

Published in final edited form as:

*Diabetes Res Clin Pract.* 2021 January 01; 171: 108625. doi:10.1016/j.diabres.2020.108625.

## The absolute and relative risk of type 2 diabetes after gestational diabetes: A systematic review and meta-analysis of 129 studies

Rebecca A Dennison<sup>a</sup>, Eileen S Chen<sup>b</sup>, Madeline E Green<sup>b</sup>, Chloe Legard<sup>b</sup>, Deeya Kotecha<sup>b</sup>, George Farmer<sup>b</sup>, Stephen J Sharp<sup>c</sup>, Rebecca J Ward<sup>a</sup>, Juliet A Usher-Smith<sup>a</sup>, Simon J Griffin<sup>a,c</sup>

Deeya Kotecha: DK537@cam.ac.uk; George Farmer: GF312@cam.ac.uk; Stephen J Sharp: Stephen.Sharp@mrc-epid.cam.ac.uk; Rebecca J Ward: rebeccajane.ward@net.net; Juliet A Usher-Smith: JAU20@medschl.cam.ac.uk; Simon J Griffin: ProfGP@medschl.cam.ac.uk

<sup>a</sup>The Primary Care Unit, Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK

<sup>b</sup>School of Clinical Medicine, University of Cambridge, UK

<sup>c</sup>MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge, Cambridge, UK

### Structured abstract

**Aims**—To estimate development of type 2 diabetes (T2DM) in women with previous gestational diabetes (GDM) and investigate characteristics associated with higher diagnoses, building on previous meta-analyses and exploring heterogeneity.

**Methods**—Systematic literature review of studies published up to October 2019. We included studies reporting progression to T2DM 6 months after pregnancy, if diagnostic methods were reported and 50 women with GDM participated. We conducted random-effects meta-analyses and meta-regression of absolute and relative T2DM risk. PROSPERO ID: CRD42017080299.

**Results**—In 129 included studies, the percentage diagnosed with T2DM was 12% (95% confidence interval 8–16%) higher for each additional year after pregnancy, with a third developing diabetes within 15 years. Development was 18% (5–34%) higher per unit BMI at follow-up, and 57% (39–70%) lower in White European populations compared to others (adjusted for ethnicity and follow-up). Women with GDM had a relative risk of T2DM of 8.3 (6.5–10.6). 17.0% (15.1–19.0%) developed T2DM overall, although heterogeneity between studies was substantial ( $I^2$  99.3%), and remained high after accounting for various study-level characteristics.

---

Correspondence to: Rebecca A Dennison; Simon J Griffin, ProfGP@medschl.cam.ac.uk.

Corresponding author: Rebecca Dennison, Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Institute of Public Health, Cambridge, UK, CB2 0SR, RL423@medschl.cam.ac.uk, Phone: 01223 330307.

#### Contributors

RW, RD, JUS and SG contributed to the study design and protocol development. RW conducted the literature search. RW and RD completed the first stage of title and abstract screening; EC, RD, GF, MG, DK and CL completed the second stage of title and abstract screening, full text screening, data extraction and quality assessment with input from SG. RD completed statistical analysis, data interpretation and first drafted this report with input from the other authors, particularly SS. All authors contributed to the writing and revision of this report, and have seen and approved the final version.

#### Competing interests

The authors declare no competing interests.

**Conclusions**—Percentage developing T2DM after GDM is highly variable. These findings highlight the need for sustained follow-up after GDM through screening, and interventions to reduce modifiable risk factors.

## Keywords

Gestational diabetes; incidence; meta-analysis; systematic review; type 2 diabetes

---

## 1 Introduction

Gestational diabetes mellitus (GDM) is one of the most common conditions of pregnancy. Prevalence estimates vary according to which diagnostic criteria are used, with up to 13% of pregnancies affected across the world based on local criteria[1]. This equated to an estimated 17.8 million live births affected by GDM in 2015[2]. The incidence continues to rise as a consequence of increasing levels of obesity, sedentary lifestyles and unhealthy diets[3].

GDM increases the risk of adverse pregnancy outcomes for both mother and baby. Furthermore, there is an association with higher long-term risk of obesity, glucose intolerance and development of type 2 diabetes mellitus (T2DM)[4]. The prevalence of T2DM is also increasing, and, along with its complications, is a major contributor to morbidity and premature mortality with an estimated worldwide cost of US\$1.3 trillion in 2015[5]. A recent systematic review of 20 studies found women with GDM to be nearly ten-times more likely to develop T2DM than unaffected women, with a cumulative incidence of up to 16.5%[6]. Factors such as elevated body mass index (BMI), multiparity and poorer pregnancy glucose tolerance have been suggested to further increase T2DM risk[7,8], while breastfeeding may have a protective effect[9].

Since earlier diagnosis of glucose abnormalities enables timely management, most national and international guidelines recommend that women who have been diagnosed with GDM are screened for T2DM shortly after pregnancy and subsequently at regular intervals[10,11]. However, uptake of postpartum screening is suboptimal, even for the first test[12,13]. Women may not attend testing due to its inconvenience or unpleasantness, but also may not have had attendance promoted appropriately by their clinicians during and after pregnancy[14]. Similarly, they may not appreciate or prioritise the need to make changes to their lifestyle to reduce diabetes risk and feel unsupported to achieve sustained behaviour change[15].

GDM management and diagnostic strategies have changed significantly over time, and prevalence of risk factors has continued to increase. Consequently, up-to-date estimates of T2DM risk are required by clinicians, patients and policy makers to inform postpartum care after GDM. Since starting our review, Vounzoulaki *et al.* have compared progression rates to T2DM in women with GDM and healthy controls[6]. In a larger, more comprehensive systematic review, we report the absolute and relative risk of T2DM among women with prior GDM from all available data from observational and experimental studies published up to October 2019. Generally, relative estimates are more frequently reported, but absolute risk (such as *n* in 1000 people) tends to be more easily understood[16,17]. We also describe the

heterogeneity of estimates of progression, and explore study-level characteristics associated with heterogeneity.

## 2 Subjects, Materials and Methods

### 2.1 Search strategy and selection criteria

The protocol for this review was registered on PROSPERO ([www.crd.york.ac.uk/prospéro](http://www.crd.york.ac.uk/prospéro)); study ID: CRD42017080299.

We searched MEDLINE, Embase, PsychINFO, CINAHL and the Cochrane Library electronic databases from inception up to 14 October 2019 as part of a group of literature reviews concerning GDM using the search strategy shown in Supplementary Table 1. No language or other restrictions were applied. Reference lists of previous reviews related to this topic were also screened for citations that were not identified by the literature search.

We included papers published in peer-reviewed journals that quantified the percentage developing T2DM in those with a history of GDM. The diagnostic method or criteria for both GDM and T2DM must have been specified. Only studies following 50 participants with GDM and 6 months postpartum follow-up were included in order to reduce the number of studies that would have little impact on the interpretation of incidence estimates and avoid over-estimating development of T2DM after GDM by including a significant proportion of women who were likely to have had pre-existing, undiagnosed T2DM. All study designs were eligible.

Titles and abstracts identified during the literature search were screened for general topic relevance after removal of duplicate citations. We then re-assessed these citations against this study's selection criteria. Secondly, full text articles were acquired and assessed for inclusion; reasons for exclusion were recorded. All authors involved reviewed 10% of the papers independently at each stage to assess agreement and clarify the selection criteria; any discrepancies in opinion were reviewed by at least two of the remaining co-authors. If an author was unsure whether to include an article, they consulted the other authors (we did not formally assess inter-rater agreement over selection of studies).

### 2.2 Data gathering

A data extraction form was developed to facilitate systematic extraction of summary study, incidence and demographic information from included citations (including control groups without GDM, if reported). Data were extracted by two authors independently and any initial differences were resolved by discussion in order to minimise error.

After extracting basic details of each study, we sought to identify whether the same study population had been reported by multiple publications (primarily comparing location and recruitment dates). When overlap occurred, we only included the publication with the most person-years of follow-up of women with GDM. Similarly, if progression to T2DM was reported at multiple timepoints within one citation, we extracted the data with the most person-years of follow-up.

