

Exploratory analysis of comparative clinical trials used for marketing approval in patients with type 2 diabetes in Japan

SAGE Open Medicine
Volume 7: 1–9
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2050312118823407
journals.sagepub.com/home/smo



Reina Kaneko  and Shunsuke Ono

Abstract

Background/objectives: The results of phase 2 and 3 clinical trials, which justify decisions regarding marketing approval for new drugs, are used for comparison of drugs in the post-marketing phase. A number of meta-analyses of approved antidiabetics have been performed, but the heterogeneity of trials has not been fully examined. The aim of this study was to explore factors that may influence baseline HbA1c in trial samples and treatment outcomes (i.e. HbA1c reductions and effect sizes), with the goal of providing unbiased and fair retrospective comparisons between different antidiabetics.

Method: We conducted three meta-regression analyses using 78 randomized or non-randomized comparative phase 2 or 3 trials of 24 approved antidiabetics in Japan, conducted from 1987 to 2012.

Results: Baseline HbA1c of each arm was higher in phase 2 trials, trials with a greater number of subjects, trials with a lower proportion of male subjects, trials of combination therapy, or trials with longer subject disease duration. Entry criteria were different among drug classes and caused variations in baseline HbA1c. HbA1c reductions were larger in non-randomized trials, trials with a shorter treatment period, or trials with a lower proportion of male subjects. Effect sizes were larger in phase 2 trials, or trials of combination therapy. Larger effect sizes were observed in drugs with later market entry for alpha-glucosidase inhibitors and glinides.

Conclusion: Baseline HbA1c, an important characteristic of subjects enrolled in trials of antidiabetics, differed significantly across trials. Differences in features of study subjects were caused by explicit stipulations in eligibility criteria of HbA1c and also by other conditions (e.g. trial design, regulatory guidance, treatment guideline) and/or interventions of investigators and pharmaceutical companies that were specific to drugs and trials. Healthcare professionals should carefully consider these heterogeneities in trials used for marketing approval review when making a retrospective comparison to select the best treatment option for patients.

Keywords

Diabetes/endocrinology, clinical pharmacology, clinical trials, clinical trial design, drug development and approval, meta-regression analysis, statistics, type 2 diabetes

Date received: 22 August 2018; accepted: 10 December 2018

Introduction

The number of individuals diagnosed with diabetes mellitus (DM) or at high risk for DM has increased tremendously, to over 100 million in the United States in 2015 and about 20 million in Japan in 2016.^{1,2} DM is a major cause of death and is an established risk factor for coronary heart disease and ischemic stroke.³ In addition to diet and exercise regimens, optimal glycemic control using antidiabetic drugs is considered a key element in the treatment of DM. Several oral antidiabetics have been introduced since 1960s, with alleged therapeutic advantages over older drugs. However, several studies suggest difficulties in comparing efficacy

between old and new drugs in a “fair” manner, because patient backgrounds, clinical trial regulations, and treatment guidelines have changed over decades.^{4,5} Trial design and characteristics of study subjects were shown to have a significant impact on the results of pivotal trials of recently

Laboratory of Pharmaceutical Regulation and Sciences, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan

Corresponding author:

Shunsuke Ono, Laboratory of Pharmaceutical Regulation and Sciences, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
Email: shun-ono@mol.f.u-tokyo.ac.jp



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

approved drugs.⁶ Antihypertensive drugs, for example, showed time trends in baseline blood pressures (BP) and treatment effects.⁷ The observed trends, particularly those in baseline BP, seem to reflect historical changes in the target population in the past decades and indicate that heterogeneity in trial results should be considered in evaluating treatment effects of different drugs.

It is inevitable that pivotal phase 3 trials conducted in different settings entail a certain level of heterogeneity. These trials are used as the rationale for regulatory decisions regarding new drug approvals and eventually exploited in post-marketing for multiple purposes such as comparative analyses of treatment including economic evaluation, advertisements, and the rationale for treatment guidelines and/or algorithms.^{5,8} Irrespective of their critical roles in approval and post-marketing activities, however, the levels and causes of heterogeneity among the trials have not been fully investigated.

We aimed to explore factors that might influence baseline HbA1c, changes in HbA1c and treatment effect sizes (ESs). It is of importance to investigate all three dependent variables because baseline HbA1c is commonly included as a covariate in the primary analysis for pivotal studies and those variables are closely related to each other. We analyzed whether baseline and clinical outcomes of phase 2 or 3 clinical trials (i.e. changes in HbA1c and treatment ESs) used for marketing approval decisions were associated with characteristics of drugs, designs of trials, characteristics of enrolled subjects, or trial start year. Our study presented candidate factors to which attention should be paid in comparing drug efficacy and also provided a general description of features associated with successful comparative trials.

Methods

We gathered data on phase 2 and 3 trials conducted in Japan for all antidiabetics available on the market. A total of 26 drugs were identified. For two, trial data were unavailable in the public domain. Five multi-regional trials conducted for marketing approval in Japan were included. We used data from a Japanese population in those studies to eliminate any ethnic differences in the results.

