

## REVIEW ARTICLE

# Interpretation of clinical trials on the cardiovascular effects of hypoglycemic drugs in people with type 2 diabetes<sup>☆</sup>



Ana Romero<sup>a</sup>, Gemma Llauradó<sup>b,c</sup>, José-Miguel González-Clemente<sup>a,c,\*</sup>

<sup>a</sup> Servicio de Endocrinología y Nutrición, Hospital Universitari Parc Taulí, Institut d'Investigació i Innovació Parc Taulí, Universitat Autònoma de Barcelona, Sabadell, Barcelona, Spain

<sup>b</sup> Servicio de Endocrinología y Nutrición, Hospital del Mar, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain

<sup>c</sup> Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain

Received 31 August 2020; accepted 1 December 2020

**KEYWORDS**

Type 2 diabetes;  
Cardiovascular  
disease;  
Cardiovascular risk;  
Randomized clinical  
trial;  
SGLT2 inhibitor;  
GLP-1 receptor  
agonist;  
DDP4 inhibitor

**Abstract** Randomized clinical trials on the cardiovascular effects of hypoglycemic drugs on people with type 2 diabetes mellitus began more than fifty years ago. In the last decade, the emergence of new classes of hypoglycemic drugs has led to the development of randomized clinical trials to assess their cardiovascular safety. Known as Cardiovascular Outcome Trials, they have provided a lot of new information that needs to be critically appraised if the knowledge obtained is to be applicable in clinical practice. To this end, the current article first comments on the guidelines to which these trials have adhered, then reviews some concepts for improving their interpretation (such as different types of analyses, the definition of objectives and the evaluation of their results), and concludes by mentioning the new guidelines to which future trials designed to evaluate the safety of new hypoglycemic drugs should adhere.

© 2021 Published by Elsevier España, S.L.U. on behalf of SEEN and SED.

**PALABRAS CLAVE**

Diabetes tipo 2;  
Enfermedad  
cardiovascular;  
Riesgo

**Interpretación de los ensayos clínicos sobre efectos cardiovasculares de los fármacos hipoglucemiantes en personas con diabetes tipo 2**

**Resumen** Hace más de cincuenta años que empezaron los ensayos clínicos aleatorizados sobre los efectos cardiovasculares de los fármacos hipoglucemiantes en personas con lo que

<sup>☆</sup> Please cite this article as: Romero A, Llauradó G, González-Clemente J-M. Interpretación de los ensayos clínicos sobre efectos cardiovasculares de los fármacos hipoglucemiantes en personas con diabetes tipo 2. *Endocrinol Diabetes Nutr.* 2021;68:741–750.

\* Corresponding author.

E-mail address: josmi.gonza@gmail.cpm (J.-M. González-Clemente).

cardiovascular;  
 Ensayo clínico  
 aleatorizado;  
 Inhibidor del  
 co-transportador  
 sodio-glucosa 2;  
 Análogos del receptor  
 del GLP1;  
 Inhibidor de la DDP4

hoy se conoce como diabetes mellitus tipo 2. Este proceso se ha acelerado en la última década con la aparición de nuevas categorías de fármacos hipoglucemiantes, lo que ha llevado al desarrollo de ensayos clínicos aleatorizados para evaluar su seguridad cardiovascular. Los denominados *Cardiovascular Outcome Trials* han proporcionado una ingente cantidad de información que debería analizarse críticamente para poderla transformar en conocimiento aplicable a la práctica clínica. Para ello, tras comentar las guías a las que se han adherido esos ensayos, se comentan una serie de conceptos para interpretarlos mejor (desde los tipos de análisis hasta la definición de objetivos y la evaluación crítica de los resultados), y se finaliza con una mención a las nuevas guías a las que se adherirán los futuros ensayos que se diseñen para evaluar la seguridad de los nuevos fármacos hipoglucemiantes.

© 2021 Publicado por Elsevier España, S.L.U. en nombre de SEEN y SED.

## Introduction: the beginnings of clinical trials on cardiovascular events

The first randomised clinical trial (RCT) on the cardiovascular effects of antidiabetic drugs in what today would be considered type 2 diabetes mellitus (T2DM) began in the 1960s. The results of that trial, the University Group Diabetes Program, suggested that insulin was safer than tolbutamide or phenformin, and sparked an intense debate for years.<sup>1</sup> In the late 1970s, the United Kingdom Prospective Diabetes Study (UKPDS) started in that country. It recruited 4,209 people with T2DM of recent onset, and its first phase ended in 1997. From a cardiovascular perspective, the study concluded that, for an endpoint made up of 12 sundry components (sudden death, death due to hypoglycaemia or hyperglycaemia, fatal or non-fatal myocardial infarction [MI], angina, heart failure, cerebrovascular accident [CVA], kidney failure, amputation of at least one finger, vitreous haemorrhage, laser-treated retinopathy, amaurosis fugax in one eye or cataract removal), more intensive blood glucose control with a mean HbA1c of 7.0% was better than standard blood glucose control with a mean HbA1c of 7.8%, with a relative risk (RR) of 0.88 (95% CI: 0.78-0.99).<sup>2</sup> In a context of an absolute lack of this type of study, we would like to point out the following limitations:

- 1 It had a long recruitment period (from 1977 to 1991) and the initial design was repeatedly changed.<sup>3</sup>
- 2 The diagnostic criteria for T2DM in that time required higher blood glucose levels than they do today.
- 3 HbA1c was not available in the first few years.
- 4 The composite endpoint mentioned included procedures (e.g. cataract removal); whether participants underwent these procedures depended on their corresponding sites.
- 5 The upper limit of the 95% CI for the result of this endpoint was 0.99 – that is, at the limit of statistical significance.
- 6 In the analysis, dozens of comparisons were made without adjusting the p value.<sup>2</sup>
- 7 The evaluation of the components of this endpoint did not take into account the time when the events appeared. It only considered whether or not they had appeared by the

end of the study; hence, the results were given in terms of RR (see the following section).