Where possible, non-English language articles were translated using an online translation tool followed by verification of the data extracted and justification of quality assessment by a native or fluent speaker of that language. If the online translation was unclear, the native speaker completed the full text review and data extraction using a simple form to guide them through this process.

We then evaluated the risk of bias in each included study using a checklist adapted from the Critical Appraisal Skills Programme and Newcastle-Ottawa Scale checklists (Supplementary Table 2)[18,19]. We compared the checklists and selected six questions that were common to both for assessing possible bias in the incidence estimate across all study designs, and used a simple scoring system to maintain comparability and internal validity with numerous studies. Studies scored one point for “yes” and zero points for “unclear” or “no”; scores of five or six were considered as high quality studies, and three or four were medium quality. Quality assessment was independently completed by at least two authors for each study.

If studies were part of a postpartum diabetes prevention intervention, only the placebo/control arm was analysed unless the intervention had no effect on the percentage developing T2DM (in accordance with study authors’ interpretations). Study locations were grouped based on geographical regions. They were classified as pregnancy before or after the year 2000 based on the median year that eligible pregnancies took place (date of GDM diagnosis or delivery; 2001 was the median year across the studies). Mean, median or planned duration of follow-up from delivery to assessment of T2DM was categorised into five groups (<3.0, 3.0–5.9, 6.0–8.9, 9.0–11.9, or ≥12.0 years); if only the range or upper limit was reported, we categorised the study based on the estimated mid-point. We determined whether glycaemic tests performed within the study, medical record reviews, or self-report had been used to identify GDM or classify T2DM. Additionally, diagnostic criteria were classified as high or low sensitivity, or a clinical diagnosis (low sensitivity defined as fasting plasma glucose ≥5.8 mmol/l for GDM and ≥7.8 mmol/l for T2DM based on author consensus; high sensitivity defined as fasting plasma glucose <5.8 mmol/l for GDM and <7.8 mmol/l for T2DM). If the criteria changed during the study, the one used for the greatest proportion of follow-up informed diagnostic sensitivity.

Study-level demographic characteristics were used as binary variables according to average (mean or median) and/or clinically relevant cut-offs, as reported below. Average age at delivery and follow-up were estimated if not reported (for example, age at delivery was calculated by adding 0.25 years to age at GDM diagnosis at 27 weeks gestation). If participants’ ethnicity was reported, this was used to generate the binary variable; if ethnicity was not reported, we inferred whether the majority of the population were White European based on the study setting.

### 2.3 Statistical analysis

The analysis was performed using STATA 15.1, using ‘metan’, ‘metareg’ and ‘forestplot’ commands. We grouped studies according to each of the characteristics described above, and combined these groups of studies using random-effects meta-analysis of the log odds of T2DM (in women with GDM), with estimates back-transformed to the percentage scale. We used meta-regression to model the association between study-level characteristics and log

odds of T2DM. We then extended the model to investigate the extent to which any of the study-level study and maternal characteristics described above explained the heterogeneity between studies, adjusting all models for ethnicity (majority White European or other) and categorised follow-up duration; that is, to investigate the extent to which any of the characteristics explained the heterogeneity between studies independent of follow-up and ethnicity. Studies were weighted by the inverse of the sum of the within and between study variance. We also calculated the relative risk of T2DM in studies that had a comparator population and combined these across studies using random-effects meta-analysis. Unadjusted relative risk was calculated from raw data, overall and stratified by study and maternal characteristics. We used a fixed continuity correction of 0.5 where no cases of T2DM were reported. Heterogeneity between studies was quantified using the  $I^2$  statistic throughout. A post hoc sensitivity analysis (proposed by a reviewer) was performed according to study quality.

### 3 Results

Our literature search identified 25,789 studies after removal of duplicates. We reviewed 518 full texts and included 129 citations from the literature search and reference lists (Figure 1). The percentage developing T2DM was reported in 310,214 women with a history of GDM, plus 4,155,247 parous women without GDM. Supplementary Tables 3 and 4 report details of the included studies.

Sixty one studies (47%) were based in Europe and 61 studies (47%) followed up 200 participants with GDM. The date of pregnancy ranged from 1979 to 2018, and 45 studies (35%) included a non-GDM comparator group. The average duration of follow-up was 5.7 years since pregnancy (range 0.6–29.9 years). Most studies fell into the <3.0 years or 3.0–5.9 years follow-up categories (n=38 [29%] and n=44 [34%], respectively).

Most cases of GDM were identified by oral glucose tolerance tests (OGTTs) performed during the study (n=36 [28%]) or recorded in medical records (n=86 [67%]). However, many different diagnostic criteria were used (varying by timing, dose of glucose administered, and glycaemic cut-offs); Carpenter and Coustan[20], Australasian Diabetes in Pregnancy Society[21], American Diabetes Association[22–25], and WHO[26–29] criteria were frequently reported alongside local protocols. According to the definitions described above, 34 studies (26%) used low sensitivity and 63 (49%) used high sensitivity tests, and the remainder were grouped as clinical diagnoses. To assess T2DM status, glycaemic tests such as 75g OGTT or HbA<sub>1c</sub> with multiple different criteria were used frequently (n=79 [61%]); high sensitivity tests were most common (n=87 [67%]). Clinical diagnoses included review of medical records, diabetes registers, or reimbursement for diabetes medication records.

Ninety one (71%) studies were medium quality (median score 4/6). The most common reasons for reduced quality were high loss to follow-up (n=63 [49%]) and no assessment of pre-existing T2DM (n=61 [47%]; Supplementary Figure 1).

Among the study populations with GDM, average age was 31.8 years at delivery (range 18.7–38.5 years; data available in n=103 studies) and 37.7 years at follow-up (30.2–52.2

years; n=96). In studies clearly reporting ethnicity, 44.9% were White European (0–100%; n=78). We estimated that 57% (n=74) studies included populations in which the majority of women were White European. At the index pregnancy, 37.6% of participants were nulliparous (9.7–100.0%; n=37). Study populations were often overweight: average BMI before pregnancy was 25.9 kg/m<sup>2</sup> (21.0–32.4 kg/m<sup>2</sup>; n=41) while the average at follow-up was 27.8 kg/m<sup>2</sup> (22.7–35.0 kg/m<sup>2</sup>; n=46). 53.4% of participants reported a family history of diabetes (7.2–100%; n=60).

### 3.1 Absolute risk of T2DM

Overall, 17.0% (95% confidence interval [CI] CI 15.1–19.0%; I<sup>2</sup> 99.3%) women across the studies developed T2DM after GDM. This ranged from 0.0% in a study with an average 1.5 years follow-up (n=68 participants followed up)[30] to 93.4% at 29.9 years follow-up in a high-risk population (n=332)[31]. Percentage developing T2DM increased in a near-linear way as study-level duration of follow-up increased (Figure 2). A third of women developed T2DM within 15 years of pregnancy. Studies in Central and South America and Africa had the highest percentage diagnosed with T2DM while those in Europe and Australasia had the lowest (Figure 3 and Supplementary Figure 2).

Table 1 (univariable analysis) and Supplementary Figure 2 show that progression to T2DM did not clearly vary with GDM or T2DM diagnostic method or sensitivity, or study quality (including when each quality domain was considered separately; Supplementary Table 5). However, studies that relied on a clinical or low sensitivity diagnosis tended to report higher percentages with T2DM than highly sensitive GDM diagnoses. Studies in which women were pregnant, on average, before the year 2000 reported higher percentages developing T2DM (Supplementary Figure 4A).

Table 1 (univariable analysis) and Supplementary Figure 3 show the percentage developing T2DM after GDM according to study-level maternal demographics. Diagnosis of T2DM was higher in studies considering women who were not of White European ethnicity, older at follow-up (reflecting longer duration of follow-up), less frequently nulliparous, and had a higher BMI at follow-up. Scatter plots showing the percentage of women developing T2DM after GDM by these variables are shown in Supplementary Figures 4B–E. Age at delivery, pre-pregnancy BMI, and the proportion of women with a family history of diabetes did not influence the estimates.

