Application and impact of run-in studies

Michael Fralick, MD, MSc[1,2,3], Jerry Avorn, MD[1,3], Jessica M. Franklin, PhD[1,3], Abdurrahman Abdurrob, BS[3], and Aaron S. Kesselheim, MD, JD, MPH[1,3]

[1]Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [2]Eliot Phillipson Clinician-Scientist Training Program, University of Toronto, Toronto, ON, Canada; [3]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

BACKGROUND: A run-in phase is often employed prior to randomization in a clinical trial to exclude non-adherent patients, placebo responders, active drug non-responders, or patients who do not tolerate the active drug. This may impact the generalizability of trial results.

OBJECTIVE: To determine if clinical outcomes differed between randomized controlled trials with run-in phases compared with randomized controlled trials of the same medication without run-in phases.

DESIGN, PARTICIPANTS: From 2006 to 2014, the Food and Drug Administration approved 258 new medications. Sitaglitpin, saxagliptin, linagliptin, and alogliptin were among the only drugs with a common mechanism of action that each had multiple clinical trials, some of which had run-in phases and some of which did not. We identified all published randomized controlled trials for these four medications from MEDLINE and EMBASE as well as prior systematic reviews.

MAIN MEASURES: We extracted key measures of medication efficacy (reduction in hemoglobin A1C) and safety (serious adverse events) from qualifying trials. Study results were pooled for each medication using random effects meta-analysis.

KEY RESULTS: We identified 106 qualifying trials for DPP4 inhibitors, of which 88 had run-in phases and 18 did not. The average run-in phase duration was 4.0 weeks (range 1–21), and 73% of run-in phases administered placebo rather than active drug. The reduction in hemoglobin A1C compared to baseline was similar for trials with and without run-in phases (0.70%, 95% confidence interval [CI] 0.65–0.75 vs 0.76%, 95% CI 0.69–0.84, $p = 0.27$). The proportion of patients with serious adverse events was also similar for trials with and without run-in phases (4%, 95% CI: 3–5% vs 3%, 95% CI: 1–4%, $p = 0.35$).

CONCLUSION: Trials with run-in phases provided similar estimates for medication efficacy and safety compared to trials without run-in phases. Because run-in phases are costly and time-consuming, these results call their utility into question for clinical trials of short duration.

KEY WORDS: run-in; lead-in; clinical trial; study design.

Randomized controlled trials (RCTs) remain the gold standard for assessing medication safety and efficacy.[1,2] Some RCTs include a pre-randomization run-in period, often justified as a way of preserving internal validity.[3] The run-in phase involves administration of a drug or placebo and generally serves to exclude non-adherent patients, placebo responders, active drug non-responders, or patients who do not tolerate the active drug.[3,4] Run-in phases can identify patients who are more likely to adhere to medications, which could improve the efficiency and statistical power of the subsequent trial.[5,6]

However, run-in phases are expensive, time-consuming, and may reduce the generalizability of the trial results.[5,6] There is little published evidence on their contribution to the clinical trial enterprise, and few analyses have focused on run-in phases specifically.[3] In one notable example, investigators examined two trials of the impact of primary prophylaxis with aspirin on the rate of myocardial infarction. In the Physicians' Health Study, which was conducted with a run-in phase, primary prophylaxis with aspirin was found to decrease the rate of myocardial infarction (relative risk, 0.56; 95% confidence interval [CI] 0.45–0.70).[7] By contrast, in the British Physicians' Study, which was conducted without a run-in, the relative risk was 0.97 (95% CI, 0.73–1.24).[8]

The inclusion of run-in phases in clinical trials has important implications for the trials' outcomes. The Food and Drug Administration (FDA) assesses new medications by weighing the benefits and risks demonstrated in the pivotal clinical trials.[9] If trials with run-in phases lead to highly selected patient populations, this could result in overestimates of a drug's benefits and underestimates of its harms relative to trials without run-in phases. Understanding the effect of run-in phases could therefore aid regulatory evaluation of new drugs and may help the FDA decide whether to request heightened post-marketing surveillance or additional post-approval studies for medications approved based on RCTs with run-in phases. Conversely, if trials with run-in phases provide similar results to trials without run-in phases, then the value of run-ins might be questioned. We identified a class of drugs that has been widely studied both with and without run-in phases and assessed for differences in the observed benefits and harms reported in those trials.