- 8 From the results in the 342 people with overweight who received metformin, this drug was inferred to have beneficial cardiovascular effects;<sup>4</sup> this has yet to be demonstrated.

At that time, benefits of intensive treatment for an endpoint consisting of fatal or non-fatal MI and sudden death were not detected ( $p = 0.052$ ), but benefits 10 years later for MI were indeed reported, especially for those who received metformin ( $p = 0.005$ ).<sup>5</sup> To summarise, the UKPDS did not show intensification of treatment of hyperglycaemia to have cardiovascular benefits. Neither did three other multicentre RCTs<sup>6-8</sup> conducted after the UKPDS, in the 2000s. In fact, one of them had to be cancelled prematurely due to an increase in total mortality [TM] in the intensive treatment group.<sup>6</sup> This did not prevent it from being included with the other two and the UKPDS in a meta-analysis that concluded that intensive blood glucose control was associated with a hazard ratio (HR) of 0.91 (95% CI: 0.78-0.99) for an endpoint consisting of cardiovascular mortality [CVM], non-fatal MI and non-fatal CVA, later termed 3-point major adverse cardiovascular events (3P-MACE).<sup>9</sup> Here, again, the result was at the limit of statistical significance.

## Questions on cardiovascular safety of thiazolidinediones

After the first thiazolidinediones were placed on the market, the PROspective pioglitAzone Clinical Trial In macroVascular Events (PROactive) RCT showed no benefits of pioglitazone versus placebo for its primary endpoint [PE] (MT, non-fatal [including silent] MI, non-fatal CVA, acute coronary syndrome, revascularisation and amputation), but did show benefits for a secondary endpoint which included only the first three components of the primary endpoint and excluded silent infarctions (HR: 0.84; 95% CI: 0.72-0.98).<sup>10</sup> It was concluded that the PE had not been significant as it included procedures (e.g. revascularisation), which could have had different indications at the different participating sites. After muraglitazar did not manage to secure market-

ing approval due to its deleterious cardiovascular adverse effects,<sup>11</sup> a meta-analysis with 42 RCTs at least 24 weeks long showed an odds ratio (OR) of 1.43 (95% CI: 1.03-1.98) for MI against rosiglitazone.<sup>12</sup> Controversy ensued, rosiglitazone was taken off the market and, in 2008, the United States Food and Drug Administration (FDA) issued guidelines on how to evaluate the cardiovascular risk of newer oral antidiabetic drugs (NOADs) for the treatment of T2DM. These guidelines can be summarised as follows:<sup>13,14</sup>

- 1 The evaluation of cardiovascular risk should involve at least one PE including examination of the three components of the 3P-MACE (CVM, MI and CVA), although other events can also be included, such as hospitalisation for acute coronary syndrome, hospitalisation for heart failure (HfHF) and emergency revascularisation.
- 2 All those events had to be defined in advance and determined by an independent committee.
- 3 That evaluation requires the NOAD development programme to include phase-2 and phase-3 RCTs sufficient size in size and duration for meta-analysis thereof to duly reflect the recommendations in the guidelines. In place of this, a specific RCT called a cardiovascular outcome trial (CVOT) can be conducted to evaluate the cardiovascular risk of NOADs. In practice, all NOADs have had their own CVOTs.
- 4 Newer insulins are not subject to these recommendations.
- 5 Fig. 1 shows the interpretation of the PE results of these CVOTs.

All CVOTs conducted since these guidelines were issued have met the criterion of a non-inferiority margin of <1.3. In some cases, as with empagliflozin, canagliflozin and semaglutide (subcutaneous or oral), the RCT was designed to exceed a cut-off point of 1.8, and the follow-up period (empagliflozin)<sup>15</sup> or the sample size (canagliflozin)<sup>16</sup> was later extended to exceed the cut-off point of <1.3, or no additional steps were taken as the cut-off point of <1.0 had already been exceeded with the initial design (semaglutide).<sup>17</sup>

## Interpretation of CVOTs

Up to 5 August 2020, 15 CVOTs were published. The comparator used was a placebo in 14 cases and another NOAD in one case. The latter will not be analysed;<sup>18</sup> those conducted after an acute coronary event<sup>19,20</sup> and those with NOADs not yet placed on the market in Spain were also not included.<sup>21</sup> Tables 1 and 2 show the 11 remaining CVOTs that will be analysed: three with sodium/glucose cotransporter 2 (SGLT2) inhibitors,<sup>15,16,22</sup> five with glucagon-like peptide 1 receptor analogues (GLP1RAs)<sup>17,23–26</sup> and three with dipeptidyl peptidase-4 (IDDP4) inhibitors.<sup>27–29</sup> In them, the following points will be considered:

### 1. Specification in advance of the entire analysis plan

When the CVOT is designed, or at least before any results are obtained, its statistical analysis plan, in which p (statistical significance) values proportionate to the degree of uncer-

tainty at that time are assumed, should be made known. This happened in each CVOT for its PE. When the results are known, they can undergo post hoc analyses, but such analyses will always be exploratory and hypothesis-generating, as they are based on the observed behaviour of the results reported.