The trial data were obtained from common technical documents (CTDs) on the Pharmaceutical and Medical Devices Agency website (<http://www.info.pmda.go.jp/>), “interview forms,” which provide healthcare professionals with comprehensive, specific, and practical guide for proper use of drugs, and publications on trials (Table S1). Baseline HbA1c and changes in HbA1c in each treatment arm were collected as explanatory and/or objective variables. National Glycohemoglobin Standardization Program (NGSP) values were used for our analyses; Japanese Diabetes Society values found in CTDs, interview forms and publications, were converted to NGSP values.⁹ We conducted regression analysis

using all the arms, as in previous studies.^{7,10} Studies were characterized by trial design, the entry criteria for HbA1c, mean age, disease duration, proportion of male subjects, and drug class.

We analyzed changes in HbA1c and ESs for each drug-placebo pair. Both are important outcomes used for various post-marketing comparisons, but the two have somewhat different practical implications. Changes in HbA1c are a direct measure of response to a drug; they reflect changes of a trial arm; they commonly appear in treatment guidelines and/or package insert of the drug. ESs, which are calculated based on changes in HbA1c and standard deviations, are a measure of how much a test drug is effective over a comparator/placebo in a comparative trial; they reflect differences between trial arms; they do not usually appear in guidelines or package inserts in their own, but in the form of various statistical measures including *p*-values because ESs form the basis of statistical analysis. We followed the method in previous studies to calculate ES using Hedges’ *g*.^{6,7} To obtain Hedges’ *g*, we divided the differences in HbA1c reductions between test drugs and placebo by the standard deviations. Negative ES indicated that the test drug was more effective than placebo. Two out of 78 trials did not report standard deviations in the publications; therefore, we imputed mean values (0.718) from other trials of drugs with the same mode of action.¹¹ We confirmed that analyses with and without the two trials lead to similar results.

Statistical analysis

We applied random-effects meta-regression models to account for both within-trial variances of patient backgrounds and treatment effects (weighted by the sample size of each trial), and the residual between-trial heterogeneity. The objective variable (*Y*) was baseline HbA1c, changes in HbA1c, or ES, depending on the models. We followed the same approach to establish the final meta-regression models, as reported recently⁷

$$Y_i = \beta_1 x_i^T + \beta_2 x_i^{ct} + \beta_3 x_i^{bl} + \beta_4 x_i^{ot} + u_i + \varepsilon_i,$$

where $u_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma_i^2)$

x_i^T : variables indicating trial start time; x_i^{ct} : variables indicating trial design; x_i^{bl} : variables indicating baseline HbA1c used in models except in case that baseline HbA1c was the objective variable; other exploratory variables (x_i^{ot}) indicating intrinsic features of test drugs or comparator drugs that includes placebo were used where applicable; u_i : random effects; i : trial arm or test drug-comparator pairs; and τ^2 : between-trial variance.

We categorized entry criteria based on lower and upper limitations of HbA1c: without criteria for a lower

Table 1. Summary statistics of explanatory variables and effect sizes for HbA1c in 78 clinical trials.

Variables	N (%) ^a	Mean (SD)
Trial start time (years)	176	2008.7 ^b (6.7)
Clinical trial design		
Ph2 dose-finding RCTs	43 (24.4)	–
Ph2 non-RCTs	4 (2.3)	–
Ph3 RCTs	123 (69.9)	–
Ph3 non-RCTs	6 (3.4)	–
Lower limit of baseline HbA1c		
No criteria	28 (15.9)	–
7.0% or less HbA1c level	119 (67.6)	–
Higher than 7.0% HbA1c level	29 (16.5)	–
Upper limit of baseline HbA1c		
No criteria	36 (20.45)	–
10.0% or less HbA1c level	55 (31.25)	–
Higher than 10.0% HbA1c level	85 (48.3)	–
Drug class of clinical trials		
Biguanides	4 (2.3)	–
Sulfonylureas	4 (2.3)	–
Alpha-glucosidase inhibitors	16 (9.1)	–
Glinides	16 (9.1)	–
Thiazolidinediones	12 (6.8)	–
DPP-4 inhibitors	90 (51.1)	–
SGLT2 inhibitors	34 (19.3)	–
Concomitant drug use		
Yes	76 (43.2)	–
No	100 (56.8)	–
Treatment period (weeks)	176	17.8 (10.8)
Number of Japanese subjects (patients)	176	75.70 (36.64)
Proportion of male subjects (%)	176	64.77 (9.35)
Mean age of subjects (years)	176	58.31 (2.37)
Subject disease duration (years)	176	7.63 (4.68)
Baseline HbA1c (%)	176	8.20 (0.59)
Changes in HbA1c (%)		
Biguanides	3 (1.7)	–1.07 (0.18)
Sulfonylureas	3 (1.7)	–1.43 (0.17)
Alpha-glucosidase inhibitors	14 (8.0)	–0.35 (0.26)
Glinides	9 (5.1)	–0.85 (0.32)
Thiazolidinediones	8 (4.55)	–1.24 (0.20)
DPP-4 inhibitors	49 (27.8)	–0.71 (0.18)
SGLT2 inhibitors	16 (9.1)	–0.63 (0.19)
Placebo	74 (42.05)	0.14 (0.20)
Effect size for HbA1c of placebo-controlled trial	74 ^c	–1.28 (0.43)

DPP-4: dipeptidyl peptidase-4; RCT: randomized controlled trial; SD: standard deviation; SGLT2: sodium glucose co-transporter 2.