METHODS

Data sources

To identify an experimental setting for our study, we systematically reviewed new drugs and biologics approved between 2006 and 2014 using the publicly available FDA Approved Drug Products database.[10] We then determined which of these drugs had been evaluated by clinical trials with and without run-in phases using a two-step process. First, we reviewed the Clinical Studies section of each drug's FDA summary basis of approval document.[10] Second, we searched MEDLINE to find RCTs involving the drug and read the methods section and appendix of each retrieved trial. Less than 5% of the 258 medications we reviewed had a combination of trials with and without run-in phases. Sitagliptin (approved in 2006), saxagliptin (2009), linagliptin (2011), and alogliptin (2013), all of which are dipeptidyl peptidase-4 (DPP4) inhibitors used for the treatment of type 2 diabetes mellitus, were some of the few medications that had multiple clinical trials with and without run-in phases.

To identify relevant RCTs relating to these DPP4 inhibitors, we compared our list of trials to the references from published meta-analyses for each drug. We then conducted a systematic review of MEDLINE and EMBASE to identify trials published after the meta-analyses (January 2015 through October 2016). Our search term was the medication's name as a MeSH heading and we limited our results to randomized controlled trials. We included only trials with a primary endpoint of change in hemoglobin A1C. We excluded phase 1, phase 2, and post-hoc studies as well articles not written in English. We also excluded conference abstracts, because we required detailed data on patient demographics, study design, and trial results.

Data extraction

An initial data collection tool was piloted and revised using ten articles. The following categories of data were extracted from each article independently by two investigators (MF, AA): study design, trial-level data, patient-level data, run-in phase characteristics, and trial results. Differences were resolved by consensus.

Study design data included blinding, concealment, method of analysis (e.g., intention-to-treat), and duration of follow-up. Trial-level data included number of patients screened for inclusion, number of patients included, and the number of patients who completed the trial. Patient-level data included age, sex, weight, baseline hemoglobin A1C, use of other diabetes medications, and use of insulin. Run-in phase characteristics included run-in product (e.g., drug, placebo), duration of run-in phase, word used to describe the run-in phase (e.g., run-in, lead-in), and reasons for exclusion during or after the run-in phase. Trial results included drop-out rate, measures of drug efficacy, and measures of drug safety. Our measure of drug efficacy was the reduction in hemoglobin A1C at the time of the study's primary assessment (the FDA's standard measure of drug efficacy for diabetes medications), as compared with the patient's baseline/pre-trial hemoglobin A1C. Drug safety data included hypoglycemia, severe hypoglycemia, frequency of the most common adverse event reported, and serious adverse events.

Statistical analysis

The primary analysis compared reported drug efficacy and safety for studies with run-in phases compared to studies without run-in phases. Efficacy results from each trial for each drug were pooled individually using a random-effects meta-analysis model.[11] This provided point estimates and 95% CIs for the reduction in hemoglobin A1C reported for each DPP4 inhibitor. For drug safety, the proportion of patients experiencing a serious adverse event, hypoglycemia, and the most common adverse event in trials with run-in phases compared to trials without run-in phases were meta-analyzed using the Clopper-Pearson interval to provide exact 95% CIs.[12]

To identify study-level predictors for the evaluated efficacy and safety endpoints, meta-regression was performed using the following variables (selected a priori): duration of the primary endpoint, baseline hemoglobin A1C, inclusion of a run-in phase, mean age, and percentage of male patients.