### 2. Non-inferiority versus superiority<sup>30,31</sup>

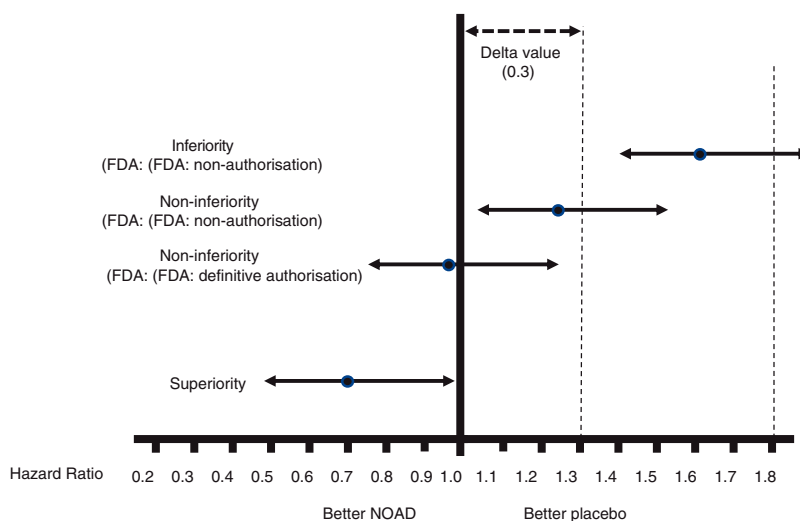
All CVOTs are non-inferiority RCTs that adhere to the 2008 FDA guidelines, except for the Researching cardiovascular Events with a Weekly Incretin in Diabetes (REWIND) trial, which was initially proposed as a superiority trial.<sup>23</sup> Unlike superiority RCTs, non-inferiority RCTs establish a delta value corresponding to the maximum acceptable difference between the NOAD and its comparator in the PE (almost always 3P-MACE) to consider whether the NOAD is non-inferior. The FDA guidelines fixed that delta value at <0.3, and therefore the cut-off point to deem a NOAD safe was <1.3 (Fig. 1). Non-inferiority does not mean equivalence, but it does mean interchangeability. In addition, it cannot be established with an RCT to demonstrate the superiority of a NOAD to its comparator, when such superiority is not demonstrated, or with multiple small-scale RCTs (e.g. n <800) showing no differences between a NOAD and its comparator in a meta-analysis. Non-inferiority RCTs require a larger sample size than superiority RCTs (see point 4); hence, within a single CVOT, it is permissible to first evaluate non-inferiority and, if confirmed, later evaluate superiority, provided that all this is specified in advance; this was common in the CVOTs analysed. If superiority is not demonstrated in a superiority RCT, possible non-inferiority cannot be evaluated afterwards, as this would require a larger sample size.

### 3. Per-protocol or as-treated analysis or intention-to-treat analysis<sup>32</sup>

Per-protocol analysis only evaluates participants who complied with the entirety of the treatment assigned, whether a NOAD or a comparator. This number is always lower than that corresponding to the intention-to-treat population, since the latter includes all participants randomised to receive a NOAD or a comparator, regardless of whether they have fully complied with it. The intention-to-treat analysis would dilute the potential effect of the NOAD and thus favour non-inferiority. According to Pocock et al.<sup>32</sup>, both types of analysis should be performed in CVOTs to confirm that they are congruent.

### 4. Sample size

When evaluating a NOAD versus a comparator, two types of error are defined: a) Type I error (alpha): the NOAD is thought to increase cardiovascular risk, but it actually does not (false positive). The alpha value in CVOTs is 0.05 (bilateral). b) Type II error (beta): the NOAD is thought not to increase cardiovascular risk, but it actually does (false negative). The power is  $(1 - \beta) \times 100$ , and is usually 90% in CVOTs.



**Figure 1** Interpretation of the results for the PE of the CVOTs according to the 2008 FDA guidelines.<sup>13</sup> This interpretation is based on: 1) Acceptance of a delta value  $<0.3$  to accept a NOAD as safe (or non-inferior). 2) Attention to the upper limit of the 95% CI (bilateral  $p$ ) for the difference between people who received the NOAD and people who did not, in relation to the PE (in general 3P-MACE): If it is  $\geq 1.8$ , the NOAD will not be authorised due to its high cardiovascular risk; if it is  $<1.3$ , it will be assumed to be safe and authorised to be marketed; if it is  $\geq 1.3$  and  $<1.8$ , it will be provisionally authorised to be marketed (or not suspended), pending more results confirming that this upper limit of the difference is  $<1.3$ . The guidelines do not recommend that the NOAD show cardiovascular benefits, or in other words that this upper limit of the difference be  $<1.0$  (superiority); they only recommend demonstrating non-inferiority (cut-off point  $<1.3$ ).

Five factors increase sample size in non-inferiority RCTs:<sup>33</sup> a) A lower rate of PE events in the control group. b) A lower rate of events in the PE with the NOAD (a lesser effect of the NOAD on the PE). c) A lower alpha risk. d) A lower delta value; the FDA recommended a delta value of  $<0.3$  (which meant recognising NOADs that could increase the risk of the PE by up to  $<30\%$  as safe) for pragmatic reasons, because for example for delta values of  $<0.1$ , CVOTs with more than 50,000 participants would be needed. e) A higher statistical power; this depends on the number of expected events in the PE, which for a HR of 1.3 is usually  $>700$ . CVOTs are designed to stop when a prespecified number of those events is reached (they are said to be event-driven trials). All the CVOTs analysed exceeded that number (Table 1).

## 5. Inclusion/exclusion criteria

The inclusion/exclusion criteria determine the external validity of the CVOT (extrapolation of its results to all people with T2DM). In addition to the demographic characteristics of the participants, one very important criterion is the percentage of participants who have not yet experienced cardiovascular events (primary prevention), since they have a lower absolute risk (AR) of experiencing them than the participants who have already had them (secondary prevention) (Table 1). When comparing CVOTs, it must be taken into consideration that their criteria for selecting participants in primary prevention are usually not identical. For example, in the CANagliflozin cardiovascular Assessment Study (CANVAS),<sup>16</sup> the group in primary prevention had a higher AR of 3P-MACE due to its members' clinical characteristics than the group in primary prevention

from the Dapagliflozin Effect on Cardiovascular Events-Thrombolysis In Myocardial Infarction 58 (DECLARE-TIMI) trial.<sup>22</sup> The REWIND trial criteria<sup>23</sup> are also not comparable to the criteria of the Trial to Evaluate Cardiovascular and Other Long-term Outcomes with Semaglutide in Subjects with Type 2 Diabetes (SUSTAIN-6)<sup>17</sup> or the *Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results* (LEADER) trial.<sup>26</sup> Furthermore, initial cardiovascular events may be asymptomatic. Thus it seems most reasonable to compare CVOTs according to the AR of the PE in the control group.<sup>34</sup> The DECLARE-TIMI trial<sup>22</sup> was the CVOT with the lowest AR (2.4% per year or 2.4 per 100 patient-years), and the Harmony Outcome (Harmony) trial was the one with the highest (5.9% per year) (Table 1).<sup>24</sup>

