Title: 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] Srikanth Bellary*[2,6]

1Institute of Metabolism and Systems Research, University of Birmingham, UK

2 Department of Diabetes and Endocrinology, University Hospitals Birmingham NHS Foundation Trust, UK

3 Institute of Applied Health Research, University of Birmingham, UK

4 Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, U.K.

5 Midlands Health Data Research, U.K.

6 School of Life and Health Sciences, Aston University, UK

*KN and SB are joint senior authors and contributed equally to this manuscript.

Corresponding authors:

K Nirantharakumar

Institute of Applied Health Research, University of Birmingham, Birmingham, B15 2TT, UK, Phone number- +447737773559

Email- K.Nirantharan@bham.ac.uk

Abstract- 245

Word Count: 5917

Tables: 3

Figures:2

Abstract

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 was undertaken. The search was restricted to studies published in the last 15 years and in the English language.

Study selection and data extraction

All studies developing or validating a prediction model for diabetes remission in adults after bariatric surgery were included. The search identified 4165 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 curve (AUC) ≥ 0.800. Two of these prediction models, ABCD and DiaRem, were widely externally validated in different populations, 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 and Conclusions

Amongst 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.

Introduction

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 glycaemic control, including achieving type 2 diabetes remission (1-3), and reduction in the risk of micro- and macro-vascular 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 beta cell function such as younger age, shorter diabetes duration, high c-peptide, lack of insulin treatment, and lower pre-operative HbA1c (15) and lower pre-operative body mass index (BMI) have been identified as predictors of type 2 diabetes remission post-surgery.

Considering the importance of predicting diabetes remission for individualising 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).

Acknowledging 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 PRISMA guidelines (19). The protocol was registered on PROSPERO, registration number CRD42019124644.

Literature search and screening

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 term for prognostic/predictive models included prediction, prognosis, sensitivity, specificity, ROC (receiver operating characteristics) curve, AUC (area under the curve) with wild cards as well as other search terms as per published guidance (20).

The search was limited to papers 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 (ADA) in 2009. The final search was performed on 26th January 2019 and updated on 8th August 2020 (Figure 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 (PS and SB) 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 that developed or validated a prediction model to predict diabetes remission (Outcome, O). Multiple definitions of diabetes remission were used, but we included only those using definitions of HbA1c of ≤6.5% (48mmol/mol) and off glucose lowering medication, with follow-up of at least a year.

We excluded review articles, studies whose participants included children/adolescents or those with gastric cancer or gastric ulcer, studies where the intervention was other than bariatric/metabolic surgery, studies which had an outcome of diabetes remission defined as HbA1c >6.5% (48 mmol/mol), studies which had a follow-up period less than 12 months, and studies where the analysis was limited to identifying predictors.

Data extraction and analysis

The data extraction template was drafted on Distiller SR software (22). Data were collected by PS and independently collected by second reviewers (JH, NJA and SB). We adapted the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARM) 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 customised version of the Prediction model risk of Bias Assessment Tool (PROBAST) to assess risk of bias and applicability (24). The assessment was done under 4 domains for risk of bias – participants, predictors, outcomes and analysis – and 3 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 involves identifying predictors, selecting the important predictors by regression analysis/modelling, proposing a model by assigning relative weights to the individual predictors included, conducting internal validation, and validating in an external cohort to avoid overfitting (26). In this review, we classed the studies which developed and internally validated a prediction model as model development studies, and studies which externally validated 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.

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 receiver operating characteristic 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 categorisations used by Zhang et al, defining 0.501–0.699 as poor discrimination, 0.700–0.799 as acceptable discrimination, 0.800–0.899 as excellent discrimination and 0.900–1 as outstanding (29). For studies which did not report AUC, where possible we calculated discrimination using published tables providing information on the participants, their score, and their outcome in terms of remission and non-remission. We used Stata for the analysis, generating AUC graphs and values (with 95% confidence interval). 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 expected / observed (E/O) ratio (30). E/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 effect meta-analyses of studies based on i) their duration of follow-up, ii) the HbA1c cut-offs used to define remission, and iii) the type of bariatric surgery (Figure 2 a-f). We excluded studies from analysis where AUC was not known or could not be estimated with 95% confidence interval.

Our first meta-analysis was based on follow- up duration; studies were grouped into those with follow-up of 1 year and those with more than 1 year. In studies where diabetes remission was defined using two HbA1c cut-off values (e.g. 6.0% (42mmol/mol) and 6.5% (48mmol/mol)), we included the AUC for the higher cut-off only to avoid duplication of data sources.

The second meta-analysis was based on HbA1c cut-offs and studies were grouped into those with HbA1c cut-offs of 6.5% (48mmol/mol) and 6% (42mmol/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 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 HbA1c cut-off and longer follow-up duration where AUC was available for more than one HbA1c cut-off or length of follow-up.

Results

Following the initial search, we retrieved 5825 papers. After removing 1660 duplicates, 4165 publications were identified for title and abstract screening. 91 publications were identified as eligible for full text screening. 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 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 pre-defined prediction model in a different population or time period from the population/time period used to develop the model.