As indicated by the I<sup>2</sup> values (I<sup>2</sup> 84.0%; reported in Supplementary Figures 2 and 3), heterogeneity remained high in subgroup analyses. After adjusting for follow-up duration and ethnicity in the study-level meta-regression, for most characteristics the associations with T2DM development from univariable analysis were slightly attenuated (Table 1), although residual heterogeneity remained high throughout (I<sup>2</sup> 95.8%). In the multivariable model, White European populations had 57% lower percentage developing T2DM compared to non-White European populations (95% CI 39–70%), and percentage developing T2DM was 12% higher for each additional year of follow-up after pregnancy (95% CI 8–16%). For each study-level unit higher BMI at follow-up the percentage developing T2DM was 18% higher (95% CI 5–34%). Only one study had follow-up ≥ 20 years[31]; when this was

excluded, percentage developing T2DM remained 10% higher for each additional year of follow-up after pregnancy (95% CI 5–15%).

### 3.2 Relative risk of T2DM

Women who had GDM were 8.3 (95% CI 6.5–10.6) times more likely to develop T2DM than women with normoglycaemic pregnancies, as shown in Figure 4 (unadjusted relative risk calculated from raw data).

Full results of the stratified relative risk analysis are shown in Supplementary Table 6. The relative risk was particularly high in studies in Europe (16.1 [95% CI 12.4–21.0]) compared to studies in other regions, and in mainly White European populations (11.2 [95% CI 9.0–13.9]) compared to non-White European populations, despite such women having a lower absolute risk of developing T2DM after GDM. The relative risk was highest before six years postpartum (15.8 [95% CI 12.6–19.9]), and in studies using clinical diagnosis of T2DM (16.5 [95% CI 12.9–21.2]). Relative risk tended not to vary with other study-level maternal characteristics (measured in women with GDM), except by BMI before pregnancy. In studies where the average pre-pregnancy BMI was  $<25 \text{ kg/m}^2$ , the relative risk was comparatively low (2.1 [95% CI 1.4–3.4]). When three low-quality studies reporting very high relative risks were excluded, the overall relative risk remained at 8.1 (95% CI 6.3–10.3).

Thirteen studies reported adjusted relative analyses (odds ratios, relative risks, hazard ratios or incidence rate ratios). A history of GDM statistically significantly increased T2DM risk in all cases, but the magnitude of increase was highly variable. In five studies, the pooled adjusted odds ratio was 8.1 (95% CI 3.0–22.1), and ranged from 2.2 (95% CI 1.5–3.1; adjusted for age, BMI and family history of diabetes)[32] to 52.5 (95% CI 26.5–103.9; adjusted for age at delivery)[33]. Engeland *et al.* 2011 reported an adjusted relative risk of 41 (95% CI 35–47; adjusted for maternal age and parity in women with GDM but not preeclampsia)[34] and Sreelakshmi *et al.* 2015 reported an adjusted relative risk of 13.2 (95% CI 1.5–116.0; variables adjusted for unclear)[35]. Five studies reported adjusted hazard ratios but two did not report the CI so could not be pooled. In the remaining three studies, the pooled adjusted hazard ratio was 14.2 (95% CI 6.6–30.4), and ranged from 5.36 in Canadian First Nation women (variables adjusted for unclear)[36] to 40.1 in overweight women (95% CI 34.4–46.6; adjusted for adjusting for maternal age, preeclampsia, parity, smoking status during pregnancy, ethnicity, socioeconomic status and GDM in a subsequent pregnancy)[37]. Daly *et al.* 2018 reported an adjusted incidence rate ratio of 22.0 (95% CI 18.3–26.3; adjusted for age, Townsend quintile, BMI and smoking status)[38].

## 4 Discussion

We have shown that progression to T2DM after GDM is both common and highly variable, and while the relative risk is highest soon after pregnancy, the number of women diagnosed with T2DM continues to increase in a near-linear and clinically important way. Women with GDM may therefore have a T2DM risk that is comparable to individuals with impaired glucose tolerance or impaired fasting glycaemia[39]. Although having lower relative risk,

non-White European women have high rates of progression, as do women who are older and overweight at follow-up.

We report considerable heterogeneity in the meta-analysis, which we investigated through stratified analyses and study-level meta-regression. The heterogeneity, measured using the  $I^2$  statistic, did not improve in the multivariable meta-regression, indicating that many of the differences between populations and studies remain unexplained. This may be due to variation in study design or exposure/outcome assessment that we did not adjust for, or due to diversity within the GDM population. A proportion of the heterogeneity may be explained by variables that were measured in just a few studies, and some may remain unmeasured or unknown. Buchanan and Xiang describe different GDM phenotypes (autoimmune, monogenic and chronic insulin resistance) that are not currently assessed in GDM diagnosis[40]. It is possible that these phenotypes have different associations with development of diabetes postpartum.

Nonetheless, our findings support sustained T2DM screening after GDM; we did not identify a time after GDM at which screening might become less clinically useful. However, low long-term attendance is often reported in routine practice[12,13]. Non-White European and overweight women developed T2DM at higher rates therefore shorter screening intervals for these populations may be considered appropriate. Further research should be done to improve precision of risk stratification and determine the clinical benefit, cost effectiveness and acceptability of stratified screening strategies.

Moreover, consistent with T2DM risk factors in the general population, women with high BMI at follow-up had higher than average progression to and relative risk of T2DM. The authors of the Diabetes Prevention Programme suggested that participants with GDM did not reduce their risk because they lost less weight than comparable high-risk women[41]. Other evidence suggests that dietary and physical activity guidelines are not adhered to after GDM[42], therefore development of effective strategies to help women to manage weight in order to reduce T2DM risk is important.

#### 4.1 Comparison to existing literature

Following publication of the protocol for our review, Vounzoulaki *et al.* reported the incidence of T2DM after GDM in 20 studies (published from 2000 to 2019)[6]. They found that cumulative T2DM incidence was higher, but not statistically significantly higher, in mixed ethnicity and non-White populations than in White populations (up to 16.5% [95% CI 16.2–16.8%]), and was higher in longer study follow-up categories. However, of note, they found that effect size was not significantly associated with mean study age, BMI, publication year or length of follow-up in the univariable meta-regression analyses and suggested this was due to a lack of power. Our larger study meant that we were able to examine potential associations between these variables and others, concluding that ethnicity, time since pregnancy and BMI at follow-up were associated with diabetes risk in a multivariable analysis.

Prior to Vounzoulaki *et al.*[6], Kim *et al.* conducted an influential literature review of the cumulative incidence of T2DM after GDM in 2002 using similar inclusion criteria to



ours[43]. Adjusting for retention, they reported that cumulative incidence increased most quickly during the first five years postpartum, plateauing after ten years. Just one study with >11 years follow-up was included. This is inconsistent with the findings of individual studies, such as Lee *et al.* and Albareda *et al.*[44,45] as well as Vounzoulaki *et al.*[6]. We have reported a more constant increase in the crude proportion developing T2DM over time, including 12 studies with >11 years follow-up in the meta-regression, which supports sustained follow-up efforts. They discuss how different exclusion criteria, particularly including women with symptomatic diabetes in the GDM cohorts, might increase T2DM diagnoses soon after pregnancy, whereas we included more studies after the immediate postpartum period. They also reported that women with GDM progressed to T2DM at similar rates independent of ethnicity. In contrast, we found that White European women were less likely to progress than women from other ethnic groups.

Our relative risk estimate of 8.3 (6.5–10.6) is based on more studies and participants (45 studies and 4,376,734 women in total) than previous recent reviews, hence is more precise but highly comparable. Bellamy *et al.* reported a relative risk of 7.4 (4.8–11.5) in 20 studies published up to 2009 including 675,455 women[46]; Song *et al.* reported a relative risk of 7.8 (5.1–11.8) in 30 studies published up to 2017 including 2,626,905 women, alongside an adjusted odd ratio of 17.9 (17.0–19.0)[47]; Vounzoulaki *et al.* reported a relative risk of 9.5 (7.1–12.7) in 20 studies published between 2000 and 2019 including 1,332,373 women[6]. These data suggest that relative risk may be increasing over time. We observed similar trends across subgroups. Unlike these previous reviews, we also considered parity and family history of diabetes but they did not convincingly affect relative risk.

