similar predictive performance and has the advantage of being validated in different bariatric procedures and for long-term diabetes remission (49,59,62). 5y-DR (37) included postoperative number of glucose-lowering medications, fasting capillary blood glucose, weight loss, and 1-year remission to predict long-term diabetes remission. Postoperative parameters will not be available in the clinical consultation setting for bariatric surgery, and, hence, use of this prediction model is limited.

Treatment with insulin has been used in many models as a predictor. Patient preference for noninsulin treatments and therapeutic inertia are recognized causes for delay in treatment with insulin (77). In instances where insulin treatment is delayed, treatment with insulin as a predictor can overestimate the chances of remission. Conversely, if insulin is initiated early, the possibility of diabetes remission may be underestimated.

The inconsistency in calibration scores with existing models either overestimating or underestimating the observed remission rates suggests that there are other variables that could influence remission. The utility of the prediction models largely depends on the clinical setting and resources available. The choice of a model used to predict remission must therefore be tailored to these factors.

Strengths and Limitations

We believe our study is the first systematic review summarizing prediction model performance for diabetes remission in patients undergoing bariatric surgery. We calculated the discrimination score (AUC) for the studies where data were available and where AUC was not reported by the authors themselves.

While the robust search strategy used in this review is a strength of our study, there are certain limitations: we restricted our search to articles published in English and published in the last 15 years. We were also unable to contact the authors for further information regarding the performance of prediction models, where pertinent information was not available. It is possible that some relevant articles were not included in the review and meta-analysis. However, not many prediction models were available before our search date and the likely impact of this on our findings would be minimal.

While the key messages were consistent, a large proportion of the studies were conducted in small cohorts of patients with short duration of follow-up. In the majority of external validation studies routinely collected data were used; consequently, follow-up data were not available for all patients who underwent bariatric surgery. While the studies had predefined inclusion and exclusion criteria for participants, with complete data at 1 year of follow-up for those included, there remains a possibility of selection bias due to lack on information on patients lost to follow-up in the routinely collected source data. Validation studies in large cohorts with longer follow-up are therefore needed to overcome these limitations.

Conclusion

This systematic review identified 16 prediction models, with DiaRem (17) and ABCD (16) as the two most widely validated models to predict diabetes remission following bariatric surgery. Newer models published in the last 3–4 years showed promising results in test cohorts, but there is a limited number of external validation studies. More external validation studies are needed for assessing the performance and clinical applicability of the new prediction models. Future studies should also examine these models in real-world clinical settings to assess the impact on patient outcomes.