Model development studies

11 of the 16 model development studies produced scoring systems while the other five were logistic regression prediction models. The scoring systems were ABCD (Lee et al) from Taiwan in 2013 (16); Robert et al from France in 2013 (31); DiaRem (Still et al) from the USA in 2014 (17, 32); Diabetes Remission Score (DRS, Ugale et al) from India in 2014 (33); Individualised Metabolic Score (IMS, Aminian et al) from the USA in 2017 (18); Advanced DiaRem (Ad-DiaRem, Aron-Wisnewsky et al) from France in 2017 (34); DiaBetter (Pucci et al) from the UK in 2017 (35); DiaRem2 (Still et al) in 2018 (36), an updated version of the pre-existing DiaRem model developed by the same group; 5y-DR (Debedat et al) from France in 2018 (37); Metabolic Surgery Diabetes remission (MDR score, Mei Ching Moh et al) from Japan in 2020 (38); and Umemura et al 2020 from Singapore (39). The five logistic regression models were Hayes et al from New Zealand in 2011 (40); Dixon et al from Taiwan in 2013 (13); Ramos-Levi et al from Spain in 2014 (41); Cotillard et al from France in 2015 (42); and Stallard et al from Canada in 2016 (43) (Table 2).

Participants

Out of 16 studies, 14 studies used retrospective data and two 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 Umemura et al (39), which had higher male representation. The mean age of participants ranged between 36.5 years 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 Umemura et al (33). Mean diabetes duration ranged from 2.1 years in Lee et al (ABCD model) (16) to 9.9 years in Ugale et al (DRS model) (33). Pre-operative HbA1c ranged from 6.8% (51mmol/mol) in Still et al (DiaRem score) (17) to 9.1% (76mmol/mol) in the Dixon et al study (13).

Follow-up duration

The median follow-up range was 1-5 years. The majority of the studies reported remission rates at one 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 (Debedat et al) (37).

Outcome definition

Different definitions for diabetes remission were noted with some focusing on complete diabetes remission (defined as HbA1c <6.0% (42mmol/mol) and no anti-diabetic medication for at least 12 months) (13, 16, 31, 37) and others combining complete and partial diabetes remission (defined as <6.5% (48mmol/mol) and off medications for 12 months) (17, 18, 34, 38, 39). DiaRem2 (36) and Stallard et al (43) defined diabetes remission as a HbA1c of less than 5.7% (39mmol/mol) and ≤5.9% (41mmol/mol) off anti-diabetic medications at 12 months, respectively.

Method/analysis and presentation

Predictors in the models varied and included age, baseline BMI, C-peptide, diabetes duration, HbA1c, insulin use, glucose lowering medications, sex, micro- and macro-vascular complications. 5y-DR (37) included post-operative variables as predictors in the prediction model. The number of predictors ranged from 2 (40) to 10 (42). Five prediction models proposed a logistic regression model; the Dixon et al (13) and Hayes et al (40) models gave a logarithmic equation; 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. DiaRem (17) and DiaRem2 (36) 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. IMS (18) used a nomogram and benchmarks selected by an expert panel. Ad-Diarem (34) and 5yDR (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 cut-offs 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

To represent the model performance, 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) presented AUC. We calculated the AUC for ABCD (16), DiaRem (17) and Umemura et al (39) (Supplement materials 7). No AUC or performance was reported for DRS (33), IMS (18), Hayes et al (40) or Cotillard et al (42), and data in the publications were insufficient to calculate these.

Out of 12 prediction models for which AUC was available, ten prediction models [Dixon et al (13), Ramos-Levi et al (41) and 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 to 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 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 summarised in Table 3. Of the 22 external validation studies, nine validated the ABCD score, six validated DiaRem, one validated IMS and the remaining six studies compared two or more models.

Participants

19 studies used retrospective data (44-48, 51, 52, 54-65), and three collected the data prospectively (49, 50, 53). The sample size ranged from 53 (53) to 2190 (61), and mean age ranged from 35.7 (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 pre-surgery HbA1c ranged from 7.2 (55mmol/mol) (60) to 9.1% (76mmol/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

To define diabetes remission, the HbA1c cut-off was taken as 5.7% (39mmol/mol) in two studies (51, 65); 6.0% (42mmol/mol) in 13 studies (44-49, 52, 54, 55, 58, 62, 64); and the remaining studies defined diabetes remission as HbA1c ≤6.5% (48mmol/mol).

Performance

Although 16 models were identified in model development studies, few of these were externally validated in more than one external cohort, and those that are predominantly scoring systems. Direct comparison of the models was seen in only six studies (48, 55, 57, 59, 61, 62).

We have presented the assessment of validation studies based on the prediction models validated. As ABCD and DiaRem scores were validated most frequently, we have presented 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 Figure 2) and calibration (E/O) as 1.01 in the external cohort. In a subsequent study, ABCD score cut-off 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 cut-offs (44) has been externally validated in 13 studies (45, 46, 48, 49, 53-59, 62, 64). Out of these 13 validation studies, five looked at long-term diabetes remission at 3-5 years (45, 49, 59, 62, 64) and the remaining eight at 1 year. One study found poor discrimination (53) while others found the discrimination 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 the 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 (Figure 2a).

At the different HbA1c cut-offs, discrimination was excellent with an AUC of 0.81 (95% CI 0.79-0.83) for a HbA1c cut-off of 6.0% (42mmol/mol), and acceptable at 0.78 (0.74-0.81) for a HbA1c cut-off of 6.5% (48mmol/mol) (Figure 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) (Figure 2c).