## 4.2 Strengths and limitations

Our meta-analysis is larger than previous ones, in part, because we did not restrict study methods, language or publication year, which enabled us to report a percentage estimate of T2DM risk in a large number of women with GDM. This increased the analysis power and consequently our opportunity for stratified and multivariable analyses to explore heterogeneity. Furthermore, we report longer follow-up than previous reviews by including new studies and updates of studies already published. Most of the studies we included had a moderate overall risk of bias in relation to T2DM outcomes (13/129 [10%] were low quality) and overall estimates of risk only changed slightly in the sensitivity analyses, which increases our confidence in the findings. There are differing views on the relative merits of ‘lumping’ and ‘splitting’ heterogeneous studies in a systematic review. We judged that by identifying, synthesising and then pooling all the evidence we could explore heterogeneity and improve understanding of the topic more than would be possible through a narrative review.

On the other hand, several limitations may have affected our findings. We only used study-level data and crude subgroups reduced accuracy. For example, diagnostic sensitivity was grouped as clinical, low or high rather than by specific criteria because numerous different criteria were used. Although we investigated incidence by 15 characteristics, some characteristics that may have explained heterogeneity were not available or not reported in a usable way for all studies. For example, few studies reported data on socioeconomic status

or other T2DM risk factors (e.g. gestational age at onset of GDM, or breastfeeding). Breastfeeding may help to prevent T2DM after GDM, although Rayanagoudar *et al.* did not find a significant association[8,9]. In the relative risk analyses, subgroups were developed according to characteristics of women with GDM only and adjusted analyses were limited. Furthermore, we combined highly variable studies, although we did not find significant differences according to many of the study characteristics included in the meta-regression (Supplementary Figure 2). We may have overlooked a few studies if they had a small sample size, short follow-up or were not included in the databases we searched, but this is unlikely to affect our conclusions.

Although it was adapted from published checklists, we used a non-validated quality assessment tool and scoring system. This allowed a more homogenized and systematic assessment of risk of bias across highly heterogeneous study designs, which may be considered as an additional strength of the study. It may, however, have introduced misclassification bias in the categorization of risk of bias of the included studies, with the possibility of underestimating the true risk of bias in the studies. Studies varied by quality, which may have influenced our analyses. In particular, a frequently observed weakness was low percentage of the study population followed up (Supplementary Figure 1). Previous studies report that women with fewer diabetes risk factors are more likely to receive follow-up than women with more risk factors[13], therefore we may have underestimated the percentage developing T2DM because those at highest risk were not tested and T2DM remained undiagnosed. However, we did not observe large differences in our estimates when only high quality studies were included according to each criteria examined in the post hoc sensitivity analysis (Supplementary Table 5). Also, the relative risk of T2DM was much higher in studies of women with clinical, as opposed to biochemical, T2DM diagnoses.

Most of the included studies had short follow-up therefore many of the women may have been yet to develop T2DM at the time of assessment. In part because we included the timepoint of studies with the most person-years follow-up, Carr *et al.* was the only study where we reported the outcome at 20 years follow-up[31]. This study was not representative of the general GDM population because all of the participants had diabetes in first-degree relatives, and inflated the estimate after 15 years postpartum. Furthermore, different factors may have confounded associations with T2DM risk. For example, older studies tend to use lower sensitivity diagnostic criteria and include women with higher glucose levels than would be the case with current criteria, therefore more of the cohort are likely to develop T2DM. These studies also tend to have longer follow-up, also increasing risk of developing T2DM. The association between follow-up duration and development of T2DM may also reflect, in part, the change in diagnostic criteria with time. Studies with longer and complete follow-up are needed in order to accurately describe progression to T2DM. Risk of diabetes may also be influenced by whether studies distinguished between diabetes in pregnancy and GDM, for example by recruiting women whose immediate postpartum tests were normal. However, excluding studies that did not attempt to exclude pre-existing/previously undiagnosed T2DM from the GDM cohort did not significantly affect the overall incidence estimate (Supplementary Table 5). Furthermore, excluding the three studies that were part of a postpartum intervention[48–50] did not influence the findings.

### 4.3 Conclusion

In this review, we have described and explored development of T2DM after GDM. Our findings strengthen the need for T2DM risk to remain on the agenda of affected women and the clinicians who care for them. Unlike previous research that suggested risk plateaued over time since pregnancy, our findings show that the number of women diagnosed with T2DM increases each year and underline the need for continued blood glucose monitoring over time. This is in line with current guidelines but is not implemented systematically, thus should be promoted. Also, the association we found between BMI and T2DM highlights the need for effective weight management strategies that are appropriate to the needs of women with a history of GDM.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank Isla Kuhn and those at the University of Cambridge Clinical School Library for their help developing the search strategy and accessing full text manuscripts. We also thank Zhirong Yang, Hannah Harrison, Julia Mannes and Parto Forouhi for their help extracting data or verifying extractions from non-English papers.

### Funding

RD was funded by a PhD studentship from the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR; SPCR-S-S102). This paper presents independent research funded by the NIHR SPCR. The views expressed are those of the author(s) and not necessarily those of the NIHR, the NHS or the Department of Health. RW was funded by an NIHR Academic Clinical Fellowship. JUS was funded by a Cancer Research UK Cancer Prevention Fellowship (C55650/A21464). SG was supported by the Medical Research Council (MC\_UU\_12015/4). The University of Cambridge has received salary support in respect of SG from the NHS in the East of England through the Clinical Academic Reserve.

### References

- [1]. Zhu Y, Zhang C. Prevalence of gestational diabetes and risk of progression to type 2 diabetes: A global perspective. *Curr Diab Rep.* 2016; 16:1–11. [PubMed: 26699764]
- [2]. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract.* 2017; 128:40–50. [PubMed: 28437734]
- [3]. [Date last accessed: 01 Sept 2020] International Diabetes Federation Diabetes Atlas. 9th 2019. Available from: <http://www.diabetesatlas.org>
- [4]. Damm P. Future risk of diabetes in mother and child after gestational diabetes mellitus. *Int J Gynecol Obstet.* 2009; 104:S25–S26.
- [5]. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Bärnighausen T, et al. The global economic burden of diabetes in adults aged 20-79 years: a cost-of-illness study. *Lancet Diabetes Endocrinol.* 2017; 5:423–30. [PubMed: 28456416]
- [6]. Vounzoulaki E, Khunti K, Abner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: Systematic review and meta-analysis. *BMJ.* 2020; 369 doi: 10.1136/bmj.m1361
- [7]. Baptiste-Roberts K, Barone BB, Gary TL, Golden SH, Wilson LM, Bass EB, et al. Risk Factors for Type 2 Diabetes Among Women with Gestational Diabetes: A Systematic Review. *Am J Med.* 2009; 122:207–214. e4 [PubMed: 19272478]
- [8]. Rayanagoudar G, Hashi AA, Zamora J, Khan KS, Hitman GA, Thangaratinam S, et al. Quantification of the type 2 diabetes risk in women with gestational diabetes: A systematic