^aTrial arm-unit data.

^bMedian.

^cTest drug and placebo pair-unit data.

limitation, 7.0% or less and higher than 7.0%; without criteria for an upper limitation, 10.0% or less and higher than 10.0%.

A rank correlation test was performed to investigate correlations between ES of drugs and the order of market entry for alpha-glucosidase inhibitors (aGI), glinides, dipeptidyl peptidase-4 (DPP-4) inhibitors, and sodium glucose co-transporter 2 (SGLT2) inhibitors.

The significance level was set at $p < 0.1$ as in previous studies.^{6,7} We performed analyses using Stata Statistical Software version 14 (StataCorp LLC, College Station, TX, USA).

Results

For 24 antidiabetics approved from 1961 to 2015 in Japan, 78 comparative phase 2 or 3 trials, conducted from 1987 to

Table 2. Meta-regression analysis of baseline HbA1c.

Exploratory variables	Model 1 Baseline HbA1c (%) (N = 176)			Model 1' Baseline HbA1c (%) (N = 176)		
	Coefficient	SE	p value	Coefficient	SE	p value
Trial start time (years)	0.0159	0.0112	0.160	0.0100	0.0098	0.307
Clinical trial design (baseline: Ph2 dose-finding RCTs)						
Ph2 non-RCTs	0.6732	0.2231	0.003***	0.1116	0.1319	0.399
Ph3 RCTs	-0.2492	0.0795	0.002***	-0.1872	0.0437	<0.001***
Ph3 non-RCTs	-0.0264	0.1804	0.884	-0.2652	0.0917	0.004***
Number of Japanese subjects per arm (patients)	0.0015	0.00090	0.095*	0.0016	0.00045	0.001***
Concomitant drug use (1: Yes, 0: No)	0.2614	0.0672	<0.001***	0.1846	0.0375	<0.001***
Proportion of male subjects (%)	-0.0130	0.0044	0.004***	-0.0022	0.0023	0.344
Mean age of subjects (years)	0.0141	0.0160	0.379	0.0048	0.0085	0.571
Subject disease duration (years)	0.0149	0.0068	0.030**	0.0073	0.0034	0.032**
Enrolled subjects						
Lower limit of baseline HbA1c (baseline: no criteria for lower limit of baseline HbA1c)						
7.0% or less HbA1c level	-0.8353	0.1966	<0.001***	-0.1930	0.1584	0.225
Higher than 7.0% HbA1c level	-0.3900	0.2002	0.053*	0.2199	0.1668	0.189
Upper limit of baseline HbA1c (baseline: no criteria for upper limit of baseline HbA1c)						
10.0% or less HbA1c level	-0.1589	0.1837	0.388	0.1271	0.1188	0.286
Higher than 10.0% HbA1c level	-0.0564	0.1711	0.742	0.1854	0.1237	0.136
Drug class of clinical trials (baseline: clinical trials for alpha-glucosidase inhibitors)						
Biguanides				-0.4339	0.1680	0.011**
Sulfonylureas				0.3636	0.1879	0.055*
Glinides				-0.5603	0.1108	<0.001***
Thiazolidinediones				1.6685	0.1154	<0.001***
DPP-4 inhibitors				-0.3460	0.1429	0.017**
SGLT2 inhibitors				-0.3718	0.1516	0.015**
Constant	-23.0983	22.3396	0.303	-11.9994	19.3386	0.536

DPP-4: dipeptidyl peptidase-4; RCT: randomized controlled trial; SE: standard error; SGLT2: sodium glucose co-transporter 2.

The objective variable was baseline HbA1c (%).

** $p < 0.05$, * $p < 0.1$.

2012, with 176 treatment arms were used for the analyses (Figure S1). Explanatory variables and ES for HbA1c in 78 clinical trials were summarized in Table 1. Baseline HbA1c for an arm ranged from 6.91% to 10.78%, and the average baseline was 8.20%. Changes in HbA1c in response to active drug treatment ranged from -1.63% to 0.16%, and the average change was -0.74%. There were significant differences in baseline HbA1c (including placebo, $p < 0.0001$ in one-way analyses of variance (ANOVA)) and changes in HbA1c (except the placebo arms, $p < 0.0001$) among different drug classes. There were 74 pairs of a test drug and placebo. The ES for HbA1c of placebo-controlled trials ranged from -2.31 to -0.18 with a mean of -1.28 and differed significantly across test drug-placebo pairs ($p < 0.0001$).