DiaRem score

The DiaRem score has been externally validated in 11 studies (47, 48, 51, 52, 55, 57, 60-62, 65, 66) (Table 1 and Figure 2). Three studies looked at long term (>1 year) (61, 62, 66) diabetes remission and the remaining focussed on remission at 1 year.

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

Calibration was presented in two studies (55, 62). We were able to calculate the E/O ratio for a further six studies: DiaRem underestimated the probability of diabetes remission in the Ahuja et al (57) (E/O ratio 0.67) and the Mahaffey et al studies (50) (E/O 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/O ratios of 1.31, 1.71, 1.14, 1.25 in 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.

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

DiaRem meta-analysis

In meta-analysis, discrimination for DiaRem was AUC 0.78 (0.75-0.81) for short-term and 0.83 (0.80-0.86) for longer-term follow-up (Figure 2b).

At HbA1c cut-offs of 6.0% (42mmol/mol) and 6.5% (48mmol/mol), the AUCs were 0.77 (0.74-0.80) and 0.81 (0.78-0.84), respectively (Figure 2d).

For RYGB, meta-analysis showed acceptable discrimination for DiaRem with AUC of 0.78 (0.74-0.82). No meta-analysis was performed for SG as there was only one study identified validating DiaRem in a SG cohort (Figure 2f).

Performance of other prediction models

The discrimination scores for other prediction model are summarised in Table 1. The IMS score was externally validated by three validation studies (55, 59, 63) and one Umemura et al model development study (39). Discrimination was found to be excellent in Shen et al, acceptable in the RYGB cohort of Chen et al (59) and in 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). The Kam et al study (62) 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 if either information on missing data was not reported or was not included in the analysis, or 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-analysed studies for DiaRem were consistent with the findings of this study.

Discussion

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 a HbA1c cut-off of 6.5% (48mmol/mol) was higher than for ABCD. The AUC for ABCD for predicting short-term remission and diabetes remission defined by a HbA1c cut-off of 6.0% (42mmol/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, confidence intervals 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 similar patient cohorts 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 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 (67). 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 HbA1c used to define diabetes remission, with cut-offs ranging from 5.7 to 6.5% (39 to 48mmol/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 when assessing the benefits of bariatric surgery in patients with type 2 diabetes. In our review, we observed that 13 out 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, 68). Moreover, diabetes remission is highest during the first year following the intervention and declines over subsequent years and with longer follow-up (7-9). The prospective SOS study, with follow-up of over 18 years, showed the incidence of diabetes remission was 72.3% at 2 years, 38.1% at 10 years and 30.4% at 15 years (69). A randomised

controlled trial with 5 years' follow-up found 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 US with 5 years' follow-up (70). 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 healthcare systems. Beside the benefit of reduction in the incidence of micro- and macro-vascular 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 varies with the type of bariatric procedure (5, 71, 72). 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 sleeve gastrectomy (61.4%), Roux-en-Y gastric bypass (59.3%), gastric banding (29.6%), and then great curvature plication (18.6%) (72). Despite this, none of the prediction models included the type of surgery as a predictor. However, when we analysed 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 non-obese individuals. In a recent systematic review and meta-analysis Ji et al 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 (73) . They found a 1.58%(~16mmol/mol) reduction in HbA1c at 2 years using a random effects model (73). However, 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 (74). We found two studies -a model

development study by Ugale et al (33) and the validation study Lee et al (46) that focused 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 5 Asian clinics. We identified one validation study by Wood et al validating the DiaRem score in a White and Hispanic population that noted an AUC of 0.84 (0.80-0.88) and 0.79 (0.71-0.86) in White and Hispanic patients, respectively (60). A meta-analysis including 14 studies showed greater weight loss in Caucasians compared to African Americans, but no difference was noted in the outcome of diabetes remission between these two ethnic groups (75, 76). None of the prediction models identified ethnicity as a predictor and data on direct comparisons between ethnic groups was 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, 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 creatinine ratio or venous blood C-peptide levels measured as random, fasting, or stimulated state (glucagon stimulation test, mixed meal tolerance test) (77). These factors may pose difficulties in standardisation and can present as a limitation when 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 post-operative number of glucose-lowering medications, fasting capillary blood glucose, weight loss and 1-year remission to predict long-term diabetes remission.

Post-operative 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 non-insulin treatments and therapeutic inertia are recognized causes for delay in treatment with insulin (78). 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, it may underestimate the possibility of diabetes remission.

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 summarising 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. The majority of external validation studies used routinely collected data; consequently, follow-up data was 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.

Conclusions

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 three to four years showed promising results in test cohorts but there are 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.

Author contributions:

KN, PS and SB developed the original study question. PS, KN and SB had full access to all the studies included in this review and take responsibility for the integrity of the data and accuracy of the data analysis. PS and SB conducted the screening. Data collection and risk of bias assessment were performed by PS, SB, JH and NJA. PS and MP performed the analysis and this was interpretated by PS, SB, KN and ATT. PS wrote the first draft of the paper, which was revised and edited by NJA, SB, KN and ATT. All authors reviewed and approved the final draft of the manuscript.

Acknowledgement: Sue Bayliss reviewed the search strategy.

Duality of Interest: Duality of Interest: KN reports funding from AstraZeneca (RSBD20464) and fees from Sanofi and Boehringer Ingelheim outside the submitted work. AAT reports personal fees and non-financial support from Novo Nordisk, Eli Lilly, AstraZeneca, and Boehringer Ingleheim, personal fees from Janssen, and non-financial support from Impeto Medical, Resmed, and Aptiva outside the submitted work. AAT is a clinician scientist supported by the National Institute for Health Research (NIHR) in the UK (CS-2013-13-029). MP was supported by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. No other potential conflicts of interest relevant to this article were reported.