- review and meta-analysis of 95,750 women. *Diabetologia*. 2016; 59:1403–11. [PubMed: 27073002]
- [9]. Tanase-Nakao K, Arata N, Kawasaki M, Yasuhi I, Sone H, Mori R, et al. Potential protective effect of lactation against incidence of type 2 diabetes mellitus in women with previous gestational diabetes mellitus: A systematic review and meta-analysis. *Diabetes Metab Res Rev*. 2017; 33:e2875
- [10]. American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetes. *Diabetes Care*. 2018; 41:S13–S27. [PubMed: 29222373]
- [11]. National Institute for Health Care and Excellence. Diabetes in pregnancy: management from preconception to the postnatal period. NICE Clin Guidel. 2015; 13:2–65.
- [12]. Carson MP, Frank MI, Keely E. Original research: Postpartum testing rates among women with a history of gestational diabetes-Systematic review. *Prim Care Diabetes*. 2013; 7:177–86. [PubMed: 23684416]
- [13]. Tovar A, Chasan-Taber L, Eggleston E, Oken E. Postpartum screening for diabetes among women with a history of gestational diabetes mellitus. *Prev Chronic Dis*. 2011; 8 A124 [PubMed: 22005617]
- [14]. Dennison RA, Fox RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on screening for Type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for increasing uptake. *Diabet Med*. 2020; 37:29–43. [PubMed: 31317569]
- [15]. Dennison RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on lifestyle changes to reduce the risk of developing Type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for practice. *Diabet Med*. 2019; 36:702–17. [PubMed: 30723968]
- [16]. Gigerenzer G, Edwards A. Simple tools for understanding risks: From innumeracy to insight. *Br Med J*. 2003; 327:741–4. [PubMed: 14512488]
- [17]. Noordzij M, Van Diepen M, Caskey FC, Jager KJ. Relative risk versus absolute risk: One cannot be interpreted without the other. *Nephrol Dial Transplant*. 2017; 32:ii13–8. [PubMed: 28339913]
- [18]. Wells, G. [Date last accessed: 01 Sept 2020] The Newcastle-Ottawa Scale (NOS) for assessing the quality of non randomised studies in meta-analyses. Available from:[http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp)
- [19]. [Date last accessed: 01 Sept 2020] Critical Appraisal Skills Programme (CASP) checklists. Available from:<https://casp-uk.net/casp-tools-checklists>
- [20]. Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. *Am J Obstet Gynecol*. 1982; 144:768–73. [PubMed: 7148898]
- [21]. Hoffman L, Nolan C, Wilson JD, Oats JN, Simmons D. Gestational diabetes mellitus-Management guidelines. The Australasian diabetes in pregnancy society. *Med J Aust*. 1998; 169:93–7. [PubMed: 9700346]
- [22]. American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care*. 2004; 27(1):S15–35. [PubMed: 14693923]
- [23]. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2013; 36:4–5. [PubMed: 23264284]
- [24]. Genuth S, Alberti KGMM, Bennett P, Buse J, Defronzo R, Kahn R, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*. 2003; 26:3160–7. [PubMed: 14578255]
- [25]. Metzger, BE; Coustan, DR. Diabetes Care; Summary and recommendations of the Fourth International Workshop-Conference on Gestational Diabetes Mellitus. The Organizing Committee; 1998.
- [26]. World Health Organization. [Date last accessed: 01 Sept 2020] Diabetes mellitus report of a WHO study group. 1985. Available from:<https://apps.who.int/iris/handle/10665/39592>
- [27]. World Health Organization. [Date last accessed: 01 Sept 2020] Laboratory diagnosis and monitoring of diabetes mellitus. 2002. Available from:<https://apps.who.int/iris/handle/10665/42642>
- [28]. World Health Organization. [Date last accessed: 01 Sept 2020] Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1,

Diagnosis and classification of diabetes mellitus. 1999. Available from:<https://apps.who.int/iris/handle>

- [29]. World Health Organization. [Date last accessed: 01 Sept 2020] Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. 2006. Available from:[http://www.who.int/diabetes/publications/diagnosis\\_diabetes2006/en](http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en)
- [30]. Zurawska-Kli M, Wójcik M, Zieleniak A, Kosiński M, Mazur B, Woźniak L, et al. The impact of lactation on glucose and insulin response and CRP concentration in women with prior GDM diagnosed according to WHO criteria-A prospective 18-month observation. *Clin Diabetol*. 2019; 8:99–109.
- [31]. Carr DB, Utzschneider KM, Hull RL, Tong J, Wallace TM, Kodama K, et al. Gestational diabetes mellitus increases the risk of cardiovascular disease in women with a family history of type 2 diabetes. *Diabetes Care*. 2006; 29:2078–83. [PubMed: 16936156]
- [32]. Minooee S, Ramezani Tehrani F, Rahmati M, Mansournia MA, Azizi F. Diabetes incidence and influencing factors in women with and without gestational diabetes mellitus: A 15 year population-based follow-up cohort study. *Diabetes Res Clin Pract*. 2017; 128:24–31. [PubMed: 28432896]
- [33]. Anderberg E, Carlsson KS, Berntorp K. Use of healthcare resources after gestational diabetes mellitus: A longitudinal case-control analysis. *Scand J Public Health*. 2012; 40:385–90. [PubMed: 22786924]
- [34]. Engeland A, Bjørge T, Daltveit AK, Skurtveit S, Vangen S, Vollset SE, et al. Risk of diabetes after gestational diabetes and preeclampsia. A registry-based study of 230,000 women in Norway. *Eur J Epidemiol*. 2011; 26:157–63. [PubMed: 21298469]
- [35]. Sreelakshmi PR, Nair S, Soman B, Alex R, Vijayakumar K, Kutty Vr. Maternal and neonatal outcomes of gestational diabetes: A retrospective cohort study from Southern India. *J Fam Med Prim Care*. 2015; 4:395.
- [36]. Shen GX, Shafer LA, Martens PJ, Sellers E, Torshizi AA, Ludwig S, et al. Does First Nations ancestry modify the association between gestational diabetes and subsequent diabetes: a historical prospective cohort study among women in Manitoba, Canada. *Diabet Med*. 2016; 33:1245–52. [PubMed: 26359622]
- [37]. Kaul P, Savu A, Nerenberg KA, Donovan LE, Chik CL, Ryan EA, et al. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: A population-level analysis. *Diabet Med*. 2015; 32:164–73. [PubMed: 25407209]
- [38]. Daly B, Toulis KA, Thomas N, Gokhale K, Martin J, Webber J, et al. Increased risk of ischemic heart disease, hypertension, and type 2 diabetes in women with previous gestational diabetes mellitus, a target group in general practice for preventive interventions: A population-based cohort study. *PLoS Med*. 2018; 15(1) e1002488 [PubMed: 29337985]
- [39]. Gerstein HC, Santaguida P, Raina P, Morrison KM, Balion C, Hunt D, et al. Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: A systematic overview and meta-analysis of prospective studies. *Diabetes Res Clin Pract*. 2007; 78:305–12. [PubMed: 17601626]
- [40]. Buchanan TA, Xiang AH. Gestational diabetes mellitus. *J Clin Invest*. 2005; 115:485–91. [PubMed: 15765129]
- [41]. Ratner RE, Christophi CA, Metzger BE, Dabelea D, Bennett PH, Pi-Sunyer X, et al. Prevention of diabetes in women with a history of gestational diabetes: Effects of metformin and lifestyle interventions. *J Clin Endocrinol Metab*. 2008; 93:4774–9. [PubMed: 18826999]
- [42]. Kaiser B, Razurel C. Determinants of postpartum physical activity, dietary habits and weight loss after gestational diabetes mellitus. *J Nurs Manag*. 2013; 21:58–69. [PubMed: 23339495]
- [43]. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: A systematic review. *Diabetes Care*. 2002; 25:1862–8. [PubMed: 12351492]
- [44]. Lee AJ, Hiscock RJ, Wein P, Walker SP, Permezel M. Gestational diabetes mellitus: Clinical predictors and long-term risk of developing type 2 diabetes-A retrospective cohort study using survival analysis. *Diabetes Care*. 2007; 30:878–83. [PubMed: 17392549]