Baseline HbA1c

The results of regression analysis regarding baseline HbA1c are summarized in Table 2. Phase 2 dose-finding randomized controlled trials (RCTs) showed higher baseline compared to phase 3 RCTs (Coefficient: -0.2492 in

Model 1 and -0.1872 in Model 1'). Baseline was higher in trials of combination therapy than in trials of monotherapy (0.2614 in Model 1 and 0.1846 in Model 1') or in trials including a larger number of Japanese subjects (0.0015 in Model 1 and 0.0016 in Model 1'). Subjects' disease duration was positively associated with baseline when we did or did not adjust for variables representing drug class (0.0149 in Model 1 and 0.0073 in Model 1'), while proportion of male subjects was negatively associated with baseline when we did not adjust for drug class (-0.0130 in Model 1). Trials without entry criteria for a lower limit of baseline HbA1c showed higher baselines compared to trials with 7.0% or less HbA1c level and higher than 7.0% HbA1c level entry criteria (-0.8353 and -0.3900 in Model 1). Trials of thiazolidinediones and sulfonylureas showed higher baselines (0.3636 and 1.6685 in Model 1'), and trials of biguanides, glinides, DPP-4 inhibitors, and SGLT2 inhibitors showed lower baselines compared to trials of aGI (-0.4339, -0.5603, -0.3460, and -0.3718 in Model 1'). The heterogeneity (I^2) between the trials outcomes was 88.0% (Model 1) and 71.0% (Model 1').

Table 3. Meta-regression analysis of changes in HbA1c.

Exploratory variables	Model 2 Changes in HbA1c (%) (N=176)		
	Coefficient	SE	p value
Trial start time (years)	0.0083	0.0076	0.276
Clinical trial design (baseline: Ph2 dose-finding RCTs)			
Ph2 non-RCTs	-0.5852	0.2290	0.012**
Ph3 RCTs	-0.0358	0.0470	0.447
Ph3 non-RCTs	-0.3516	0.1679	0.038**
Number of Japanese subjects per arm (patients)	0.00038	0.00049	0.438
Concomitant drug use (1: Yes, 0: No)	-0.0175	0.0373	0.639
Treatment period (weeks)	0.0075	0.0035	0.034**
Proportion of male subjects (%)	0.0039	0.0023	0.088*
Mean age of subjects (years)	-0.0040	0.0083	0.635
Subject disease duration (years)	-0.0027	0.0035	0.434
Baseline HbA1c (%)	-0.0057	0.0522	0.913
Enrolled subjects			
Lower limit of baseline HbA1c (baseline: no criteria for lower limit of baseline HbA1c)			
7.0% or less HbA1c level	-0.0846	0.1246	0.499
Higher than 7.0% HbA1c level	-0.1019	0.1277	0.426
Upper limit of baseline HbA1c (baseline: no criteria for upper limit of baseline HbA1c)			
10.0% or less HbA1c level	-0.1391	0.0961	0.150
Higher than 10.0% HbA1c level	-0.0645	0.0872	0.461
Drug class per arm (baseline: placebo)			
Biguanides	-1.3260	0.1308	<0.001***
Sulfonylureas	-1.4323	0.1858	<0.001***
Alpha-glucosidase inhibitors	-0.4516	0.0636	<0.001***
Glinides	-0.9548	0.0761	<0.001***
Thiazolidinediones	-1.1810	0.1426	<0.001***
DPP-4 inhibitors	-0.8494	0.0388	<0.001***
SGLT2 inhibitors	-0.8148	0.0569	<0.001***
Constant	-16.3938	14.9761	0.275

DPP-4: dipeptidyl peptidase-4; RCT: randomized controlled trial; SE: standard error; SGLT2: sodium glucose co-transporter 2.

The objective variable was changes in HbA1c (%).

** $p < 0.05$, * $p < 0.1$.

Treatment effects of drugs (HbA1c changes and ESs)

Absolute changes in HbA1c after drug treatment were associated with some variables related to study design and subject profile (Table 3). Phase 2 and phase 3 non-RCTs showed larger HbA1c reduction compared to phase 2 dose-finding RCTs (Coefficient: -0.5852 and -0.3516 in Model 2). Treatment period and proportion of male subjects were positively associated with HbA1c changes (i.e. smaller reduction; 0.0075 and 0.0039 in Model 2).

The ES of placebo-controlled trials were also related to some variables (Table 4). Phase 2 dose-finding RCTs showed larger ES compared to phase 3 RCTs (Coefficient: 0.2280 in Model 3). Concomitant drug use including insulin combination therapy in three trials was negatively associated with ES (i.e. larger treatment effect in combination therapy than monotherapy; -0.2223 in Model 3). Trials of

sulfonylureas and thiazolidinediones showed larger treatment effects compared to trials of aGIs (-1.3726 and -0.9084 in Model 3).

Neither baseline HbA1c nor subject entry criteria were associated with HbA1c changes or ES. Trial start time indicating the time trend variable did not show statistical significance in any regression analyses, although it was negatively correlated with baseline HbA1c and ES, and positively correlated with absolute changes (data not shown). The heterogeneity (I^2) between the trials outcomes was 82.7% (Model 2) and 60.1% (Model 3).