Funding: None.

References:

1. Mingrone G, Panunzi S, De Gaetano A, Guidone C, Iaconelli A, Leccesi L, et al. Bariatric surgery versus conventional medical therapy for type 2 diabetes. The New England journal of medicine. 2012;366(17):1577-85.
2. Schauer PR, Bhatt DL, Kirwan JP, Wolski K, Aminian A, Brethauer SA, et al. Bariatric Surgery versus Intensive Medical Therapy for Diabetes - 5-Year Outcomes. N Engl J Med. 2017;376(7):641-51.
3. MistryP, CurrieV, SuperP, CW lR, TahraniAA, SinghalR. 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. Clinical obesity. 2018;8(3).
4. Avenell A, Robertson C, Skea Z, Jacobsen E, Boyers D, Cooper D, et al. Bariatric surgery, lifestyle interventions and orlistat for severe obesity: the REBALANCE mixed-methods systematic review and economic evaluation. 2018.
5. SinghP, SubramaniaA, AdderleyN, GokhaleK, SinghalR, BellaryS, et al. Impact of Bariatric Surgery on Cardiovascular Outcomes and Mortality: A Population-Based Cohort Study. The British journal of surgery. 2020;107(4):432- 42.
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. 2021;44 (1):116-24.
7. Brethauer SA, Aminian A, Romero-Talamas H, Batayyah E, Mackey J, Kennedy L, et al. Can diabetes be surgically cured? Long-term metabolic effects of bariatric surgery in obese patients with type 2 diabetes mellitus. Annals of surgery. 2013;258(4):628-36; discussion 36.
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(6):902-11.
9. Mingrone G, Guidone C, Iaconelli A, Nanni G, Castagneto M, Panunzi S, 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. The Lancet. 2015;386(9997):964-73.
10. RubinoF, NathanDM, EckelRH, SchauerPR, AlbertiKG, ZimmetPZ, et al. Metabolic Surgery in the Treatment Algorithm for Type 2 Diabetes: A Joint Statement by International Diabetes Organizations. Diabetes care. 2016;39(6).
11. Mahmood F, Munasinghe A, MacAno C, Sharples A, Etumnu N, Katreddy V, et al. Factors predicting remission or improvement of type 2 diabetes after gastric bypass. 33rd American Society for Metabolic Surgery Annual Meeting, ASMBS 2016 United States. 2016;12(7 Supplement 1):S165-S6.
12. Panunzi S, Carlsson L, De Gaetano A, Peltonen M, Rice T, Sjostrom L, et al. Determinants of Diabetes Remission and Glycemic Control After Bariatric Surgery. Diabetes care. 2016;39(1):166-74LID.
13. Dixon JB, Chuang LM, Chong K, Chen SC, Lambert GW, Straznicky NE, et al. Predicting the glycemic response to gastric bypass surgery in patients with type 2 diabetes. Diabetes care. 2013;36(1):20-6LID.
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. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2012;8(5):548-55LID.
15. Park JY. Prediction of Diabetes Remission after Bariatric or Metabolic Surgery. 2018.
16. Lee W-J, Hur Kyung Y, Lakadawala M, Kasama K, Wong Simon KH, Chen S-C, et al. Predicting success of metabolic surgery: age, body mass index, C-peptide, and duration score. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2013;9(3):379-84.
17. Still CD, Wood GC, Benotti P, Petrick AT, Gabrielsen J, Strodel WE, et al. A probability score for preoperative prediction of type 2 diabetes remission following RYGB surgery. Lancet Diabetes Endocrinol. 2014;2(1):38-45.
18. Aminian A, Brethauer SA, Andalib A, Nowacki AS, Jimenez A, Corcelles R, et al. Individualized Metabolic Surgery Score: Procedure Selection Based on Diabetes Severity. Annals of surgery. 2017;266(4):650-7LID.