- [45]. Albareda M, Caballero A, Badell G, Piquer S, Ortiz A, de Leiva A, et al. Diabetes and abnormal glucose tolerance in women with previous gestational diabetes. *Diabetes Care*. 2003; 26:1199–205. [PubMed: 12663597]
- [46]. Bellamy L, Casas J-P, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: A systematic review and meta-analysis. *Lancet*. 2009; 373:1773–9. [PubMed: 19465232]
- [47]. Song C, Lyu Y, Li C, Liu P, Li J, Ma RC, et al. Long-term risk of diabetes in women at varying durations after gestational diabetes: a systematic review and meta-analysis with more than 2 million women. *Obes Rev*. 2018; 19:421–9. [PubMed: 29266655]
- [48]. Ferrara A, Peng T, Kim C. Trends in postpartum diabetes screening and subsequent diabetes and impaired fasting glucose among women with histories of gestational diabetes mellitus: A report from the translating research into action for diabetes (TRIAD) study. *Diabetes Care*. 2009; 32:269–74. [PubMed: 18984776]
- [49]. Aroda VR, Christophi CA, Edelstein SL, Zhang P, Herman WH, Barrett-Connor E, et al. The effect of lifestyle intervention and metformin on preventing or delaying diabetes among women with and without gestational diabetes: The diabetes prevention program outcomes study 10-year follow-up. *J Clin Endocrinol Metab*. 2015; 100:1646–53. [PubMed: 25706240]
- [50]. Pérez-Ferre N, Del Valle L, Torrejón MJ, Barca I, Calvo MI, Matía P, et al. Diabetes mellitus and abnormal glucose tolerance development after gestational diabetes: A three-year, prospective, randomized, clinical-based, Mediterranean lifestyle interventional study with parallel groups. *Clin Nutr*. 2015; 34:579–85. [PubMed: 25262459]

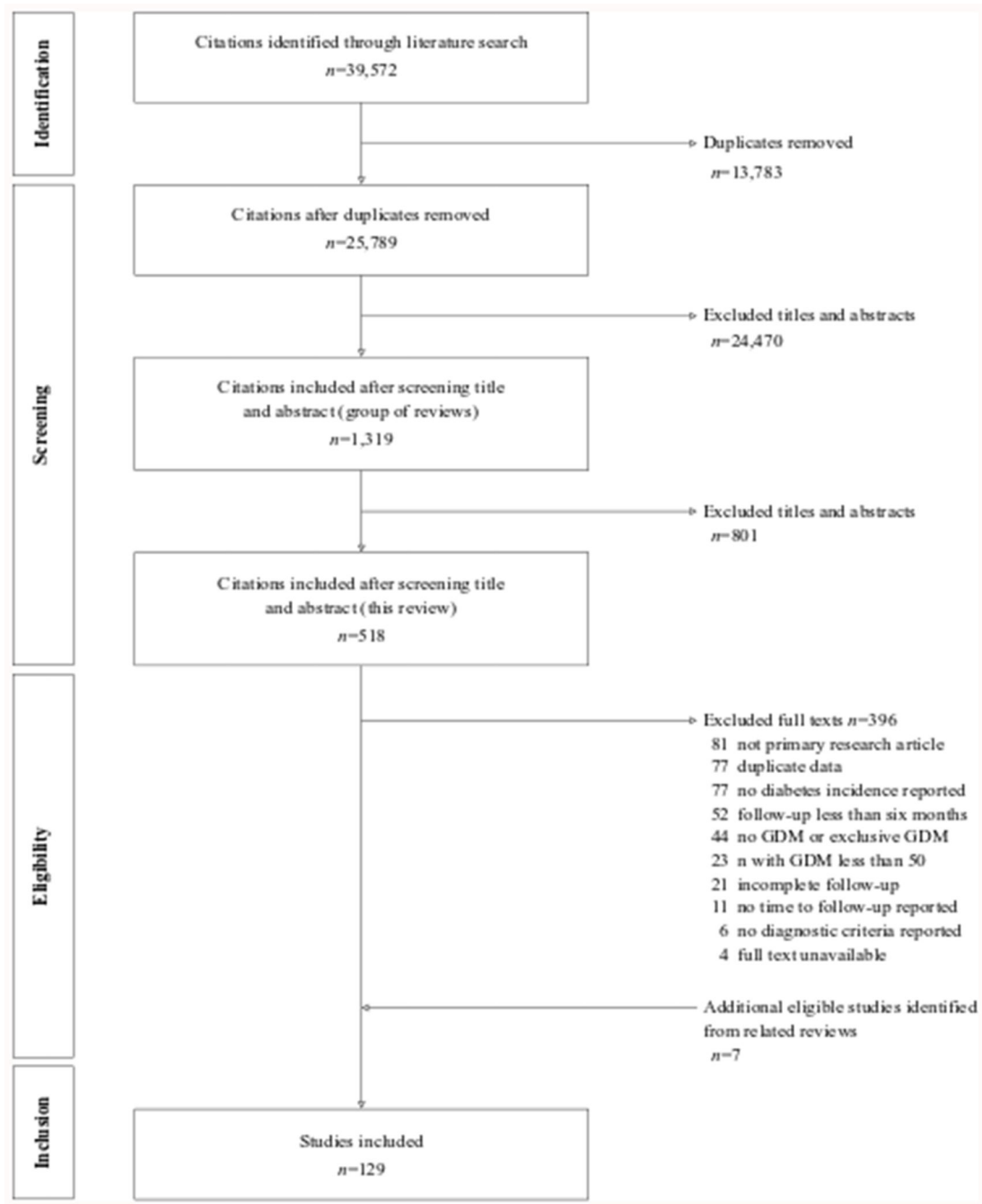
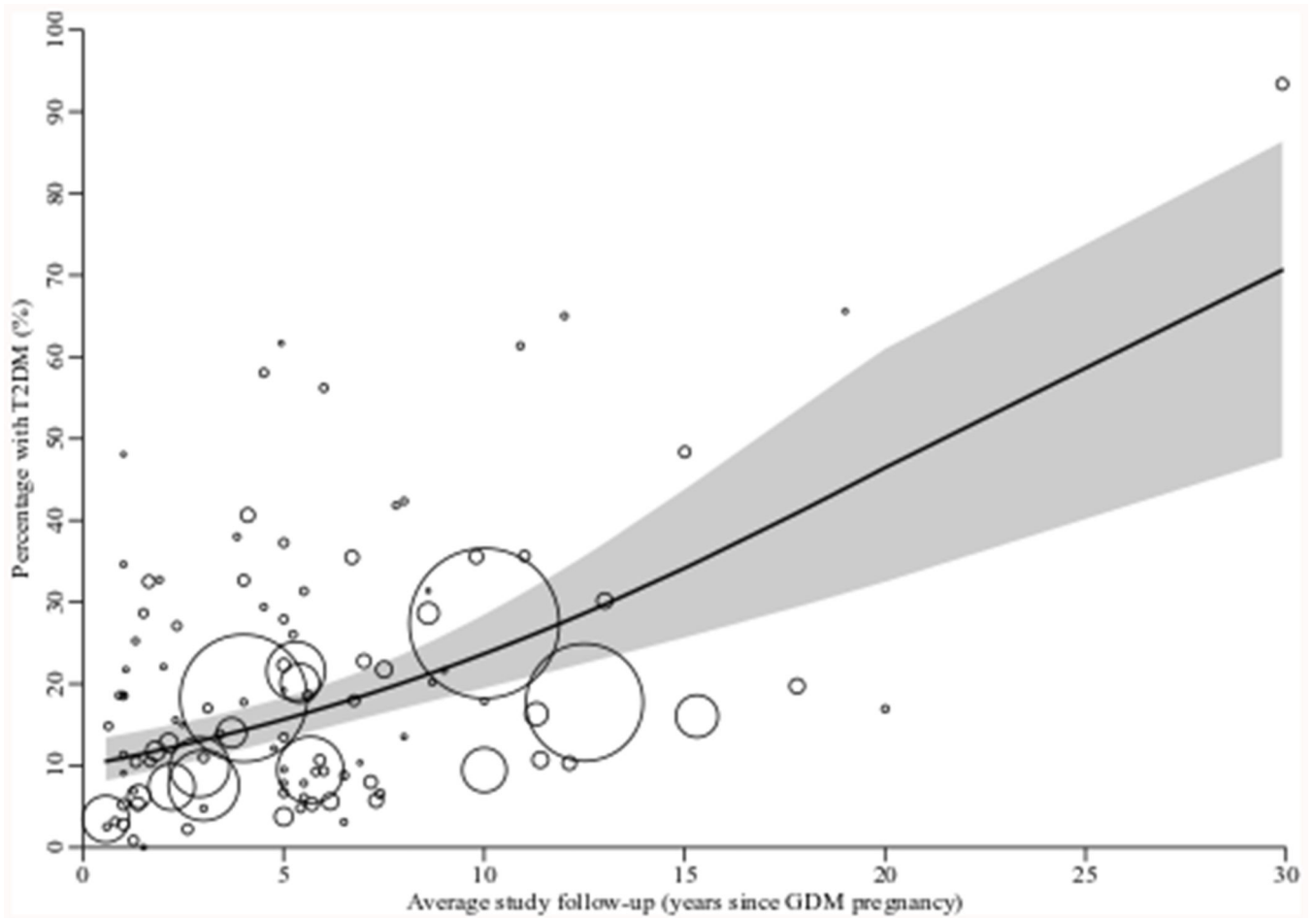
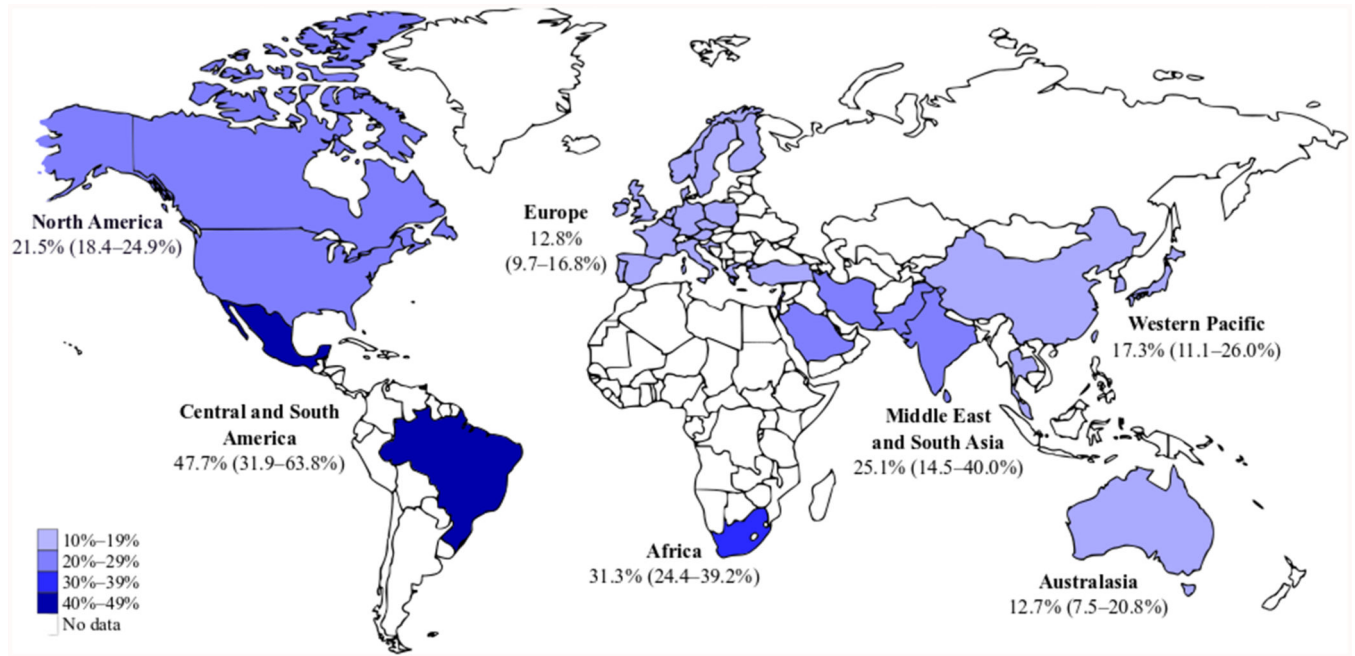