Descriptive statistics were used to characterize patient-level characteristics, study-level characteristics, and trial-level characteristics. Categorical data were compared with chi-square test, and continuous data were compared with the student t test with unequal variance. Statistical analyses were performed using Stata IC version 14.2 (StataCorp LP, College Station, Texas).

Patient involvement

Patients were not involved in the design or implementation of our study.

Role of the funding source

The funding sources had no role in the study design; the collection, analysis, and interpretation of data; the writing of the report; and the decision to submit the paper for publication.

RESULTS

We identified 106 randomized trials of DPP4 inhibitors, 88 with run-in phases and 18 without run-in phases. Characteristics of patients enrolled and randomized in trials with or without run-in phases were similar (Table 1). Demographic characteristics of patients who entered the run-in phases, but were excluded and not randomized, were not reported in the manuscripts or appendices of any article. Trials with run-in phases were more likely to be double-blind and to have concealed allocation. They also screened about twice as many patients for inclusion compared to trials without run-in phases (Table 1). The percentage of patients completing trials with or

Table 1 Trial and patient-level characteristics

	Run-in (N = 88)	No run-in (N = 18)	p
Age in years, mean (SD)	56.6 (4.9)	56.0 (5.9)	0.70
Percent male sex, mean (SD)	54 (10)	62 (8)	< 0.01
Percent Caucasian race, mean (SD)	66 (18)	61 (34)	0.72
Baseline weight in kg, mean (SD)	80.4 (9.3)	78.5 (11.8)	0.59
Baseline hemoglobin A1C, mean (SD)	8.2% (0.5)	8.3% (1.2)	0.55
Trials including patients on insulin, %	6	0	0.30
No. of weeks for assessment of primary endpoint, mean (SD)	27 (2.0)	20 (2.4)	0.03
Patients assessed for trial, mean (SD)	1212 (988)	684 (706)	0.03
Patients randomized, mean (SD)	582 (408)	385 (354)	0.047
Percent of patients completing trial, mean (SD)	86 (11)	89 (9)	0.12
Double-blind trials, %	98	33	< 0.01
Trials with concealed allocation, %	56	28	0.052
Trials reporting serious adverse events, %	97	78	< 0.01
Trials reporting severe hypoglycemia, %	98	78	< 0.01

SD standard deviation, kg kilogram

without run-in phases did not differ on average (86%, standard deviation [SD] 11% vs. 89%, SD 9%; $p = 0.12$).

The mean duration of a run-in phase was 4.0 weeks (range 1–21). Most run-in phases (74%) administered placebo rather than active drug and 9% of run-in phases administered neither (Table 2). Of the studies that administered active drug, the most common medication administered was metformin, not a DPP4 inhibitor. About two-thirds of studies used the word "run-in" in describing their methods, while the remaining either used other terminology (e.g., "lead-in"), or did not

Table 2 Characteristics of run-in phases (N = 88)

Mean run-in duration, weeks	4.0 (SD 3.9; range 1–21)
Intervention administered during run-in phase	
Placebo	47 (54%)
Placebo plus exercise	17 (19%)
Drug	7 (8%)
Drug plus exercise or diet	8 (9%)
Diet or exercise	8 (9%)
Unknown	1 (1%)
Word used to describe run-in phase	
"Run-in"	66%
"Lead-in"	20%
"Screening"	9%
"Washout"	3%
"Open-label"	1%
Most common reason for exclusion after run-in	
Not reported	84%
Withdrew consent	8%
No longer met study criteria	6%
Poor adherence	0%
Adverse event	< 1%

SD standard deviation

explicitly label that phase of the study. The reasons why patients were excluded during or after run-in phases were not reported in 84% of studies (Table 2), including in the supplementary documents and appendices.

Drug efficacy

As demonstrated in Table 3, the reduction in hemoglobin A1C at the primary endpoint for a given medication was similar for trials with run-in phases compared to trials without run-in phases (e.g., for linagliptin: 0.67, 95% CI 0.53–0.81 vs 0.64, 95% CI 0.40–0.89; $p = 0.88$). When trial results for all drugs were pooled, the overall reduction in hemoglobin A1C was also similar for trials with and without run-in phases (0.70, 95% CI 0.65–0.75 vs 0.76, 95% CI 0.69–0.84; $p = 0.27$).