19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. 2009.
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(2):e32844.
21. EPPI-Reviewer 4: systematic review software 2017 [Available from: http://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4.
22. DistillerSR. DistillerSR 2020 [Available from: https://www.evidencepartners.com/products/distillersr-systematic-review-software/.
23. Karel G. M. Moons JAHdG, Walter Bouwmeester, Yvonne Vergouwe,. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. 2014 October.
24. Wolff RF, Kleijnen Systematic Reviews Y, United Kingdom (R.F.W., M.W.), Moons KGM, Julius Center for Health Sciences and Primary Care and Cochrane Netherlands UMCU, Utrecht University, Utrecht, the Netherlands (K.G.M., J.B.R.), Riley RD, Centre for Prognosis Research RIfPCaHS, Keele University, Keele, United Kingdom (R.D.R.), et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Annals of Internal Medicine. 2019;170(1):51-8.
25. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ (Clinical research ed). 2009;338:b606.
26. 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(9):683-90.
27. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010;21(1):128-38.
28. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010;5(9):1315-6.
29. Zhang R, Borisenko O, Telegina I, Hargreaves J, Ahmed AR, Sanchez Santos R, et al. Systematic review of risk prediction models for diabetes after bariatric surgery. The British journal of surgery. 2016;103(11):1420-7.
30. Hanson RK. Assessing the Calibration of Actuarial Risk Scales: A Primer on the E/O Index. http://dxdoiorg/101177/0093854816683956. 2016.
31. Robert M, Ferrand-Gaillard C, Disse E, Espalieu P, Simon C, Laville M, et al. Predictive factors of type 2 diabetes remission 1 year after bariatric surgery: impact of surgical techniques. Obes Surg. 2013;23(6):770-5.
32. Still Christopher D, Wood GC, Benotti P, Petrick Anthony T, Gabrielsen J, Strodel William E, et al. Preoperative prediction of type 2 diabetes remission after Roux-en-Y gastric bypass surgery: a retrospective cohort study. The lancet Diabetes & endocrinology. 2014;2(1):38-45.
33. 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. Journal of Diabetes & Metabolic Disorders. 2014;13(1):89.
34. Aron-Wisnewsky J, Sokolovska N, Liu Y, Comaneshter DS, Vinker S, Pecht T, et al. The advanced-DiaRem score improves prediction of diabetes remission 1 year post-Roux-en-Y gastric bypass. Diabetologia. 2017;60(10):1892-902.
35. Pucci A, Tymoszuk U, Cheung WH, Makaronidis JM, Scholes S, Tharakan G, et al. Type 2 diabetes remission 2 years post Roux-en-Y gastric bypass and sleeve gastrectomy: The role of the weight loss and comparison of DiaRem and DiaBetter scores. Diabetic Medicine. 2017.
36. StillCD, BenottiP, T M, CookA, WoodGC. DiaRem2: Incorporating Duration of Diabetes to Improve Prediction of Diabetes Remission After Metabolic Surgery. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2019;15(5).

37. Debedat J, Sokolovska N, Coupaye M, Panunzi S, Chakaroun R, Genser L, et al. Long-term Relapse of Type 2 Diabetes After Roux-en-Y Gastric Bypass: Prediction and Clinical Relevance. Diabetes Care. 2018;41(10):2086-95.
38. Moh Mei C, Cheng A, Tan Chun H, Lim Boon K, Tan Bo C, Ng D, et al. Metabolic Surgery Diabetes Remission (MDR) Score: a New Preoperative Scoring System for Predicting Type 2 Diabetes Remission at 1 Year After Metabolic Surgery in the Singapore Multi-ethnic Asian Setting. Obesity surgery. 2020;30(9):3387-93.
39. Umemura A, Sasaki A, Nitta H, Nikai H, Baba S, Takahara T, et al. Prognostic factors and a new preliminary scoring system for remission of type 2 diabetes mellitus after laparoscopic sleeve gastrectomy. Surgery Today. 2020.
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(7):910-6.
41. Ramos-Levi AM, Matia P, Cabrerizo L, Barabash A, Sanchez-Pernaute A, Calle-Pascual AL, et al. Statistical models to predict type 2 diabetes remission after bariatric surgery. Journal of diabetes. 2014;6(5):472-7LID.
42. Cotillard A, Poitou C, Duchateau-Nguyen G, Aron-Wisnewsky J, Bouillot JL, Schindler T, et al. Type 2 Diabetes Remission After Gastric Bypass: What Is the Best Prediction Tool for Clinicians? Obesity surgery. 2015;25(7):1128-32LID.
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. 2016;42(3):573-80.
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. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2015;11(5):991-6.
45. Lee MH, Lee W-J, Chong K, Chen J-C, Ser K-H, Lee Y-C, et al. Predictors of long-term diabetes remission after metabolic surgery. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2015;19(6):1015-21.
46. Lee W-J, Almulaifi A, Chong K, Chen S-C, Tsou JJ, Ser K-H, 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). Obesity surgery. 2015;25(10):1772-8.
47. Sampaio-Neto J, Nassif LS, Branco-Filho AJ, Bolfarini LA, Loro LS, de SM, et al. EXTERNAL 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, Zachariah J, Ser K-H, Lee Y-C, et al. Preoperative Prediction of Type 2 Diabetes Remission After Gastric Bypass Surgery: a Comparison of DiaRem Scores and ABCD Scores. Obesity surgery. 2016;26(10):2418-24.
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 journal of surgery. 2017;41(1):216-23.
50. Mehaffey JH, Mullen MG, Mehaffey RL, Turrentine FE, Malin SK, Kirby JL, et al. Type 2 diabetes remission following gastric bypass: does diarem stand the test of time? Surgical endoscopy. 2017;31(2):538-42LID.
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. Clinical obesity. 2017;7(3):176-82LID.
52. Tharakan G, Scott R, Szepietowski O, Miras AD, Blakemore AI, Purkayastha S, 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. Obesity surgery. 2017;27(3):782-6LID.
53. Praveen Raj P, Bhattacharya S, Saravana Kumar S, Sabnis SC, Parthasarathi R, Swamy PDK, et al. Do Bariatric Surgery-Related Type 2 Diabetes Remission Predictors Add Clinical Value? A Study on Asian Indian Obese Diabetics. Obesity surgery. 2017;27(8):2113-9LID.