Figure 1. PRISMA diagram for the systematic review.

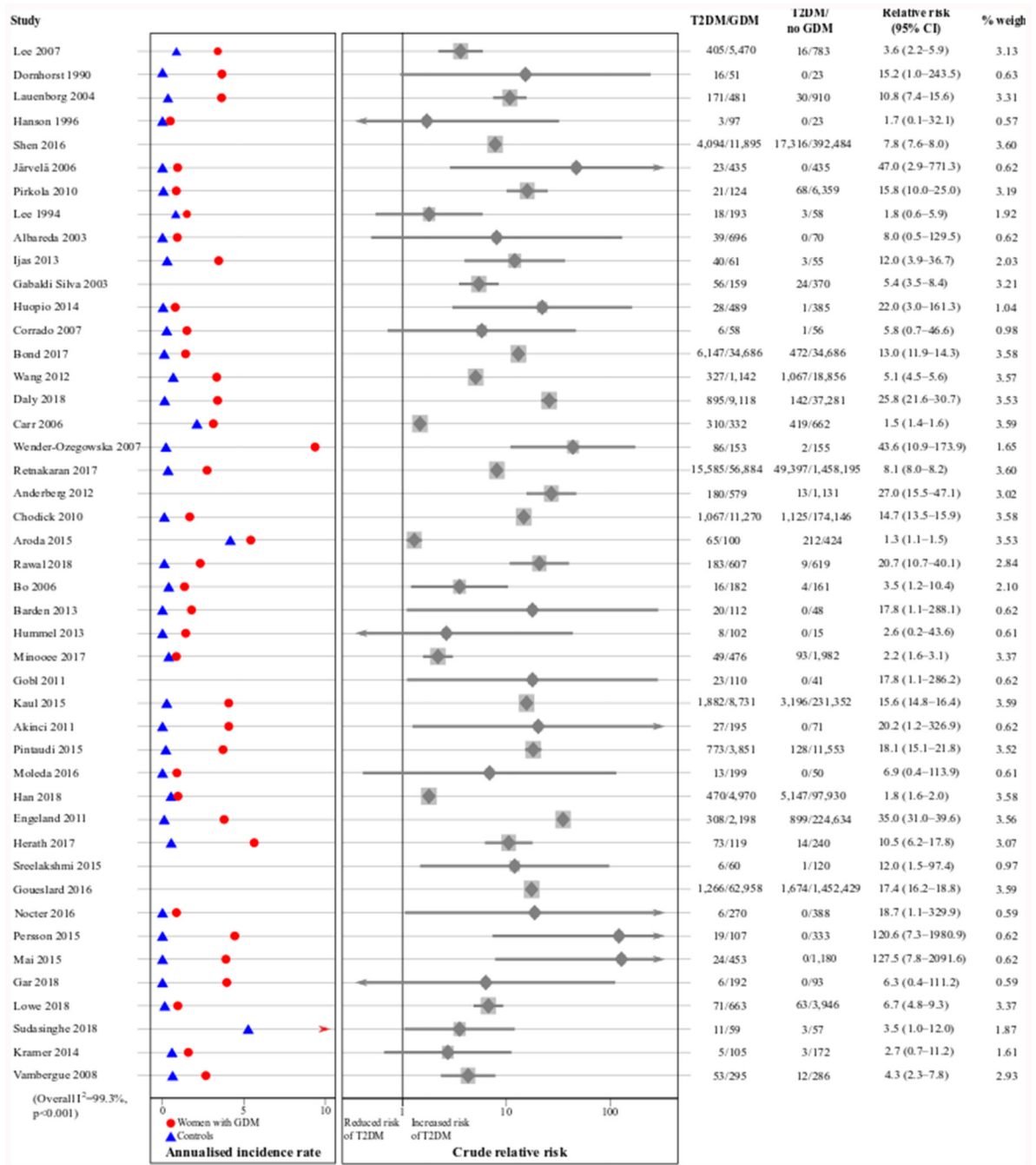


**Figure 2. Scatter plot showing the percentage of women developing T2DM after GDM by average study follow-up duration.** Circle size indicates weight given to each study; line of best fit and 95% confidence region (grey shaded area) estimated from meta-regression. N=108 studies and 226,497 women.





**Figure 3. Map showing the crude percentage and 95% confidence intervals of women with T2DM after GDM by region, estimated using random-effects meta-analysis.**



**Figure 4. The crude relative risk of T2DM after GDM compared to normoglycaemic pregnancies.**  
Annualised incidence rates are only presented for studies reporting average follow-up.  
Studies are ordered by date of pregnancy.