References

1. Mingrone G, Panunzi S, De Gaetano A, et al. Bariatric surgery versus conventional medical therapy for type 2 diabetes. N Engl J Med 2012;366:1577–1585
2. Schauer PR, Bhatt DL, Kirwan JP, et al.; STAMPEDE Investigators. Bariatric surgery versus intensive medical therapy for diabetes - 5-year outcomes. N Engl J Med 2017;376:641–651
3. Mistry P, Currie V, Super P, le Roux CW, Tahrani AA, Singhal R. Changes in glycaemic control, blood pressure and lipids 5 years following laparoscopic adjustable gastric banding combined with medical care in patients with type 2 diabetes: a longitudinal analysis. Clin Obes 2018;8:151–158
4. Avenell A, Robertson C, Skea Z, et al. Bariatric surgery, lifestyle interventions and orlistat for severe obesity: the REBALANCE mixed-methods systematic review and economic evaluation. Health Technol Assess 2018;22:1–246
5. Singh P, Subramanian A, Adderley N, et al. Impact of bariatric surgery on cardiovascular outcomes and mortality: a population-based cohort study. Br J Surg 2020;107:432–442
6. Singh P, Adderley N, Subramanian A, Gokhale K, Singhal R, Toulis KA, et al. The impact of bariatric surgery on incident microvascular complications in patients with type 2 diabetes: a matched controlled population-based retrospective cohort study. Diabetes Care 2021;44:116–124
7. Brethauer SA, Aminian A, Romero-Talamás H, et al. Can diabetes be surgically cured? Long-term metabolic effects of bariatric surgery in obese patients with type 2 diabetes mellitus. Ann Surg 2013;258:628–636
8. Schauer PR, Mingrone G, Ikramuddin S, Wolfe B. Clinical outcomes of metabolic surgery: efficacy of glycemic control, weight loss, and remission of diabetes. Diabetes Care 2016;39:902–911
9. Mingrone G, Panunzi S, De Gaetano A, et al. Bariatric-metabolic surgery versus conventional medical treatment in obese patients with type 2 diabetes: 5 year follow-up of an open-label, single-centre, randomised controlled trial. Lancet 2015;386:964–973
10. Rubino F, Nathan DM, Eckel RH, et al.; Delegates of the 2nd Diabetes Surgery Summit. Metabolic surgery in the treatment algorithm for type 2 diabetes: a joint statement by international diabetes organizations. Diabetes Care 2016;39:861–877
11. Mahmood F, Munasinghe A, MacAno C, et al. Factors predicting remission or improvement of type 2 diabetes after gastric bypass. Surg Obes Relat Dis 2016;12(7 Suppl.):S165–S166
12. Panunzi S, Carlsson L, De Gaetano A, et al. Determinants of diabetes remission and glycemic control after bariatric surgery. Diabetes Care 2016;39:166–174
13. Dixon JB, Chuang L-M, Chong K, et al. Predicting the glycemic response to gastric bypass surgery in patients with type 2 diabetes. Diabetes Care 2013;36:20–26
14. Blackstone R, Bunt JC, Cortes MC, Sugerman HJ. Type 2 diabetes after gastric bypass: remission in five models using HbA1c, fasting blood glucose, and medication status. Surg Obes Relat Dis 2012;548–555
15. Park JY. Prediction of diabetes remission after bariatric or metabolic surgery. J Obes Metab Syndr 2018;27:213–222
16. Lee W-J, Hur KY, Lakadawala M, et al. Predicting success of metabolic surgery: age, body mass index, C-peptide, and duration score. Surg Obes Relat Dis 2013;9:379–384
17. Still CD, Wood GC, Benotti P, 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
18. Aminian A, Brethauer SA, Andalib A, et al. Individualized Metabolic Surgery score: procedure selection based on diabetes severity. Ann Surg 2017;266:650–657
19. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097
20. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7:e32844
21. EPPI-Reviewer. Accessed 8 November 2017. Available from https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4
22. Distiller SR, 2020. Accessed 13 April 2020. Available from https://www.evidencepartners.com/products/distillersr-systematic-review-software/
23. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744
24. Wolff RF, Moons KGM, Riley RD, et al.; PROBAST Group. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–58
25. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606
26. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012;98:683–690
27. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–138
28. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315–1316
29. Zhang R, Borisenko O, Telegina I, et al. Systematic review of risk prediction models for diabetes after bariatric surgery. Br J Surg 2016;103:1420–1427
30. Hanson RK. Assessing the calibration of actuarial risk scales: a primer on the E/O index, 2016. Accessed on 5 November 2020. Available from https://journals.sagepub.com/doi/10.1177/0093854816683956
31. Robert M, Ferrand-Gaillard C, Disse E, et al. Predictive factors of type 2 diabetes remission 1 year after bariatric surgery: impact of surgical techniques. Obes Surg 2013;23:770–775
32. Still CD, Wood GC, Benotti P, et al. A probability score for preoperative prediction of type 2 diabetes remission following RYGB surgery. Lancet Diabetes Endocrinol. 1 January 2015 [Epub ahead of print]. Accessed 28 October 2019. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3932625/
33. Ugale S, Gupta N, Modi KD, 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
34. Aron-Wisnewsky J, Sokolovska N, Liu Y, et al. The advanced-DiaRem score improves prediction of diabetes remission 1 year post-Roux-en-Y gastric bypass. Diabetologia 2017;60:1892–1902
35. Pucci A, Tymoszuk U, Cheung WH, 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;35:360–367
36. Still CD, Benotti P, Mirshahi T, Cook A, Wood GC. DiaRem2: incorporating duration of diabetes to improve prediction of diabetes remission after metabolic surgery. Surg Obes Relat Dis 2019;15:717–724
37. Debédat J, Sokolovska N, Coupaye M, et al. Long-term relapse of type 2 diabetes after Roux-en-Y gastric bypass: prediction and clinical relevance. Diabetes Care 2018;41:2086–2095
38. Moh MC, Cheng A, Tan CH, 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. Obes Surg 2020;30:3387–3393
39. Umemura A, Sasaki A, Nitta H, et al. Prognostic factors and a new preliminary scoring system for remission of type 2 diabetes mellitus

after laparoscopic sleeve gastrectomy. Surg Today 2020;50:1056–1064

40. 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

41. Ramos-Levi AM, Matia P, Cabrerizo L, et al. Statistical models to predict type 2 diabetes remission after bariatric surgery. J Diabetes 2014;6:472–477