Meta-regression, performed to identify factors associated with reduction in hemoglobin A1C, confirmed that the inclusion of run-in phases did not explain differences in reductions in hemoglobin A1C. The inclusion of run-in phases was associated with observed reductions in hemoglobin A1C virtually identical to that observed in trials with no run-in phase (a difference of 0.09%, 95% CI − 0.03, 0.21). The percentage of patients in the trial that were male and the baseline hemoglobin A1C had larger effects on the reduction in hemoglobin A1C.

Drug safety

The proportion of patients with serious adverse events was nearly identical for trials with run-in phases compared to trials without run-in phases (4%, 95% CI 3–5% vs 3%, 95% CI 1–4%; $p = 0.35$; see Table 4). Similar results were also observed for rates of hypoglycemia, and the reported frequency of the most common adverse event. Results for severe hypoglycemia could not be analyzed, because the outcome was too rare to quantify. Meta-regression confirmed that the inclusion of run-in phases did not explain differences in the observed proportions of patients with hypoglycemia, as the inclusion of run-in phases was associated with an observed proportion of patients experiencing hypoglycemia very similar in trials with run-in phases (− 0.01%, 95% CI − 0.06, 0.04) and trials without run-in phases (Table 5).

Table 3 Reduction in hemoglobin A1C within clinical trials of study drugs

	Run-in	No run-in	p
Alogliptin	0.68 (95% CI 0.56–0.81)	1.16 (95% CI 0.26–2.58)	0.08
Linagliptin	0.67 (95%CI 0.53–0.81)	0.64 (95% CI 0.40–0.89)	0.88
Saxagliptin	0.74 (95% CI 0.62–0.87)	0.84 (95% CI 0.46–1.23)	0.63
Sitagliptin	0.70 (95% CI 0.65–0.76)	0.78 (95% CI 0.71–0.85)	0.19
Overall	0.70 (95% CI 0.65–0.75)	0.76 (95% CI 0.69–0.84)	0.27

Table 4 Proportion of patients with adverse events

Proportion of patients		
Hypoglycemia (non-severe)		
Non run-in	3% (95% CI 2–5%)	p = 0.40
Run-in	4% (95% CI 3–5%)	
Most common adverse event		
Non run-in	7% (95% CI 4–9%)	p = 0.08
Run-in	9% (95% CI 8–11%)	
Serious adverse event		
Non run-in	3% (95% CI 1–4%)	p = 0.35
Run-in	4% (95% CI 3–5%)	

CI confidence interval

DISCUSSION

Our review of DPP4 inhibitors showed that trials with run-in phases provided results in terms of efficacy, safety, and the percentage of patients completing the trial that were nearly identical to those seen in trials without run-in phases. These findings differ from the conventional wisdom—supported by past studies[3–6] and textbooks[13,14]—that run-in phases improve efficiency and potentially statistical power, by improving rates of medication adherence and study completion.

The purported benefits of run-in phases, which can also be found in guidance documents from the FDA,[15] often cite the contrasting results of the British Physicians' Study and the Physicians' Health Study, one of the first RCTs to include a run-in phase.[7] However, the observed difference in the studies' main findings could be explained by different baseline risk of myocardial infarction among patients enrolled in the