When the groups are compared it is interesting to observe the changes over time in the cumulative percentage of participants with any PE event with Kaplan–Meier curves. In the Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG), the curves started to diverge within a matter of months in favour of NOADs,<sup>15</sup> whereas they diverged after almost two years in the LEADER trial.<sup>26</sup> Consequently, it was suggested that the NOAD in the first study might have beneficial effects through a more haemodynamic mechanism, while the NOAD in the second study might have beneficial effects through a more anti-arteriosclerotic mechanism.

## 6. Run-in period

This is a period prior to randomisation intended to ensure that the participants to be randomised adhere well to the trial treatments, though at the expense of reducing the trial's external validity. The best way to assess it is by

**Table 1** Characteristics of the CVOTs analysed.

Name NOAD n	Expected/observed PE events	Age (years)/Females (%)/BMI (kg/m <sup>2</sup> )	Years since onset of T2DM/kidney failure (%)	Primary prevention (%)	Control group absolute risk (% per year)
<i>SGLT2 inhibitors</i>					
DECLARE-TIMI <sup>22</sup>	1,390/1,559	64	11	59	2.4
Dapagliflozin 17,160		37 32	7		
CANVAS <sup>16</sup>	698/1011	63	14	44	3.2
Canagliflozin 10,142		36 32	20		
EMPA-REG <sup>15</sup>	691/792	63	57% > 10	1	4.4
Empagliflozin 7,020		29 31	26		
<i>GLP1RAs</i>					
REWIND <sup>23</sup>	1,200/1,257	66	11	69	2.7
Dulaglutide 9,901		46 32	22		
Harmony <sup>24</sup>	661/776	64	14	0	5.9
Albiglutide 9,463		31 32	23		
EXSCEL <sup>25</sup>	1,360/1,744	62	12	27	4.0
Exenatide ER 14,752		38 32	22		
SUSTAIN-6 <sup>17</sup>	122/254	65	14	29	4.2
Semaglutide 3,297		39 33	24		
LEADER <sup>26</sup>	611/1,302	64	13	19	3.9
Liraglutide 9,340		36 33	23		
<i>DPP4 inhibitors</i>					
CARMELINA <sup>27</sup>	611/854	66	15	43	5.6
Linagliptin 6,979		37 31	62		
TECOS <sup>28</sup>	1,300/1,690	65	12	0	4.2
Sitagliptin 14,671		29 30	23		
SAVOR-TIMI <sup>29</sup>	1,040/1,222	65	10	21	3.6
Saxagliptin 16,492		33 31	16		

CARMELINA: Cardiovascular and Renal Microvascular Outcome Study with LINAGliptin; DPP4 inhibitors: dipeptidyl-peptidase 4 inhibitors; GLP1RAs: glucagon-like peptide 1 receptor analogues; NOAD: newer oral antidiabetic drug; PE: primary endpoint; SGLT2 inhibitors: sodium/glucose cotransporter 2 inhibitors; T2DM: type 2 diabetes mellitus.

examining the participant flow chart that appears in these publications. It is a good idea to then evaluate both the percentage of potential participants who did not get past the run-in phase and the reason why, although sometimes this is not specified in large numbers of cases.<sup>23</sup>

## 7. Interventions, randomisation and blinding

In all CVOTs, the intervention consisted of randomly adding the NOAD or its placebo to participants' regular treatment for T2DM, thus minimising between-group differences in blood glucose control — in other words, pursuing glycaemic equipoise. This is because the goal is to evaluate the cardiovascular safety of the NOAD independently of its antidiabetic

effects. To achieve this equilibrium, other antidiabetic drugs tend to be used, the use of which is usually not authorised until several weeks/months after randomisation. This means that the mean HbA1c level is always somewhat lower in participants with the NOAD (Table 2). Some CVOTs have used two different NOAD doses but have specified in advance that those doses would be considered jointly to evaluate the PE.<sup>15–17</sup> The CVOT with the greatest between-group difference in HbA1c was the SUSTAIN-6,<sup>17</sup> consistent with its NOADs having the greatest effects on HbA1c compared to all the other antidiabetic drugs with which it has been compared to date.<sup>35–37</sup>

All the CVOTs had another three characteristics:<sup>33</sup> a) The investigator did not know in advance which treatment would