42. Cotillard A, Poitou C, Duchâteau-Nguyen G, et al. Type 2 diabetes remission after gastric bypass: what is the best prediction tool for clinicians? Obes Surg 2015;25:1128–1132

43. 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;42:573–580

44. Lee W-J, Almulaifi A, Tsou JJ, Ser K-H, Lee Y-C, Chen S-C. Laparoscopic sleeve gastrectomy for type 2 diabetes mellitus: predicting the success by ABCD score. Surg Obes Relat Dis 2015;11:991–996

45. Lee MH, Lee W-J, Chong K, et al. Predictors of long-term diabetes remission after metabolic surgery. J Gastrointest Surg 2015;19:1015–1021

46. Lee W-J, Almulaifi A, Chong K, 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

47. Sampaio-Neto J, Nassif LS, Branco-Filho AJ, et al. Eternal validation of the DiaRem score as remission predictor of diabetes mellitus type 2 in obese patients undergoing Roux-en-Y gastric bypass. Arq Bras Cir Dig 2015;28(Suppl. 1):19–22

48. Lee W-J, Chong K, Chen S-C, et al. Preoperative prediction of type 2 diabetes remission after gastric bypass surgery: a comparison of DiaRem scores and ABCD scores. Obes Surg 2016;26:2418–2424

49. Lee W-J, Chong K, Aung L, Chen S-C, Ser K-H, Lee Y-C. Metabolic surgery for diabetes treatment: sleeve gastrectomy or gastric bypass? World J Surg 2017;41:216–223

50. Mehaffey JH, Mullen MG, Mehaffey RL, et al. Type 2 diabetes remission following gastric bypass: does diarem stand the test of time? Surg Endosc 2017;31:538–542

51. Honarmand K, Chetty K, Vanniyasingam T, Anvari M, Chetty VT. Type 2 diabetes remission rates 1-year post-Roux-en-Y gastric bypass and validation of the DiaRem score: the Ontario Bariatric Network experience. Clin Obes 2017;7:176–182

52. Tharakan G, Scott R, Szepietowski O, et al. Limitations of the DiaRem score in predicting remission of diabetes following Roux-En-Y gastric bypass (RYGB) in an ethnically diverse population from a single institution in the UK. Obes Surg 2017;27:782–786

53. Raj PP, Bhattacharya S, Saravana Kumar S, et al. Do bariatric surgery-related type 2 diabetes remission predictors add clinical value? A study on Asian Indian obese diabetics. Obes Surg 2017;2113–2119

54. Seki Y, Kasama K, Yasuda K, Yokoyama R, Porciuncula JP, Kurokawa Y. The effects of laparoscopic sleeve gastrectomy with duodenojejunal bypass on Japanese patients with BMI < 35 kg/m² on type 2 diabetes mellitus and the prediction of successful glycemic control. Obes Surg 2018;28:2429–2438

55. Shen SC, Wang W, Tam KW, et al. Validating risk prediction models of diabetes remission after sleeve gastrectomy. Obes Surg 2019;29:221–229

56. Naitoh T, Kasama K, Seki Y, et al. Efficacy of sleeve gastrectomy with duodenal-jejunal bypass for the treatment of obese severe diabetes patients in Japan: a retrospective multicenter study. Obes Surg 2018;28:497–505

57. Ahuja A, Tantia O, Chaudhuri T, et al. Predicting remission of diabetes post metabolic surgery: a comparison of ABCD, diarem, and DRS scores. Obes Surg 2018;28:2025–2031

58. Almalki OM, Lee W-J, Chong K, Ser K-H, Lee Y-C, Chen S-C. Laparoscopic gastric bypass for the treatment of type 2 diabetes: a comparison of Roux-en-Y versus single anastomosis gastric bypass. Surg Obes Relat Dis 2018;14:509–515