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. Obesity surgery. 2018;28(8):2429-38.
55. Shen SC, Wang W, Tam KW, Chen HA, Lin YK, Wang SY, et al. Validating Risk Prediction Models of Diabetes Remission After Sleeve Gastrectomy. Obes Surg. 2018.
56. Naitoh T, Kasama K, Seki Y, Ohta M, Oshiro T, Sasaki A, et al. Efficacy of Sleeve Gastrectomy with Duodenal-Jejunal Bypass for the Treatment of Obese Severe Diabetes Patients in Japan: a Retrospective Multicenter Study. Obesity Surgery. 2018;28(2):497-505.
57. Ahuja A, Tantia O, Chaudhuri T, Khanna S, Seetharamaiah S, Majumdar K, et al. Predicting remission of diabetes post metabolic surgery: a comparison of ABCD, diarem, and DRS scores. Obesity Surgery. 2018:1-7.
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. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2018;14(4):509-15.
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. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2018;14(5):640-5.
60. Craig Wood G, Horwitz D, Still CD, Mirshahi T, Benotti P, Parikh M, 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. Obesity Surgery. 2018;28(1):61-8.
61. Dicker D, Golan R, Aron-Wisnewsky J, Zucker JD, Sokolowska N, Comaneshter DS, et al. Prediction of Long-Term Diabetes Remission After RYGB, Sleeve Gastrectomy, and Adjustable Gastric Banding Using DiaRem and Advanced-DiaRem Scores. Obes Surg. 2019;29(3):796-804.
62. Kam H, Tu Y, Pan J, Han J, Zhang P, Bao Y, et al. Comparison of Four Risk Prediction Models for Diabetes Remission after Roux-en-Y Gastric Bypass Surgery in Obese Chinese Patients with Type 2 Diabetes Mellitus. Obesity surgery. 2020;30(6):2147-57.
63. Park DG CY, 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;9:13-8.
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. Obesity Surgery. 2020.
65. Guerron AD, Perez JE, Risoli T, Lee HJ, Portenier D, Corsino L. Performance and improvement of the DiaRem score in diabetes remission prediction: a study with diverse procedure types. Surgery for Obesity and Related Diseases. 2020.
66. Mehaffey JH, Mullen MG, Mehaffey RL, Turrentine FE, Malin SK, Schirmer B, et al. Diabetes remission following gastric bypass: Does diarem stand the test of time? 2016 Scientific Session of the Society of American Gastrointestinal and Endoscopic Surgeons, SAGES 2016 Boston, MA United States. 2016;30(SUPPL. 1):S461.
67. FinerS, RobbP, CowanK, DalyA, RobertsonE, FarmerA. Top ten research priorities for type 2 diabetes: results from the Diabetes UK-James Lind Alliance Priority Setting Partnership. The lancet Diabetes & endocrinology. 2017;5(12).
68. JansA, NäslundI, OttossonJ, SzaboE, NäslundE, StenbergE. Duration of type 2 diabetes and remission rates after bariatric surgery in Sweden 2007-2015: A registry-based cohort study. PLoS medicine. 2019;16(11).
69. Sjostrom L, Peltonen M, Jacobson P, Ahlin S, Andersson-Assarsson J, Anveden A, et al. Association of bariatric surgery with long-term remission of type 2 diabetes and with microvascular and macrovascular complications. JAMA. 2014;311(22):2297-304LID.

70. Arterburn DE, Bogart A, Sherwood NE, Sidney S, Coleman KJ, Haneuse S, et al. A multisite study of long-term remission and relapse of type 2 diabetes mellitus following gastric bypass. Obesity surgery. 2013;23(1):93-102LID.
71. Dogan K, Gadiot RP, Aarts EO, Betzel B, van Laarhoven CJ, Biter LU, 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(7):1110-8.
72. DingL, FanY, LiH, ZhangY, QiD, TangS, et al. Comparative Effectiveness of Bariatric Surgeries in Patients With Obesity and Type 2 Diabetes Mellitus: A Network Meta-Analysis of Randomized Controlled Trials. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2020.
73. JiG, LiP, LiW, SunX, YuZ, LiR, et al. The Effect of Bariatric Surgery on Asian Patients with Type 2 Diabetes Mellitus and Body Mass Index 2: a Systematic Review and Meta-analysis. Obesity surgery. 2019;29(8).
74. Segal-Lieberman G, Segal P, Dicker D. Revisiting the Role of BMI in the Guidelines for Bariatric Surgery. 2016.
75. WM A, F C, VE G, RM D, JB H, F H. Ethnic differences in weight loss and diabetes remission after bariatric surgery: a meta-analysis. Diabetes care. 2012;35(9).
76. 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. Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery. 2015;11(1):94-100LID.
77. Leighton E, Sainsbury CA, Jones GC. A Practical Review of C-Peptide Testing in Diabetes. Diabetes Ther. 82017. p. 475-87.
78. PeyrotM, RubinRR, LauritzenT, SkovlundSE, SnoekFJ, MatthewsDR, et al. Resistance to Insulin Therapy Among Patients and Providers: Results of the Cross-National Diabetes Attitudes, Wishes, and Needs (DAWN) Study. Diabetes care. 2005;28(11).