Table 5 Meta-regression of study-level characteristics on the primary outcomes

Reduction in A1C for a given study characteristic	Run-in design	0.09 (95% CI − 0.03, 0.21)
	Mean age	− 0.01 (95% CI − 0.021,− 0.001)
	Percentage male	0.63 (95% CI 0.1, 1.2)
	Baseline A1C	0.26 (95% CI 0.17, 0.35)
	Duration of trial	− 0.001 (95% CI − 0.004, 0.001)
Proportion of patients with hypoglycemia	Run-in design	− 0.01 (95% CI − 0.06, 0.04)
	Mean age	0.008 (95% CI 0.003, 0.012)
	Percentage male	0.050 (95% CI − 0.18, 0.27)
	Baseline A1C	0.050 (95% CI 0.01, 0.08)
	Duration of trial	0.001 (95% CI 0.0004, 0.0014)
Proportion of patients with serious adverse events	Run-in design	− 0.11 (95% CI − 0.70, 0.48)
	Mean age	0.02 (95% CI − 0.02, 0.05)
	Percentage male	0.46 (95% CI − 1.41, 2.34)
	Baseline A1C	0.038 (95% CI − 0.36, 0.44)
	Duration of trial	0.002 (95% CI − 0.01, 0.01)

CI confidence interval

Physicians' Health Study—not the existence, or lack thereof, of a run-in phase. Results from our study also support the conclusion that baseline risk, rather than the inclusion of a run-in phase, is likely to be a more important determinant in differences of observed drug efficacy. Specifically, a patient's baseline hemoglobin A1C, a marker of disease severity, and the percentage of male patients enrolled were two particularly important study-level characteristics associated with estimated differences in hemoglobin A1C reduction. Conversely, the inclusion of run-in phases was associated with an estimated reduction in hemoglobin A1C (0.09%), hardly different from that observed in trials with no run-in phase. Considering a clinically meaningful change in hemoglobin A1C is about 1 to 2%, these data indicate that including a run-in phase does not have a clinically meaningful impact on the reported change in hemoglobin A1C.

These results may renew the debate on the advisability of run-in phases, which are often justified based on their potential to "weed out" patients who would not adhere well to the regimen being tested. If run-in phases did eliminate study subjects who poorly tolerated the experimental drug, it would raise the question of whether the increased internal validity in assessing a drug's biologic effect that is achieved through their exclusion is worth the loss of generalizability concerning how the drug would actually perform in the "real-world" if many patients stopped taking it because of adverse effects. Run-ins may be inadvisable, in particular for shorter RCTs, because they are time and resource intensive and do not appear to affect the study's results.

We found no differences in the effect estimates for drug safety outcomes between the trials with or without run-in phases. This was true regardless of the class (e.g., serious adverse event) or type (e.g., hypoglycemia) of adverse event reported. Notably, recent RCTs of other medications with run-in phases[16–18] have reported rates of adverse events that some have speculated are artificially low.[19–22] For example, in PARADIGM-HF, patients were randomized to receive a neprilysin inhibitor (LCZ696) or enalapril after they completed three phases: a screening period, a run-in phase with enalapril, and then a run-in phase with LCZ696.[16] The authors stated that the multiple run-in phases were intended to "ensure an acceptable side-effect profile of the study drugs."[16] Of the 10,513 patients who entered the run-in phase, 1138 (11%) were excluded due to an adverse event.[16] Some have argued that the rates of adverse events reported in PARADIGM-HF therefore underestimated the actual risks that might be expected of neprilysin inhibitors among "real-world" patients after FDA approval.[19–21,23]

Our study has several limitations. We assessed the impact of run-in phases within a single class of medications, and thus our results might not be generalizable to other classes of medications. Of the 106 studies we identified, only 18 did not have a run-in phase. Thus, we may have been underpowered to detect the impact of run-in phases. In addition, most run-in phases for DPP4 inhibitors

involved administering placebo, so the results of our study may not be generalizable to studies like PARADIGM-HF in which an active drug is administered during the run-in phase. Since the DPP4 inhibitor trials in our sample lasted only about 30 weeks on average, the impact of run-in phases may be apparent only for trials of longer duration. Finally, understanding why patients were excluded during or after run-in phases is important to understand the impact of run-in phases. However, among the studies we included, over 80% did not report this information.