**Table 2** Characteristics of the CVOTs analysed.

Name NOAD n	Baseline HbA1c	Mean control/NOAD HbA1c difference	Median follow-up (years)	Did not complete placebo or NOAD (%)	Hazard ratio (95% CI) for PE
<i>SGLT2 inhibitors</i>					
DECLARE-TIMI <sup>22</sup> Dapagliflozin 17,160	8.3	-0.4	4.2	23	0.93 (0.84-1.03) 0.83 (0.73-0.95)
CANVAS <sup>16</sup> Canagliflozin 10,142	8.2	-0.58	3.7-2.4	30	0.86 (0.75-0.97)
EMPA-REG <sup>15</sup> Empagliflozin 7,020	8.1	-0.3	3.1	25	0.86 (0.74-0.99)
<i>GLP1RAs</i>					
REWIND <sup>23</sup> Dulaglutide 9,901	7.3	-0.61	5.4	27	0.88 (0.79-0.99)
Harmony <sup>24</sup> Albiglutide 9,463	8.7	-0.63	1.6	26	0.78 (0.68-0.90)
EXSCEL <sup>25</sup> Exenatide ER 14,752	8.0	-0.53	3.2	44	0.91 (0.83-1.00)
SUSTAIN-6 <sup>17</sup> Semaglutide 3,297	8.7	-0.7/-1.0	2.1	20	0.74 (0.55-0.95)
LEADER <sup>26</sup> Liraglutide 9,340	8.7	-0.4	3.8	20	0.87 (0.78-0.97)
<i>DPP4 inhibitors</i>					
CARMELINA <sup>27</sup> Linagliptin 6,979	8.0	0.36	2.2	26	1.02 (0.89-1.17)
TECOS <sup>28</sup> Sitagliptin 14,671	7.2	-0.3	3.0	23	0.99 (0.89-1.12)
SAVOR-TIMI <sup>29</sup> Saxagliptin 16,492	8.0	-0.3	2.1	19	1.00 (0.89-1.12)

CARMELINA: Cardiovascular and Renal Microvascular Outcome Study with LINAgliptin; DPP4 inhibitors: dipeptidyl-peptidase 4 inhibitors; GLP1RAs: glucagon-like peptide 1 receptor analogues; NOAD: newer oral antidiabetic drug; PE: primary endpoint; SGLT2 inhibitors: sodium/glucose cotransporter 2 inhibitors.

be assigned to the participant if included (allocation concealment). b) Both the investigator and the participant, at a minimum, did not know whether the product that the participant received was a NOAD or a placebo (double blinding). c) NOADs and placebos were randomly assigned, thus ensuring that the participants in the different treatment groups were comparable in terms of cardiovascular risk factors, whether known or unknown.

## 8. Follow-up and exposure

CVOTs are long-term studies in which it is essential to evaluate exposure to the NOAD — i.e. to determine how many participants received it and for how long (Table 2). The CVOT

with the most participants was the DECLARE-TIMI trial.<sup>22</sup> The one with the fewest was the SUSTAIN-6.<sup>17</sup> The one with the longest mean follow-up period was the REWIND trial,<sup>23</sup> and the one with the shortest was the Harmony trial.<sup>24</sup> The one with the lowest rate of treatment dropout was the Saxagliptin Assessment of Vascular Outcomes Recorded in patients with diabetes mellitus-Thrombolysis in Myocardial Infarction (TIMI) 53 (SAVOR-TIMI) trial,<sup>29</sup> and the one with the highest was the Exenatide Study of Cardiovascular Event Lowering (EXSCEL).<sup>25</sup> The best way to report exposure is in participant-years, although this is often not reported or difficult to find in publications. Low exposure to the NOAD should be taken into consideration in assessing the effects on the PE. The percentage of randomised patients with information

on their vital status should also be evaluated at the end of the trial, although it was very high in all the CVOTs.

## 9. Primary endpoint

All the CVOTs had as a PE time to onset of one of the events comprising that PE, usually 3P-MACE, except in the Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS), which added hospitalisation for unstable angina to the 3P-MACE in its PE,<sup>28</sup> and the DECLARE-TIMI trial, which ultimately evaluated a co-primary endpoint of 3P-MACE and a composite endpoint of CVM or HfHF.<sup>22</sup> For each participant: a) Both whether and when the event appears are evaluated. b) Only the first PE component to appear, if any, is considered. For example, if a participant first has an MI and then has a CVA or CVM, he or she will only be considered to have had an MI. Between-group comparison is always done with Cox regression (or proportional risk) models, which provide results in terms of HR. Although HR is similar to RR, they are not exactly the same, as the time when the event occurs is used to calculate HR, whereas RR is calculated at a particular point in follow-up (e.g. at two years) considering the ratio of the proportion of events between the two groups at that time but not considering when the events occurred.<sup>38</sup> In evaluating HR it is also necessary to consider the upper limit of its 95% CI, as it is a measure of its statistical significance (i.e. its p value). This limit was 0.99 in two of the CVOTs, indicating a result at the limit of statistical significance (Table 2).<sup>15,23</sup>

Another less common way of reporting the beneficial cardiovascular effects of a NOAD is to provide the number of participants who must be treated in order to prevent a PE event (number needed to treat). This is of great clinical interest. It cannot be easily calculated as the inverse of the reduction in the AR that the NOAD achieves,<sup>39</sup> and it should be provided by the authors. In addition, it should only be calculated for the PE of the CVOT. It should not be calculated for the secondary endpoints. It should also not be calculated for the results of subgroup analyses, as was recently done in one of the CVOTs analysed.<sup>23</sup>

All the CVOTs have composite PEs. This enables detection of a larger number of events in a shorter period of time, thus increasing the study's power. It also allows the effects of the NOAD to be better captured, although the milder components (such as hospitalisation for unstable angina in the TECOS)<sup>28</sup> tend to appear earlier, thus diluting the effects of the NOAD and favouring non-inferiority.<sup>32</sup> In any case, the results for each of the components of the PE should be shown to see whether they all have the same meaning. The only CVOT in which that did not happen was the EMPA-REG, which with its NOADs found a non-significant increase in non-fatal CVA with a non-significant drop in non-fatal MI and a significant drop in CVM, thus rendering its results less sound.<sup>15</sup>

## 10. Sensitivity analyses and subgroup analyses

To increase the robustness of the results for the PE, it is advisable to perform sensitivity analyses that confirm it.<sup>40</sup> This means performing analyses considering only the per-protocol or intention-to-treat population, or evaluating

the data based on the participants lost. If these analyses yield the same results as the initial analysis, they would confirm that initial analysis. In the EMPA-REG, the per-protocol results were not significant, but the intention-to-treat results were significant.<sup>15</sup>