Figures and tables in manuscript

Figure 1: PRISMA Flowchart

Figure 2 a-f: Meta-analysis

Table 1: Model development studies with their predictors

Table 2: Study characteristics of model development studies

Table 3: Study characteristics of validation studies

Table 1: Model development studies with their predictors

Prediction Model	Predictors included	Discrimination in study development	Discrimination in external validation study
ABCD; Lee et al 2013 (16)	Age, BMI, C-peptide, and diabetes duration	0.792 (0.728-0.856)*	Figure 2 a, c,and e
DiaRem; Still et al 2014 (17)	Age, HbA1c, diabetes medication other than metformin, and insulin use	0.840 (0.795-0.886)*	Figure 2 b, d and f
Robert et.al 2013 (31)	BMI, diabetes duration, HbA1c, fasting glucose, and diabetes medication	0.950 (0.838-0.992)	Shen et al- 0.681 ± 0.056
DRS; Ugale S 2014 (33)	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 2017 (34)	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 0.85 (0.76-0.93) Kam et al@1 year 0.752 (0.688-0.808) Kam et al @3 years 0.794 (0.715-0.860) 5y- DR 84%
DiaBetter; Pucci et.al 2017 (35)	HbA1c, diabetes duration, and kind of diabetes medication	0.867 (0.817-0.916)	Shen et al 0.826 ± 0.041 Kam et al @1 year 0.760 (0.697-0.815) Kam et al @3 years 0.804 (0.726-0.868)

IMS; Aminian et.al 2017 (18)	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 0.516 (0.330-0.702)*
DiaRem2; Still et al 2018 (36)	Age, HbA1c, diabetes medication other than metformin, and insulin use, diabetes duration	0.876	NA
5y- DR, 2018 (37)	Pre-op factors- diabetes dur, no of medication, HbA1c	90%	NA
	Post-op factors- No of medication, fasting CBG, weight loss, 1 year remission		
Metabolic surgery diabetes remission (MDR), 2020 (38)	Age, HOMA2-B, diabetes duration and HbA1c	0.79 (0.71- 0.88)	NA
Umemura et al 2020 (39)	Insulin, diabetes duration	0.865 (0.775-0.954)*	NA
Hayes et.al 2011 (40)	Insulin use and HbA1c	NA	0.632 ± 0.059
Dixon el.al 2013 (13)	BMI and diabetes duration, c-peptide	0.90 (0.84-0.95)	0.800 ± 0.047
Ramos-Levi et al 2014 (41)	Model1: age, sex, FG, diabetes duration, insulin	0.838 (0.725-0.951)	
	Model2: age, sex, FG, diabetes duration, insulin, c-peptide	0.923 (0.852-0.996)	0.811 ± 0.047
	Model3: age, sex, FG, diabetes duration, insulin, % wt loss	0.923 (0.851-0.996)	
	Model4: age, sex, FG, diabetes duration, insulin, % wt loss, c-peptide	0.981 (0.951-1.000)	
Cotillard et al 2015 (42)	Age, sex, BMI, fasting glycemia, HbA1c, hypertension, diabetes duration, insulin therapy, number of anti-diabetic drugs and C-peptide	NA	NA

Stallard et al 2016 (43)	diabetes duration, FPG, use of non-insulin anti-diabetic medications and use of insulin.	0.860 (0.763-0.957)	NA

* = calculated by authors of this systematic review. BMI = body mass index, HbA1c = glycosylated haemoglobin, HOMA2-B = Homeostasis Model Assessment 2-Beta cell, FG = fasting glucose.

Table 2: Study characteristics of model development studies

Publication Reference	Source of data	Participant characteristics					Outcomes	Types of Surgery	Presentation	Validation	
			Age	BMI	Diabetes Duration	HbA1c				V Dev	Ext V
ABCD; Lee et al 2013 (16)	Retrospective Taiwan Multi-centre 2005- 2010	N=63 17 M:56 F R NR	36.5± 10.7 44.5 ± 7.7	40.9 ± 8.9 33.3 ±7.4	2.1 ± 3.7 4.1 ± 4.5	8.2± 1.8 8.5 ± 1.8	n=48 (76%) FU=1 year	RYGB	Scoring system	Y	Y
DiaRem; Still et al 2014 (17)	Retrospective USA Multi-centre 01/01/04 – 02/11	N=690 184 M:506 F NI (n=438) I (n=252)	48.8± 10.3 53.6 ± 8.9	49.5± 8.0 49.2 ± 8.8	6.8± 1.2 8.2 ± 1.7	NA NA	n=463 (67%) FU=14 months	RYGB	Scoring system	Y	Y
Robert et.al 2013 (31)	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 S 2014 (33)	Retrospective India Single	N=75 49 M:26 F IISG	51.7 ± 13.3	23.4 ± 4.5	9.9 ± 4.8	8.1 ± 0.59	n=42 (56%) FU=1-2.5 years	SG	Scoring system	N	Y

	01/02/08 – 03/10	IIDSG	57.6 ± 11.5	25.6 ± 4.5	10.1 ± 5	9 ± 0.78					
Ad-DiaRem; Aron-Wisnewsky et.al 2017 (34)	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 2017 (35)	Retrospective UK Single 01/01/08 – 12/15	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 2017 (18)	Retrospective USA 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 2018 (36)	Retrospective USA 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
5y- DR, 2018 (37)	Retrospective	N=175	48.3 ± 10.3	47.37± 7.43	6.75 ± 6.53	7.5 ± 1.6	66 (37.7) @1	RYGB	Scoring system	Y	N

	France	F=136 (77.71%)					94 (53.7) @ 5 FU=5.1 ± 0.7year				
Metabolic Surgery Diabetes remission (MDR); Moh 2020 (38)	Retrospective Singapore 2007-2018	N=114	A= 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 2020 (39)	Retrospective Single 2008- 2018 Japan	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 2011 (40)	New Zealand Single 01/11/97 – 05/07	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 2013 (13)	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 2014 (41)	Retrospective Spain Single	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