Our study calls into question the utility of including run-in phases for RCTs of short duration. We have also provided an analytic framework that can be applied to evaluate the utility of a run-in phase for other classes of medications in which there is a combination of trials with and without run-in phases. If these results are consistent in other investigational settings, clinical trial investigators may conclude that the considerable time, energy, and money spent on run-in phases would be better allocated to recruiting more participants for inclusion for the study itself. With an increasing amount of pressure and scrutiny on trial cost and time to completion, including a run-in phase might not be a good use of resources for certain RCTs.

Corresponding Author: Michael Fralick, MD, MSc; Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont St, Ste 3030, Boston, MA 02120, USA (e-mail: mif823@mail.harvard.edu).

REFERENCES

1. Pocock S. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, part i. *Statistician.* 1982;31:1–18.
2. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard—lessons from the history of rcts. Malina D, ed. *N Engl J Med.* 2016;374(22):2175–2181.
3. Pablos-Méndez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. *JAMA.* 1998;279(3):222–225.
4. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials.* 2006;1(1):e9.
5. Lang J, Buring J, Rosner B, Cook N, Hennekens C. Estimating the effect of the run-in on the power of the physicians' health study. *Stat Med.* 1991;10:1585–1593.
6. Lang J. The use of a run-in to enhance compliance. *Stat Med.* 1990;9:87–95.
7. Physicians' Health Study. Final report on the aspirin component of the ongoing physicians' health study. *N Engl J Med.* 1989;321:129–135.
8. Peto R, Gray R, Collins R, et al. Randomised trial of prophylactic daily aspirin in british male doctors. *BMJ.* 1988;296:313–316.
9. Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. *JAMA.* 2011;305(22):2320–2326.
10. Food and Drug Administration. Drugs@fda: fda approved drug products. Available at: http://www.accessdata.fda.gov/scripts/cder/daf/. Accessed January 24, 2017.
11. Dersimonian R, Laird N. Meta-analysis in clinical trials. *Stat Med.* 1986;188:177–188.
12. Newcombe R. Two-sided confidence intervals for single proportion: comparison of seven methods. *Stat Med.* 1998;17:857–872.
13. Merill R. *Introduction to epidemiology.* 7th ed. Burlington, MA: Jones & Bartlett Learning LLC; 2017.
14. Oleckno WA. *Concepts and Methods.* Long Grove, IL: Waveland Press, Inc; 2008.
15. Office of New Drugs In the Center for Drug Evaluation and Research at the Food and Drug Administration. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/UCM377108.pdf. Good review practice: clinical review of investigational new drug applications. Accessed February 22, 2017.
16. McMurray JJV, Packer M, Desai AS, et al. Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N Engl J Med.* 2014;371(11):993–1004.
17. Lam KSL, Chow CC, Tan KCB, et al. Practical considerations for the use of sodium–glucose co-transporter type 2 inhibitors in treating hyperglycemia in type 2 diabetes. *Curr Med Res Opin.* 2016;32(6):1097–1108.
18. Wu JHY, Foote C, Blomster J, et al. Effects of sodium-glucose cotransporter-2 inhibitors on cardiovascular events, death, and major safety outcomes in adults with type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol.* 2016;4(5):411–419.
19. Chen CH. Critical questions about paradigm-hf and the future. *Acta Cardiol Sin.* 2016;32(4):387–396.
20. Correia LCL, Rassi Jr. A. Paradigm-hf: a paradigm shift in heart failure treatment? *Arq Bras Cardiol.* 2016:77–79.
21. Tyler J, Teerlink J. The safety of sacubitril-valsartan for the treatment of chronic heart failure. *Expert Opin Drug Saf.* 2017;16(2):257–263.
22. Vilela-Martin JF. Spotlight on valsartan-sacubitril fixed-dose combination for heart failure: the evidence to date. *Drug Des Devel Ther.* 2016;10:1627–1639.
23. Kang G, Banerjee D. Neprilysin inhibitors in cardiovascular disease. *Curr Cardiol Rep.* 2017;19(2):16.