It is also a good idea to perform subgroup analyses.<sup>32</sup> These analyses explore the possible heterogeneity of the effects of the NOAD on the PE in different participant subgroups defined by their baseline characteristics, such as age group, sex, geographic origin or lack of prior cardiovascular events (primary prevention). They are performed by calculating a p for interaction: if significant, it would indicate that the participants would behave differently depending on the subgroup-defining characteristic. This happened, for example, in the EMPA-REG when the PE was analysed based on the geographic origin of the participants.<sup>15</sup> This type of analysis is exploratory only and raises the problem of multiple comparisons. When 10 analyses of this type are performed in a single study considering a p for interaction <0.05 to be significant, the risk of obtaining at least one false positive is >40%.<sup>41</sup> The management of this problem is described in point 12. As it is exploratory, if, for example, the subgroup of patients in primary prevention is confirmed to exhibit behaviour equal to the subgroup of patients in secondary prevention, this does not mean that it can be affirmed that the beneficial effects on the PE detected when the two subgroups are analysed at the same time can be extrapolated to the subgroup in primary prevention (in which it is always more difficult to demonstrate any cardiovascular benefits). This occurred in the REWIND trial, the CVOT with the highest percentage of participants in primary prevention out of all those conducted, which noted that its NOADs significantly improved the PE in the subgroup in primary prevention.<sup>23</sup> A subsequent meta-analysis analysed the cardiovascular effects of all the CVOTs with GLP1RAs according to whether the participants were in primary prevention (the majority were from the REWIND trial) or secondary prevention, and concluded that there were no benefits in primary prevention and that there were benefits in secondary prevention.<sup>42</sup>

## 11. Adaptive designs

Adaptive designs can improve the efficiency of RCTs, with potential benefits for participants, as they reduce costs and increase the chances of finding any potential benefits. There are various types, but they must always be specified in advance or, if the RCT is already in progress, used without knowing the trial's results.<sup>43</sup> This type of design was used in the EMPA-REG, when silent infarctions were removed from the PE before the results were published. Silent infarctions were more common in the NOAD groups than in the placebo group (7.0 versus 5.4 per 1,000 participants per year), but it cannot be known whether not having made that change would or would not have modified the end result of the study, which was at the limit of statistical significance, as the amount of time they took to appear was not reported.<sup>15</sup> The DECLARE-TIMI trial represents another example; it was initially proposed with the PE of non-inferiority for 3P-MACE.<sup>22</sup> During the trial, in view of the results of the EMPA-REG in terms of HfHF, without knowing its results, a composite end-

**Table 3** March 2020 FDA proposal for studying the safety of NOADs, replacing the 2008 guidelines for studying the cardiovascular safety of NOADs.<sup>46</sup>**March 2020 FDA recommendations for evaluating the safety of newer oral antidiabetic drugs (NOADs)**

1. Before being authorised to be placed on the market, NOADs should have controlled clinical trials (including extension studies) with at least 4,000 patient-years of exposure to NOADs in phase-3 trials (all doses are valid), with at least 1,500 patients exposed for at least one year and at least 500 patients exposed for at least two years
2. Inclusion of patients with special characteristics:
  - At least 500 with stage-3 or stage-4 chronic kidney failure
  - At least 600 with established cardiovascular disease (prior myocardial infarction, cerebrovascular accident or peripheral artery disease, or documented coronary disease)
  - And at least 600 patients over 65 years of age
  - At least 1,200 patients should fall under one of those three categories
3. Evaluation of cardiovascular adverse effects remains important. If an adverse effect particular to a NOAD is anticipated, it is advisable to contact the FDA for the design of phase-3 trials.
4. Evaluation of possible adverse effects of the NOAD continues to require independent committees.

point of CVM or HfHF was added as a co-primary endpoint to the initial 3P-MACE. To adjust the alpha error, its initial value of 0.05 was halved, such that, at the end of the trial, both co-primary endpoints were going to need a p value <0.0231 to be considered significant, and both were going to have to be significant to hierarchically analyse the effects on a number of secondary endpoints specified in advance at the start of the trial (with  $p < 0.05$ ).<sup>44</sup> In this trial, 3P-MACE was not significant ( $p = 0.17$ ), but the other co-primary endpoint of CVM or HfHF was significant ( $p = 0.005$ ).

## 12. Secondary and exploratory endpoints

All the CVOTs included secondary endpoints. All are exploratory and hypothesis-generating, although if analyses thereof yielded a very low p value (e.g. <0.001), they would have to be considered potentially significant. Both in the EMPA-REG<sup>15</sup> and in the CANVAS,<sup>16</sup> HfHF was a secondary endpoint in which substantial benefits of the respective NOAD were seen, leading to specific trials for the treatment of heart failure.<sup>45</sup>

It is recommended that all of the components of the PE be included as secondary endpoints. For agencies like the FDA to consider their results sufficient to amend the package insert/summary of product characteristics (as occurred for example with CVM and the EMPA-REG), two strategies can be employed:<sup>32</sup> a) Prespecify a hierarchy for evaluation of secondary endpoints without modifying the initial p value, but such that, to move from one endpoint to the next, the first of them must be significant; this is the strategy most commonly used in CVOTs. b) Do not prepare this hierarchy but do adjust the p value downwards based on the number of secondary endpoints (multiple comparisons) that are specified in advance. For example, if the initial p value is <0.05 and there are five secondary endpoints, for any of them to be accepted as significant, its p value should be <0.01. This strategy would probably be the most neutral one.<sup>44</sup>

## 13. Evaluation of adverse events

CVOTs can confirm known adverse effects of the NOADs evaluated, such as genital infections with SGLT2 inhibitors and

gastrointestinal intolerance with GLP1RAs. In addition, as they include thousands of participants, they can also aid in detecting potential new adverse effects, which may or may not be confirmed later on, such as the development of retinopathy<sup>17</sup> or an increase in amputations.<sup>16</sup> Their onset must also be analysed in the placebo group, since if for example an increase in episodes of hypoglycaemia is seen in the NOAD group, such analysis could help to partly explain the results.<sup>26</sup>