Rank correlations of ESs

Rank correlations of ES by the order of market entry were tested for certain drug class where the number of drugs was three or more. Significant negative rank correlations were shown in aGIs and glinides (i.e. larger ES for later market

Table 4. Meta-regression analysis of effect sizes for HbA1c.

Exploratory variables	Model 3 Effect size for HbA1c (N=74) ^a		
	Coefficient	SE	p value
Trial start time (years)	-0.0076	0.0265	0.777
Clinical trial design (baseline: Ph2 dose-finding RCTs)			
Ph3 RCTs	0.2280	0.1268	0.078*
Ph3 non-RCTs	Dropped	–	–
Number of Japanese subjects per trial (patients)	0.000041	0.00085	0.962
Concomitant drug use (1: Yes, 0: No)	-0.2223	0.1085	0.045**
Treatment period (weeks)	0.0095	0.0101	0.351
Proportion of male subjects (%)	-0.0077	0.0071	0.537
Mean age of subjects (years)	-0.0044	0.0232	0.743
Subject disease duration (years)	0.0040	0.0115	0.732
Baseline HbA1c (%)	0.2260	0.2124	0.292
Enrolled subjects			
Lower limit of baseline HbA1c (baseline: no criteria for lower limit of baseline HbA1c)			
7.0% or less HbA1c level	0.0672	0.4190	0.873
Higher than 7.0% HbA1c level	-0.1794	0.4215	0.672
Upper limit of baseline HbA1c (baseline: no criteria for upper limit of baseline HbA1c)			
10.0% or less HbA1c level	-0.4704	0.4014	0.246
Higher than 10.0% HbA1c level	-0.5898	0.4071	0.153
Test drug-comparator pairs ^b (baseline: alpha-glucosidase inhibitors)			
Biguanides	-0.3033	0.4949	0.543
Sulfonylureas	-1.3726	0.5338	0.013**
Glinides	-0.0043	0.3999	0.991
Thiazolidinediones	-0.9084	0.4023	0.028**
DPP-4 inhibitors	-0.0511	0.4663	0.913
SGLT2 inhibitors	0.0032	0.4805	0.995
Constant	13.0740	52.1658	0.803

DPP-4: dipeptidyl peptidase-4; RCT: randomized controlled trial; SE: standard error; SGLT2: sodium glucose co-transporter 2.

The objective variable was effect size for HbA1c.

^aPlacebo-controlled trials were used for this analysis.

^bTest drug × placebo.

** $p < 0.01$, * $p < 0.05$, $p < 0.1$.

entry; Kendall's tau-b = -0.539, $p = 0.0453$ for aGIs; Kendall's tau-b = -0.725, $p = 0.0060$ for glinides), but not for DPP-4 inhibitors or SGLT2 inhibitors (Kendall's tau-b = 0.173, $p = 0.1092$ for DPP-4 inhibitors, Kendall's tau-b = 0.346, $p = 0.0627$ for SGLT2 inhibitors).

Discussion

We explored how the clinical trial results of antidiabetics were associated with trial backgrounds and subjects' characteristics. We focused on phase 2 and 3 trials that were submitted for regulatory approval and commonly used as clinical evidence of efficacy in marketing activities. Our analysis clarified that characteristics of study subjects were determined by stipulations in the protocol and other conditions specific to drugs and trials and that trial results were associated with such characteristics of enrolled subjects and trial design components. The associations between trial results

and demographics of enrolled subjects or trial design components suggest that the "observed" treatment ES does not solely depend on the efficacy of a drug per se, but also on other factors including study conditions that can be determined by sponsors or investigators, study subjects actually enrolled, clinical trial regulations, and disease treatment guidelines at the time of a trial reflecting public health needs regarding the underlying DM population, as reported in previous studies for DM and also for antihypertensives.^{7,12}

We investigated whether study subjects enrolled in clinical trials have changed over time. The results of this analysis indicated that the mean baseline HbA1c of a study arm was higher and showed considerably more variation in the 1990s than in the 2000s (without adjustment for other variables). This implies that study subjects were different between studies at least in baseline HbA1c, one of the most important patient characteristics in DM. Baseline was significantly associated with entry criteria in relation to HbA1c level and

study design, as explained later, and a time trend was not observed in regression analysis (Table 2). These results suggest that study subjects were selected based on eligibility criteria in each protocol, which seem to reflect pertinent diagnosis guidelines, therapeutic interventions, and regulations at the time of a trial.

In this regard, it is also of interest to examine the extent to which subjects in clinical trials were similar to real-world patients. According to a cross-sectional study and the Japan National Diabetes Surveys, mean HbA1c increased from 1997 to 2002 and decreased from 2002 to 2011 in the general population of patients with type 2 DM.^{13,14} Observed changes in average HbA1c levels in the trials in our study did not contradict the overall trend in patients, suggesting that the study results may be extrapolatable for a specific time, but not for all periods.