**Table 1**  
**Associations of categorical and/or continuous study and maternal characteristics with the incidence of T2DM after GDM.**

|   | N studies | N women | Univariable odds ratio [95% CI] | Multivariable odds ratio [95% CI] |
|---|-----------|---------|---------------------------------|-----------------------------------|
| <b>Study characteristics</b>                                |           |         |                                 |                                   |
| <b>Region</b>   |           |         |                                 |                                   |
| Australasia   | 8         | 7,081   | 0.51 [0.22–1.18]                | 0.78 [0.35–1.73]                  |
| Europe  | 61        | 96,773  | 0.54 [0.34–0.86] *              | 0.95 [0.57–1.57]                  |
| Western Pacific   | 13        | 8,416   | 0.77 [0.39–1.52]                | 0.80 [0.41–1.55]                  |
| North America   | 29        | 183,533 | Ref                             | Ref                               |
| Middle East and South Asia                                  | 13        | 13,327  | 1.24 [0.62–2.46]                | 1.28 [0.65–2.51]                  |
| Africa  | 1         | 150     | 1.67 [0.21–13.28]               | 1.66 [0.25–11.23]                 |
| Central and South America                                   | 3         | 271     | 3.39 [0.97–11.83]               | 3.51 [1.10–11.23] *               |
| Multiple  | 1         | 663     | 0.44 [0.056–3.44]               | 0.22 [0.03–1.54]                  |
| <b>Average duration of follow-up (per year)<sup>a</sup></b> |           |         |                                 |                                   |
| <3 years  | 108       | 226,497 | 1.11 [1.07–1.15] ***            | 1.12 [1.08–1.16] ***              |
| 3–5.9 years   | 38        | 28,734  | Ref                             | Ref                               |
| 6–8.9 years   | 44        | 152,531 | 1.19 [0.75–1.89]                | 1.39 [0.91–2.15]                  |
| 9–11.9 years  | 22        | 7,706   | 1.49 [0.85–2.60]                | 1.75 [1.04–2.93] *                |
| 12 years  | 13        | 67,167  | 1.91 [0.99–3.70]                | 2.44 [1.32–4.52] **               |
|   | 12        | 54,076  | 3.58 [1.81–7.05] ***            | 5.15 [2.71–9.80] ***              |
| <b>Method to identify GDM</b>                               |           |         |                                 |                                   |
| Medical records or self-report                              | 93        | 303,047 | Ref                             | Ref                               |
| Glycaemic test  | 36        | 7,167   | 0.79 [0.51–1.21]                | 0.89 [0.61–1.31]                  |
| <b>Sensitivity of GDM diagnosis</b>                         |           |         |                                 |                                   |
| Clinical  | 32        | 248,111 | Ref                             | Ref                               |
| Low   | 34        | 15,190  | 0.93 [0.55–1.57]                | 0.81 [0.50–1.33]                  |
| High  | 63        | 46,913  | 0.63 [0.40–1.00]                | 0.76 [0.50–1.17]                  |
| <b>Method to classify T2DM</b>                              |           |         |                                 |                                   |
| Medical records or self-report                              | 50        | 290,678 | Ref                             | Ref                               |
| Glycaemic test  | 79        | 19,536  | 1.29 [0.87–1.90]                | 1.24 [0.87–1.75]                  |
| <b>Sensitivity of T2DM diagnosis</b>                        |           |         |                                 |                                   |
| Clinical  | 26        | 253,865 | Ref                             | Ref                               |
| Low   | 16        | 3,715   | 1.58 [0.79–3.13]                | 1.37 [0.72–2.61]                  |
| High  | 87        | 52,634  | 1.22 [0.76–1.97]                | 1.08 [0.69–1.71]                  |
| <b>Median year of pregnancy (per year)</b>                  |           |         |                                 |                                   |
| Before 2000   | 119       | 308,085 | 0.97 [0.95–0.99] *              | 0.98 [0.96–1.00]                  |
| During/after 2000   | 50        | 71,967  | Ref                             | Ref                               |
|   | 69        | 236,109 | 0.55 [0.37–0.81] **             | 0.68 [0.46–1.00] *                |
| <b>Quality assessment score</b>                             |           |         |                                 |                                   |

|   | N studies | N women | Univariable odds ratio [95% CI] | Multivariable odds ratio [95% CI] |
|---|-----------|---------|---------------------------------|-----------------------------------|
| Low quality (score 0–2/6)   | 13        | 42,826  | Ref                             | Ref                               |
| Medium quality (score 3–4/6)                                      | 91        | 184,308 | 1.69 [0.88–3.24]                | 1.03 [0.56–1.90]                  |
| High quality (score 5–6/6)  | 25        | 83,080  | 1.60 [0.76–3.38]                | 1.18 [0.59–2.35]                  |
| <b>Maternal demographics</b>                                      |           |         |                                 |                                   |
| <b>Ethnicity (per 10% White European)</b>                         | 78        | 139,398 | 0.90 [0.85–0.95] ***            | 0.87 [0.83–0.92] ***              |
| Estimated majority not White European                             | 55        | 75,897  | Ref                             | Ref                               |
| Estimated majority White European                                 | 74        | 234,317 | 0.54 [0.37–0.79] **             | 0.43 [0.30–0.61] ***              |
| <b>Average age at delivery (per year)</b>                         | 103       | 302,579 | 0.91 [0.84–0.99] *              | 0.97 [0.89–1.05]                  |
| <32 years   | 56        | 165,708 | Ref                             | Ref                               |
| 32 years  | 47        | 136,871 | 0.79 [0.50–1.25]                | 0.94 [0.62–1.43]                  |
| <b>Average age at follow-up (per year)</b>                        | 96        | 224,169 | 1.09 [1.04–1.14] **             | 1.04 [0.90–1.20]                  |
| <38 years   | 56        | 61,607  | Ref                             | Ref                               |
| 38 years  | 40        | 162,562 | 1.94 [1.25–3.00] **             | 1.20 [0.70–2.05]                  |
| <b>Average pre-pregnancy BMI (per kg/m<sup>2</sup>)</b>           | 41        | 14,904  | 1.04 [0.91–1.18]                | 1.10 [0.93–1.31]                  |
| <25 kg/m <sup>2</sup>   | 14        | 8,752   |                                 | Ref                               |
| 25 kg/m <sup>2</sup>  | 27        | 6,152   | 1.08 [0.56–2.07]                | 1.32 [0.62–2.80]                  |
| <b>Average BMI at follow-up (per kg/m<sup>2</sup>)</b>            | 46        | 12,956  | 1.25 [1.13–1.39] ***            | 1.18 [1.05–1.34] **               |
| <25 kg/m <sup>2</sup>   | 6         | 2,072   | Ref                             | Ref                               |
| 25 kg/m <sup>2</sup>  | 40        | 10,884  | 1.57 [0.51–4.83]                | 1.50 [0.50–4.56]                  |
| <b>Average percentage who were nulliparous at index pregnancy</b> | 37        | 124,252 | 0.99 [0.97–1.01]                | 0.98 [0.96–1.00]                  |
| <35%  | 17        | 67,430  | Ref                             | Ref                               |
| 35%   | 20        | 56,822  | 0.51 [0.29–0.91] *              | 0.44 [0.24–0.81] *                |
| <b>Average percentage with family history of diabetes</b>         | 60        | 19,428  | 1.01 [0.99–1.03]                | 1.01 [0.99–1.02]                  |
| <50%  | 24        | 10,997  | Ref                             | Ref                               |
| 50%   | 36        | 8,431   | 1.06 [0.58–1.95]                | 1.18 [0.68–2.04]                  |

Data were transformed to the logit scale for analyses. Multivariable meta-regression adjusted for whether the majority of the study population is White European ethnicity and categorised duration of follow-up.  $I^2$  remained high (87.9–99.4% in the univariable model; 95.8–99.2% in the multivariable model).

<sup>a</sup>Only adjusted for whether the majority of the study population is White European. Ref: reference.

\*  $p < 0.05$ ;

\*\*  $p < 0.01$

\*\*\*  $p < 0.001$ .