## The new FDA proposal

After a meeting of an expert committee in October 2018, which determined that none of the CVOTs conducted with the 2008 guidelines detected an increase in cardiovascular risk with NOADs, as well as the need to evaluate people with T2DM with a higher risk of adverse effects and other types of adverse effects (not just cardiovascular ones), in March 2020, the FDA issued new recommendations on how to evaluate the safety of NOADs; they are summarised in Table 3.<sup>46</sup> These new recommendations replace the 2008 ones. They do not propose cut-off points for establishing the safety of the NOAD and they indicate that the safety assessment should be adaptive. These recommendations establish as a fulcrum a minimum exposure of participants to the NOAD of 4,000 patient-years, which in general would be lower than that recorded in the above-mentioned CVOTs; however, they call for the participation of minimum numbers of people with kidney failure, cardiovascular disease and age over 65 years. With those recommendations, new clinical trials could accurately evaluate fewer cardiovascular events to obtain authorisation of their respective NOADs.<sup>47</sup> Surprisingly, the HR cut-off points for the PE have been abandoned, and the conduct of pragmatic trials (trials that preserve randomisation but are embedded in routine clinical practice) have not been recommended as a strategy for including participants as similar as possible to potential candidates for receiving the NOAD.<sup>48</sup>

In summary, to better assess the potential impact on clinical practice of CVOTs, it must be considered that:



- 1 All analyses must be specified in advance; post hoc analyses are always exploratory.
- 2 A larger sample size is needed to demonstrate non-inferiority than to demonstrate superiority.
- 3 It is more important to know the rate of events in the control (placebo) group than the percentage of participants in primary prevention, and it is also important to assess NOAD exposure.
- 4 Run-in periods detract from the external validity of the study results.
- 5 CVOTs always attempt to minimise between-group differences in HbA1c, and are not suitable for measuring the efficacy of NOADs in relation to blood glucose.
- 6 All end when a prespecified number of events is reached.
- 7 The PE always has multiple components, and in each participant only the first component to occur and the timing of its occurrence are taken into account. In assessing the HR of the PE, if the upper limit of its 95% CI is close to 1.0 (e.g. 0.99), this indicates that the result was at the limit of statistical significance. Clinically, it is very interesting to know the number of participants who should be treated to prevent a component of the PE.
- 8 Sensitivity analyses can lend robustness to the results obtained for the PE; subgroup analyses in relation to the PE and analyses of secondary endpoints are exploratory and hypothesis-generating.
- 9 It is important to confirm whether changes occurred in the PE over the course of the trial. These considerations may aid in more objective evaluation of CVOTs, a task that remains dependent on human critical thinking and cannot be assumed by artificial intelligence.<sup>49</sup>

## Funding

This study was funded by PI15/00567 (PI: JMCG) from the Fondo de Investigación Sanitaria [Healthcare Research Fund] (FIS), in turn co-funded by the Instituto de Salud Carlos III [Carlos III Health Institute]/General Evaluation Branch (Spanish Ministry of Economics and Competitiveness) and the European Regional Development Fund (ERDF).

## Conflicts of interest

AR: none. GL has received funding for lectures and attendance at medical conferences/courses since 2018 from Abbott, Janssen and Sanofi. JMGC has received funding for lectures as well as attendance at medical expert committees and conferences/courses since 2018 from AstraZeneca, Janssen, Menarini, MSD, Novo Nordisk and Sanofi. JMGC participated in the CANVAS as an investigator.

## References

1. Blackburn H, Jacobs DR Jr. The University Group Diabetes Program 1961-1978: Pioneering randomized controlled trial. *Int J Epidemiol.* 2017;46:1354–64.
2. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet.* 1998;352:837–53.
3. Ewart RM. The case against aggressive treatment of type 2 diabetes: Critique of the UK prospective diabetes study. *BMJ.* 2001;323:854–8.
4. McCormack J, Greenhalgh T. Seeing what you want to see in randomized controlled trials: versions and perversions of UKPDS data. United Kingdom prospective diabetes study. *BMJ.* 2000;320:1720–3.
5. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med.* 2008;359:1577–89.
6. Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, Buse JB, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med.* 2008;358:2545–59.
7. Patel A, MacMahon S, Chalmers J, Neal B, Billot L, Woodward M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med.* 2008;358:2560–72.
8. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, et al. Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med.* 2009;360:129–39.
9. Turnbull FM, Abraira C, Anderson RJ, Byington RP, Chalmers JP, Duckworth WC, et al. Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia.* 2009;52:2288–98.
10. Dormandy JA, Charbonnel B, Eckland DJ, Erdmann E, Massi-Benedetti M, Moules IK, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): A randomized controlled trial. *Lancet.* 2005;366:1279–89.
11. Nissen SE, Wolski K, Topol EJ. Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. *JAMA.* 2005;294:2581–6.
12. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med.* 2007;356:2457–71.
13. U.S. Food and Drug Administration. Guidance for industry: Diabetes mellitus—evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes [accessed 29 Feb 2020]. Available from: [www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf).
14. Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, et al. Cardiovascular outcomes trials in type 2 diabetes: Where do we go from here? Reflections from a diabetes care editors' expert forum. *Diabetes Care.* 2018;41:14–31.
15. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med.* 2015;373:2117–28.
16. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med.* 2017;377:644–57.
17. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, et al. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med.* 2016;375:1834–44.
18. Rosenstock J, Kahn SE, Johansen OE, Zinman B, Espeland MA, Woerle HJ, et al. Effect of linagliptin vs glimepiride on major adverse cardiovascular outcomes in patients with type 2 diabetes: The CAROLINA randomized clinical trial. *JAMA.* 2019;322:1155–66.
19. Pfeffer MA, Claggett B, Diaz R, Dickstein K, Gerstein HC, Køber LV, et al. Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. *N Engl J Med.* 2015;373:2247–57.
20. White WB, Cannon CP, Heller SR, Nissen SE, Bergenstal RM, Bakris GL, et al. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. *N Engl J Med.* 2013;369:1327–35.
21. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, et al. Oral semaglutide and cardiovascular