Baseline HbA1c in phase 3 non-RCTs/RCTs was statistically lower than that in phase 2 RCTs (Table 2), but baseline HbA1c was unexpectedly not associated with HbA1c reduction or ES (Tables 3 and 4). Results of phase 3 RCTs were somewhat different from those in phase 2 and/or non-RCTs; HbA1c reductions were smaller in RCTs than those in non-RCTs (Table 3), and ES were smaller in phase 3 RCTs than those in phase 2 RCTs (Table 4). Smaller effectiveness outcomes in phase 3 might be explained in part by “regression to the mean,” because our data reflect only successful development pathways. However, when considering the clear differences in baseline between phase 2 and 3, as well as in outcomes between RCTs and non-RCTs, we surmise that study subjects differed somewhat between different phases and objectives. A possible explanation is that patients were selected more carefully in phase 3 RCTs, probably with the aim of creating homogeneous samples with small variances, rather than samples highly responsive to test drugs.

Baseline was higher in subjects enrolled in trials of combination therapy, and ES were also larger in such trials, even when baseline level was adjusted (Tables 2 and 4).^{14,15} These results are consistent with previous findings that combination therapies showed larger effects in reducing the HbA1c level than did monotherapy for patients with severe conditions in whom monotherapy was unsuccessful.¹⁶ Trials with longer treatment periods demonstrated less absolute changes in HbA1c (Table 3), which is in line with previous studies reporting a progressive deterioration of diabetes control and a cumulative incidence of failure of monotherapy with long-term antidiabetic treatment.^{17,18}

Demographic features of subjects were associated with baseline HbA1c and treatment effects in complicated ways. Proportion of male subjects was negatively associated with baseline (Model 1 in Table 2), but the association disappeared when the drug class variables were included (Model 1' in Table 2). In terms of treatment effects, HbA1c reductions were smaller in trials with a higher male proportion, but ES were not (Tables 3 and 4). A model without baseline yielded similar results (data not shown). These results indicate that sex

ratio may have some impact on HbA1c reductions, but the size of the possible impact appears minimal when trial backgrounds are adjusted. Previous studies report that mean HbA1c value at study enrollment was significantly higher in Japanese women with type 2 diabetes than in men,^{13,19,20} and our findings without adjustment were consistent with these results. In the entire Japanese population (i.e. healthy adults included), baseline was higher for men than for women.² It has been reported that women with type 2 diabetes are less likely to reach the goals for recommended HbA1c levels compared to men.^{21–25} Female patients with higher baseline need larger HbA1c reductions, but adverse events and/or intrinsic hormonal circumstances sometimes prevent achievement of therapeutic goals.²⁶

Mean age of subjects was not associated with baseline or treatment effects. Disease duration of subjects, one of the severity measures for DM, showed a positive association with HbA1c baseline, but did not show an association with treatment effects (Tables 2 and 3). Other studies also reported that disease duration of subjects was associated with HbA1c baseline but not with treatment effects.^{12,27}

Our results indicated that different patient entry criteria led to different study samples in terms of baseline HbA1c. Pre-specified lower limits of HbA1c level in entry criteria seemed to enhance the enrollment of patients with poor glycemic control and lead to lower baseline levels in study samples, as was intended (Model 1 in Table 2). However, the association disappeared when the drug class variables were added (Model 1' in Table 2). This probably reflects the fact that there are differences (at least) in entry criteria among trial protocols of different drug classes. In contrast, upper limits in entry criteria were not associated with observed baselines. The sponsor and investigators choose appropriate inclusion/exclusion criteria to achieve their study objective, but their choices have been constrained inevitably by the pertinent diagnostic and therapeutic guidelines.^{5,28} Sponsors and investigators should also follow domestic and international drug development guidelines, for trials to be submitted for New Drug Application (NDA) purposes.⁴ Although all trials in our analyses were used to obtain regulatory approval in general type 2 diabetes populations, the entry criteria were not consistent because applicable guidelines were not the same during the period. Prior to 2000, few trials had entry criteria for HbA1c level, while almost all trials had some entry criteria for HbA1c after 2000. These historical changes were inevitable as a response to changes in the requirements for clinical development. However, in comparison of trial results, it is essential to pay attention to what criteria were applied to specific trials, because such factors could lead to heterogeneity in sample populations.

Average baseline HbA1c in each arm was not related to treatment effects in our meta-regression models. The correlation between average baseline and HbA1c reduction was weak ($r=-0.146$, $p<0.1$). These findings were not consistent with the results of previous studies reporting significant

correlations between baseline level and HbA1c changes with treatment.^{10,29,30} The inconsistency seems to stem from differences in controlling variables adopted in analysis; owing to the exploratory nature of this research, we used several variables showing study background and characteristics of drugs and patients that may be associated with both baseline and treatment outcomes. These control variables were not included in previous studies, however, because they focused more on the therapeutic implications for patients and not necessarily on possible associations between study background and outcomes in general.