	2006- 2011										
Cotillard et al 2015 (42)	France Single	N=84 15 M:45 F	46.96 ±9.14	46.96 ± 9.14	3.86±4.64	7.01 ± 1.03	n=50 (59.5%) FU=1 year	RYGB	Logistic regression	N	N
Stallard et al 2016 (43)	Retrospective Canada Single 01/01/11 – 06/14	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 out of 77 (67.5%) FU=1 year	RYGB SG	Logistic regression	N	N

N = Total number of participants; n = Number of participants achieving diabetes remission; M = Male; F = Female; R = Remitters; NR = Non-remitters; NI = Non-Insulin; I = Insulin; FU = Follow-up; V Dev = validated in internal/external cohort while model development stage; Ext V = external validation; Y= Yes; RYGB = gastric bypass; SG = sleeve gastrectomy; GB = gastric band.

Table 3: Study characteristics of validation studies

Publication Reference	Source of data	Participant characteristics					Outcomes /Events	Types of Surgery	Study validated
			Age	BMI	Diabetes Duration	HbA1c			
Lee et al 2015 modified (44)	Retrospective Single Taiwan 2006-2013	N=85 F:M	41.9±10.9	39.0±7.4	2.7±3.1	8.1 ±1.7	n(CR+PR) =63/85 (74.1%) FU=1 year	SG	ABCD
Lee et al 2015 (45)	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%) @1 year n=97 (71.3%) @5 years FU=6 (5-8) years	RYGB SG	ABCD
Lee et al 2015 (46)	Single Taiwan 2007-2013	N=80 out of 512 had BMI<30 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 2015 (47)	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 2016 (48)	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 2017 (49)	Prospective Single Taiwan 2007-2014	N=579 (230 M:349 F) 48 M:61 F (SG) 182 M:288 F (RYGB)					n=361 (62.3%)/579 @1 year n=71 (49.7%)/143 @5 year	RYGB SG	ABCD
		SG (N=109)	43.2±11.0	35.7±7.2	3.3±3.5	8.8±1.5			

		RYGB (N=470)	41.8±10.9	36.9±7.2	4.5±4.8	8.6±1.7			
Mehaffey et al 2017 (50)	Prospective USA 2004-2013	N=57 75%F (2 years FU) N=31 55% F (10 years FU)					n@2 years =37(65%) n@10 years=18(58%)	RYGB	DiaRem
		2 year FU	49.2	52.1 ±1.5	NA	8.3			
		10 year FU	45.8	48.0 ±6.6	NA	7.7			
Honarmand et al 2017 (51)	Retrospective MC USA 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
Tharakan et al 2017 (52)	Retrospective Single UK 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 2017 (53)	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 2018 (54)	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)=49% FU=1 year	SG	ABCD
Shen et al 2018 (55)	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 AdDiaRem DiaBetter
Naitoh et al 2018 (56)	Retrospective	N=298					n(CR+PR)=247(82.9%)	LSG	ABCD

	MC Japan 2005-2015	140M:158 F LSG (N=177) LSG/DJB (N=131)	 45.2 45.2	 45.2 43.5	 5.9±6.1 7.7±1.7	 7.3±6.0 8.3±1.7	FU=1 year	LSG/DJB	
Ahuja et al 2018 (57)	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 2018 (58)	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 2018 (59)	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 2018 (60)	Retrospective Single USA 2002-2014	N=520 124 M:396 F	46.7 to 53	NA	NA	7.2 to 7.5	n=249 n=173 (RYGB) n=63 (SG) n=13 (GB)	RYGB SG GB	DiaRem
Dicker et al 2019 (61)	Retrospective Israel 1999- 2011	N=2190 64.8% F	47.1±10.9	43.5±6.3	NA	7.7±1.6	n=897 (59.7%)/1502 @ 2 years n=782 (53.6%)/1459 @ 5 years	RYGB SG GB	AdDiaRem DiaRem
Kam et al 2020 (62)	Retrospective China	N=214@1 year 131@ 3 year 117 (54.7) F	48 (37-57)	30.6 (28.7-32.9)	6.0 (3-10)	8.0 (7.1-9.8)	113 of 214 (52.8%) @1 59 of 131 (45.0%) @ 3	RYGB	ABCD, DiaRem, AdDiaRem, DiaBetter
Park 2020 (63)	Retrospective Korea	N=135 103 (76%) F	40 ± 11	39.0 ± 6.3	1 (0-5) year	7.5 (6.8- 8.9)	n= 88 (65.2%) FU= 1 year	RYGB SG	IMS

Lee et al 2020 (64)	Retrospective 2006- 2014	N=59 F	47.7± 12.4	37.6 ± 5.1	2.7 ±3.1 years	8.3± 2.2%	37 (62.7%) @ 1 year 24 (42.4%) @ 5 year	SG	ABCD
Guerron et al 2020 (65)	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

N = total number of participants; n = number of participants achieving diabetes remission; Single = single centre; M = male; F = female; MC = multi-centre; FU = follow-up; RYGB = gastric bypass; SG = sleeve gastrectomy; GB = gastric band; LSG = laparoscopic sleeve gastrectomy; LSG/DJB = LSG with duodenal-jejunal bypass.