59. Chen J-C, Hsu N-Y, Lee W-J, Chen S-C, Ser K-H, Lee Y-C. 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;14:640–645

60. Wood GC, Horwitz D, Still CD, et al. Performance of the DiaRem score for predicting diabetes remission in two health systems following bariatric surgery procedures in Hispanic and non-Hispanic white patients. Obes Surg 2018;28:61–68

61. Dicker D, Golan R, Aron-Wisnewsky J, 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;29:796–804

62. Kam H, Tu Y, Pan J, 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. Obes Surg 2020;30:2147–2157

63. Park DG, Chung Y, Kim SH, Kim YJ. Diabetes remission rate after sleeve gastrectomy or Roux-en-Y gastric bypass; utilizing Individualized metabolic surgery score for Korean patients. J Metab Bariatr Surg 2020;9:13–18

64. Lee MH, Almalki OM, Lee WJ, Chen SC, Chen JC, Wu CC. Laparoscopic sleeve gastrectomy for type 2 diabetes mellitus: long-term result and recurrence of diabetes. Obes Surg 2020;30:3669–3674

65. Guerron AD, Perez JE, Risoli T Jr, Lee HJ, Portenier D, Corsino L. Performance and improvement of the DiaRem score in diabetes remission prediction: a study with diverse procedure types. Surg Obes Relat Dis 2020;16:1531–1542

66. Finer S, Robb P, Cowan K, Daly A, Robertson E, Farmer A. Top ten research priorities for type 2 diabetes: results from the Diabetes UK-James Lind Alliance Priority Setting Partnership. Lancet Diabetes Endocrinol 2017;5:935–936

67. Jans A, Näslund I, Ottosson J, Szabo E, Näslund E, Stenberg E. Duration of type 2 diabetes and remission rates after bariatric surgery in Sweden 2007-2015: a registry-based cohort study. PLoS Med 2019;16:e1002985

68. Sjöström L, Peltonen M, Jacobson P, et al. Association of bariatric surgery with long-term remission of type 2 diabetes and with microvascular and macrovascular complications. JAMA 2014;311:2297–2304

69. Arterburn DE, Bogart A, Sherwood NE, et al. A multisite study of long-term remission and relapse of type 2 diabetes mellitus following gastric bypass. Obes Surg 2013;23:93–102

70. Dogan K, Gadiot RP, Aarts EO, et al. Effectiveness and safety of sleeve gastrectomy, gastric bypass, and adjustable gastric banding in morbidly obese patients: a multicenter, retrospective, matched cohort study. Obes Surg 2015;25:1110–1118

71. Ding L, Fan Y, Li H, et al. Comparative effectiveness of bariatric surgeries in patients with obesity and type 2 diabetes mellitus: a network meta-analysis of randomized controlled trials. Obes Rev 2020;21:e13030

72. Ji G, Li P, Li W, et al. The effect of bariatric surgery on Asian patients with type 2 diabetes mellitus and body mass index < 30 kg/m²: a systematic review and meta-analysis. Obes Surg 2019;29:2492–2502

73. Segal-Lieberman G, Segal P, Dicker D. Revisiting the role of BMI in the guidelines for bariatric surgery. Diabetes Care 2016;39(Suppl. 2):S268–S273

74. Admiraal WM, Celik F, Gerdes VE, Dallal RM, Hoekstra JB, Holleman F. Ethnic differences in weight loss and diabetes remission after bariatric surgery: a meta-analysis. Diabetes Care 2012;35;1951–1958

75. Ng J, Seip R, Stone A, Ruano G, Tishler D, Papasavas P. Ethnic variation in weight loss, but not co-morbidity remission, after laparoscopic gastric banding and Roux-en-Y gastric bypass. Surg Obes Relat Dis 2015;94–100

76. Leighton E, Sainsbury CA, Jones GC. A practical review of C-peptide testing in diabetes. Diabetes Ther 2017;8:475–487

77. Peyrot M, Rubin RR, Lauritzen T, et al.; International DAWN Advisory Panel. Resistance to insulin therapy among patients and providers: results of the cross-national Diabetes Attitudes, Wishes, and Needs (DAWN) study. Diabetes Care 2005;28:2673–2679