Differences in treatment effects among drug classes were clearly observed in this analysis. These differences reflect not only a drug's effectiveness per se, but all other factors that could affect how clinical trials are designed and implemented. Drug class was significantly associated with baseline HbA1c level even when critical entry criteria (i.e. upper and lower limits of HbA1c) and some patient backgrounds were controlled (Table 2). The observed imbalance in baseline occurred before the administration of drugs and was not attributed to the effectiveness of drugs, which apparently implies heterogeneous study subjects among trials with respect to baseline, at least. Heterogeneity in study samples may reflect the competitive environment of trials. We found significant rank correlations of ES by the order of market entry for aGIs and glinides (Kendall's tau-b = -0.539, $p=0.0453$ for aGIs; Kendall's tau-b = -0.725, $p=0.0060$ for glinides), but not for DPP-4 inhibitors or SGLT2 inhibitors. Approval lags between the first market entry to the last for aGIs and glinides were more than 10 years, while those for DPP-4 inhibitors and SGLT2 inhibitors were 6 years and 1 year, respectively. Different drug development strategies might have been implemented between them; the former drugs were developed as a possibly best-in-class drug aiming to demonstrate superiority to their predecessors, while the latter were developed as a first-in-class drug applying a highly competitive development approach.

One implication of this study is that treatment outcomes depend on the intention, implementation, and environment of trials in DM, as observed in other therapeutic fields. Trial design components that were declared in the protocol prior to recruitment, including eligibility criteria, treatment period, and the usage of concomitant medications (i.e. monotherapy or combination therapy), would affect results. In addition, we confirmed that trial results might be influenced by patient demographics in each trial. Investigators and pharmaceutical companies can also intervene in the actual process of subject enrollment to some extent. For example, they can enroll patients for whom preferable outcomes would be expected as long as eligibility criteria are satisfied and an apparent imbalance in patient characteristics does not occur. Heterogeneity in clinical trials, irrespective of whether it is intended, is a long-standing issue related to the extent to which trial results

can be extrapolated. The possible impact of heterogeneity should be carefully considered when trial results are retrospectively compared.

This study had limitations as follows: our findings are based solely on the results of phase 2 and 3 trials of approved antidiabetics and cannot automatically be extrapolated to antidiabetic trials in general. Results for unapproved drugs are not considered in this research. Our regression analysis was for exploratory purposes with only a small number of variables, and it would thus be inappropriate to infer any definitive causal links. We collected only data presented in publications; undisclosed information may affect our conclusion, for instance, the concomitant medications for complications such as statins and thiazides which may have effect on HbA1c level during the trial.^{31,32}

Conclusion

Baseline HbA1c level of study subjects was significantly different among trials of antidiabetic drugs, which reflected the fact that eligibility criteria had been changing over time. Changes in HbA1c were related to proportion of male subjects enrolled and treatment period. Study sample characteristics, trial status in clinical development, and trial design features were associated with observed treatment effects. Our results suggested that differences among study subjects were caused by explicit stipulations in eligibility criteria as well as by other conditions and/or interventions that were specific to drugs. Such differences directly or indirectly affected trial results and eventually statements on efficacy in guidelines, package inserts, and/or advertising materials. Impact of heterogeneities in trials should be considered in the retrospective evaluation of therapeutic outcomes, especially when medical professionals compare treatment effects of different antidiabetic drugs and select the best one for patients.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: R.K. is an employee of Daiichi Sankyo Co., Ltd, Tokyo, Japan. S.O. declared no potential conflict of interest in preparing this article.

Ethical approval

Ethical approval was not sought for this study because it is not a research involving human participants.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by a Japanese government-based grant-in-aid from the Ministry of Education, Culture, Sports, Science and Technology, Tokyo, Japan (Grant KAKENHI: 26460215).

Informed consent

Informed consent was not sought for this study because it is not a research involving human participants.

Supplemental material

Supplemental material for this article is available online.