- outcomes in patients with type 2 diabetes. *N Engl J Med.* 2019;381:841–51.
22. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2019;380:347–57.
  23. Gerstein HC, Colhoun HM, Dagenais GR, Diaz R, Lakshmanan M, Pais P, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): A double-blind, randomised placebo-controlled trial. *Lancet.* 2019;394:121–30.
  24. Hernandez AF, Green JB, Janmohamed S, D’Agostino RB Sr, Granger CB, Jones NP, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): A double-blind, randomised placebo-controlled trial. *Lancet.* 2018;392:1519–29.
  25. Holman RR, Bethel MA, Mentz RJ, Thompson VP, Lokhnygina Y, Buse JB, et al. Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2017;377:1228–39.
  26. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JF, Nauck MA, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2016;375:311–22.
  27. Rosenstock J, Perkovic V, Johansen OE, Cooper ME, Kahn SE, Marx N, et al. Effect of linagliptin vs placebo on major cardiovascular events in adults with type 2 diabetes and high cardiovascular and renal risk: The CARMELINA randomized clinical trial. *JAMA.* 2019;321:69–79.
  28. Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, et al. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. *N Engl J Med.* 2015;373:232–42.
  29. Scirica BM, Bhatt DL, Braunwald E, Steg PG, Davidson J, Hirshberg B, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med.* 2013;369:1317–26.
  30. Mauri L, d’Agostino RB Sr. Challenges in the design and interpretation of noninferiority trials. *N Engl J Med.* 2017;377:1357–67.
  31. Pocock SJ, Clayton TC, Stone GW. Challenging issues in clinical trial design: Part 4 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015;66:2886–98.
  32. Pocock SJ, McMurray JJV, Collier TJ. Statistical controversies in reporting of clinical trials: Part 2 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015;66:2648–62.
  33. Pocock SJ, Clayton TC, Stone GW. Design of major randomized trials: Part 3 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015;66:2757–66.
  34. Home P. Cardiovascular outcome trials of glucose-lowering medications: An update. *Diabetologia.* 2019;62:357–69.
  35. Aroda VR, Ahmann A, Cariou B, Chow F, Davies MJ, Jódar E, et al. Comparative efficacy, safety, and cardiovascular outcomes with once-weekly subcutaneous semaglutide in the treatment of type 2 diabetes: Insights from the SUSTAIN 1-7 trials. *Diabetes Metab.* 2019;45:409–18.
  36. McCrimmon RJ, Catarig AM, Frias JP, Lausvig NL, le Roux CW, Thielke D, et al. Effects of once-weekly semaglutide vs once-daily canagliflozin on body composition in type 2 diabetes: A substudy of the SUSTAIN 8 randomized controlled clinical trial. *Diabetologia.* 2020;63:473–85.
  37. Capehorn MS, Catarig AM, Furberg JK, Janez A, Price HC, Tadayon S, et al. Efficacy and safety of once-weekly semaglutide 1.0 mg vs once-daily liraglutide 1.2 mg as add-on to 1-3 oral antidiabetic drugs in subjects with type 2 diabetes (SUSTAIN 10). *Diabetes Metab.* 2020;46:100–9.
  38. Pocock SJ, McMurray JJ, Collier TJ. making sense of statistics in clinical trial reports: Part 1 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol.* 2015;66:2536–49.
  39. Gómez-Acebo I, Dierssen-Sotos T, Llorca J. [Number needed to treat: Interpretation and estimation in multivariable analyses and censored data]. *Med Clin (Barc).* 2014;142:451–6.
  40. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol.* 2013;13:92.
  41. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med.* 2007;357:2189–94.
  42. Giugliano D, Maiorino MI, Bellastella G, Longo M, Chiodini P, Esposito K. GLP-1 receptor agonists for prevention of cardiorenal outcomes in type 2 diabetes: An updated meta-analysis including the REWIND and PIONEER 6 trials. *Diabetes Obes Metab.* 2019;21:2576–80.
  43. Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med.* 2016;375:65–74.
  44. Pocock SJ, Collier TJ. Statistical appraisal of 6 recent clinical trials in cardiology: JACC state-of-the-art review. *J Am Coll Cardiol.* 2019;73:2740–55.
  45. Solomon SD, Jhund PS, Claggett BL, Dewan P, Køber L, Kosiborod MN, et al. Effect of dapagliflozin in patients with HFrEF treated with sacubitril/valsartan: The DAPA-HF trial. *JACC Heart Fail.* 2020;30254–7. S2213-1779.
  46. U.S. Food and Drug Administration. Type 2 Diabetes Mellitus: Evaluating the Safety of New Drugs for Improving Glycemic Control Guidance for Industry; 2020 [accessed 6 Jun 2020]. Available from: <https://www.fda.gov/media/135936/download>.
  47. Chong WH, Yanoff LB, Andraca-Carrera E, Thanh Hai M. Assessing the safety of glucose-lowering drugs—A new focus for the FDA. *N Engl J Med.* 2020;383:1199–202.
  48. Khunti K, Davies MJ, Marx M, Buse J, Draft FDA. Guidance for assessing the safety of glucose-lowering therapies: A missed opportunity. *Lancet Diabetes Endocrinol.* 2020;8:810–1.
  49. García Aller M. Lo imprevisible: Todo lo que la tecnología quiere y no puede controlar. *Planeta.* 2020.