ORCID iD

Reina Kaneko  <https://orcid.org/0000-0002-3557-1423>

References

- National Center for Chronic Disease Prevention and Health Promotion. National diabetes statistics report, 2017, <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf> (accessed 7 July 2018).
- Ministry of Health, Labour and Welfare. National health and nutrition survey [in Japanese], <http://www.mhlw.go.jp/toukei/itiran/gaiyo/k-eisei.html> (accessed 7 July 2018).
- The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* 2010; 375: 2215–2222.
- Ministry of Health, Labour and Welfare. On release of the guideline for clinical evaluation of oral hypoglycemic agents, <https://www.pmda.go.jp/files/000208194.pdf> (accessed 7 July 2018).
- The Japan Diabetes Society. *Treatment guide for diabetes 2018–2019* [in Japanese]. Tokyo, Japan: Bunkodo, 2018.
- Fukunaga S, Kusama M and Ono S. The effect size, study design, and development experience in commercially sponsored studies for new drug applications in approved drugs. *Springerplus* 2014; 3: 740.
- Kaneko R, Sano K and Ono S. Heterogeneity of clinical trials for antihypertensive drugs in Japan: exploratory analysis of confirmatory phase III trials used for marketing approval. *Clin Pharmacol Ther* 2018; 104: 120–129.
- Upadhyay J, Polyzos SA, Perakakis N, et al. Pharmacotherapy of type 2 diabetes: an update. *Metabolism* 2018; 78: 13–42.
- Kashiwagi A, Kasuga M, Araki E, et al. International clinical harmonization of glycated hemoglobin in Japan: from Japan Diabetes Society to National Glycohemoglobin Standardization Program values. *Diabetol Int* 2012; 3: 8–10.
- Esposito K, Chiodini P, Bellastella G, et al. Proportion of patients at HbA1c target <7% with eight classes of antidiabetic drugs in type 2 diabetes: systematic review of 218 randomized controlled trials with 78 945 patients. *Diabetes Obes Metab* 2012; 14: 228–233.
- Donders AR, van der Heijden GJ, Stijnen T, et al. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; 59: 1087–1091.
- Sherifali D, Nerenberg K, Pullenayegum E, et al. The effect of oral antidiabetic agents on A1C levels: a systematic review and meta-analysis. *Diabetes Care* 2010; 33: 1859–1864.
- Yamamoto Y, Nakagami T, Yokoyama T, et al. Temporal changes in mean HbA1c in a general population: comparisons between 1997 and 2002, the Japan National Diabetes Surveys. *Diabetol Int* 2015; 6: 83–90.
- Oishi M, Yamazaki K, Okuguchi F, et al. Changes in oral antidiabetic prescriptions and improved glycemic control during the years 2002–2011 in Japan (JDDM32). *J Diabetes Investig* 2014; 5: 581–587.
- Kobayashi M, Yamazaki K, Hirao K, et al. The status of diabetes control and antidiabetic drug therapy in Japan—a cross-sectional survey of 17,000 patients with diabetes mellitus (JDDM 1). *Diabetes Res Clin Pract* 2006; 73: 198–204.
- Bolen S, Feldman L, Vassy J, et al. Systematic review: comparative effectiveness and safety of oral medications for type 2 diabetes mellitus. *Ann Intern Med* 2007; 147: 386–399.
- Turner RC, Cull CA, Frighi V, et al. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). *JAMA* 1999; 281: 2005–2012.
- Kahn SE, Haffner SM, Heise MA, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med* 2006; 355: 2427–2443.
- Tajima N, Nishimura R, Izumi K, et al. A large-scale, observational study to investigate the current status of diabetes complications and their prevention in Japan: research outline and baseline data for type 2 diabetes—JDCP study 1. *Diabetol Int* 2015; 6: 243–251.
- Yokomichi H, Nagai A, Hirata M, et al. Serum glucose, cholesterol and blood pressure levels in Japanese type 1 and 2 diabetic patients: BioBank Japan. *J Epidemiol* 2017; 27: S92–S97.
- Wexler DJ, Grant RW, Meigs JB, et al. Sex disparities in treatment of cardiac risk factors in patients with type 2 diabetes. *Diabetes Care* 2005; 28: 514–520.
- Tang YH, Pang SM, Chan MF, et al. Health literacy, complication awareness, and diabetic control in patients with type 2 diabetes mellitus. *J Adv Nurs* 2008; 62: 74–83.
- Aryan S and Itamar R. Gender-specific care of diabetes mellitus: particular considerations in the management of diabetic women. *Diabetes Obes Metab* 2008; 10: 1135–1156.
- McGill JB, Vlajnic A, Knutsen PG, et al. Effect of gender on treatment outcomes in type 2 diabetes mellitus. *Diabetes Res Clin Pract* 2013; 102: 167–174.
- Arnetz L, Ekberg NR and Alvarsson M. Sex differences in type 2 diabetes: focus on disease course and outcomes. *Diabetes Metab Syndr Obes* 2014; 7: 409–420.
- Franconi F and Campesi I. Sex and gender influences on pharmacological response: an overview. *Expert Rev Clin Pharmacol* 2014; 7: 469–485.
- Blaum CS, Velez L, Hiss RG, et al. Characteristics related to poor glycemic control in NIDDM patients in community practice. *Diabetes Care* 1997; 20: 7–11.
- American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. *Diabetes Care* 2018; 41: S13–S27.
- Bloomgarden ZT, Dodis R, Viscoli CM, et al. Lower baseline glycemia reduces apparent oral agent glucose-lowering efficacy: a meta-regression analysis. *Diabetes Care* 2006; 29: 2137–2139.
- DeFronzo RA, Stonehouse AH, Han J, et al. Relationship of baseline HbA1c and efficacy of current glucose-lowering therapies: a meta-analysis of randomized clinical trials. *Diabetic Med* 2010; 27: 309–317.
- The Japan Diabetes Society. *Training guidebook for medical specialist in diabetes*. 7th ed. Tokyo, Japan: SHINDAN TO CHIRYO SHA, Inc., 2017.
- Kadowaki T, Araki E, Inagaki N, et al. *The diabetes: basic and clinical*. 1st ed. Tokyo, Japan: Nishimura Co., Ltd., 2015.