

TRANSLATIONAL PERSPECTIVES

Detection and Management of Geographic Disparities in the TOPCAT Trial



Lessons Learned and Derivative Recommendations

Michael R. Bristow, MD, PhD,^a Jorge Silva Enciso, MD,^b Bernard J. Gersh, MB, ChB, DPHIL,^c Christine Grady, RN, PhD,^d Madeline Murguia Rice, PhD,^e Steven Singh, MD,^f George Sopko, MD, MPH,^g Robin Boineau, MD, MA,^h Yves Rosenberg, MD, MPH,^g Barry H. Greenberg, MD^b

SUMMARY

TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial) was a multinational clinical trial of 3,445 heart failure with preserved ejection fraction patients that enrolled in 233 sites in 6 countries in North America, Eastern Europe, and South America. Patients with a heart failure hospitalization in the last 12 months or an elevated B-type natriuretic peptide were randomized to the mineralocorticoid receptor antagonist spironolactone versus placebo. Sites in Russia and the Republic of Georgia provided the majority of early enrollment, primarily based on the hospitalization criterion because B-type natriuretic peptide levels were initially unavailable there. With the emergence of country-specific aggregate event rate data indicating lower rates in Eastern Europe and differences in patient characteristics there, the Data Safety and Monitoring Board recommended relatively increasing enrollment in North America plus other corrective measures. Although final enrollment reflected the increased contribution from North America, a plurality of the final cohort came from Russia and Georgia (49% vs. 43% in North America). B-type natriuretic peptide measurements from Russia and Georgia, available later in the trial, suggested no or a mild level of heart failure consistent with low event rates. The primary results showed no significant spironolactone treatment effect overall (primary endpoint hazard ratio [HR]: 0.89; 95% confidence interval [CI]: 0.77 to 1.04), with a significant hazard ratio in North and South America (HR: 0.82; 95% CI: 0.69 to 0.98; $p = 0.026$) but not in Russia and Georgia (HR: 1.10; 95% CI: 0.79 to 1.51; interaction $p = 0.12$). This report describes the Data Safety and Monitoring Board's detection and management recommendations for regional differences in patient characteristics in TOPCAT and suggests methods of surveillance and corrective actions that may be useful for future trials. (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Trial [TOPCAT]; [NCT00094302](https://doi.org/10.1016/j.jacbts.2016.03.001)) (J Am Coll Cardiol Basic Trans Science 2016;1:180-9) © 2016 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

From the ^aCardiovascular Institute, University of Colorado, Boulder and Aurora, Colorado; ^bDepartment of Medicine, Division of Cardiology University of California, San Diego Medical Center, La Jolla, California; ^cDepartment of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota; ^dDepartment of Bioethics, National Institutes of Health Clinical Center, Bethesda, Maryland; ^eGeorge Washington University Biostatistics Center, Washington, DC; ^fDepartment of Medicine, Division of Cardiology Washington Veterans Affairs Medical Center, Washington, DC; ^gDivision of Cardiovascular Sciences, National Heart, Lung and Blood Institute, Bethesda, Maryland; and the ^hOffice of Clinical and Regulatory Affairs, National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, Maryland. The TOPCAT Trial was funded by the National Institutes of Health, National Heart, Lung, and Blood Institute (contract N01 HC45207). The TOPCAT trial was supported by NHLBI contract HHSN268200425207C awarded to New England Research Institutes Clinical Trials Coordinating Center. This work is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or National Institutes of Health. Dr. Bristow is the TOPCAT Data Safety and Monitoring Board (DSMB) chair; and he is an officer and director of Arca Biopharma, and consultant to Miragen Therapeutics. Drs. Gersh, Grady, Rice, Singh, and Greenberg are DSMB voting members. Dr. Gersh has consulting relationships with Mount Sinai St. Lukes, Boston Scientific Corporation, Teva Pharmaceutical, Janssen Scientific Affairs, St. Jude Medical, Janssen Research and Development, Baxter Healthcare Corporation, Cardiovascular Research Foundation, Medtronic Inc., Xenon Pharmaceuticals, Cipla Limited, Thrombosis Research Institute, and

Country- or region-specific differences in outcomes are frequently observed in multinational clinical trials (1-3) and may or may not be indicative of true differences in drug response (4). Geographies may vary with respect to genetics (4,5); nongenetic racial characteristics (5,6); medical practice, training, or infrastructure patterns that may influence outcomes despite general adherence to a clinical trial protocol (4); and other factors. In the planning and conduct of multinational clinical trials, the potential impact of geographic differences in outcomes is among many important factors that needs to be considered, but often is not taken into account (4).

The TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist Trial) (7,8) was a large-scale, multinational National Heart, Lung, and Blood Institute (NHLBI)-sponsored clinical trial conducted in 233 sites in 6 countries located in 3 distinct geographic regions: North America (United States and Canada); Eastern Europe (Russia and the Republic of Georgia); and South America (Argentina and Brazil). The target disease indication investigated, heart failure with preserved ejection fraction (HFpEF) of the left ventricle, has been a challenge to define and enroll in clinical trials and has thus far eluded the development of definitive therapy (9,10). The TOPCAT Data and Safety Monitoring Board (DSMB) worked closely with the NHLBI, Trial Executive/Steering Committee and the Data/Clinical Coordinating Center to deal with several major challenges during the trial. In this report we attempt to provide insight into regional heterogeneity issues in TOPCAT, including how they were detected and the recommendations made by the DSMB to resolve them. Based on this experience we offer suggestions for DSMB oversight of potential geographic disparities in future multinational trials.

METHODS

DSMB ORGANIZATION. The TOPCAT DSMB organization and responsibilities are given in the [Supplemental Appendix](#).

DATABASE. The database consisted of Clinical Trial Coordinating Center reports provided to the open and closed sessions of the various DSMB meetings, the meeting minutes, monthly safety reports viewed by

the DSMB chair, correspondence of the chair with the NHLBI, and published data from the entire TOPCAT cohort (8,11). Interval enrollment data were the most recent figures available at the various reviews or from more extensive “data freeze” analyses performed within 2 months of the meetings.

STATISTICAL METHODOLOGY. Analyses were as described for the TOPCAT trial (7,8,11) using an unadjusted model. Additional analyses were conducted using chi-square/contingency table tests with a Bonferroni correction.

TOPCAT TRIAL

OVERVIEW. The TOPCAT trial enrolled 3,445 patients with HFpEF between August 10, 2006, and January 31, 2012, with a mean follow-up of 3.3 years ending on June 30, 2013. Overall (8) and regional (8,11) differences in outcomes have been previously reported. The primary endpoint of TOPCAT was the composite of time to cardiovascular death, aborted cardiac arrest, or HF hospitalization, with each component adjudicated by a clinical events committee. Inclusion of non-U.S. sites was an integral part of the original study design to enhance generalizability of the results and manage trial costs.

In addition to monthly safety monitoring that was focused on the class-adverse effects of mineralocorticoid receptor antagonists (MRAs) or MRAs in combination with other renin-angiotensin system inhibitors (hyperkalemia, renal function, gynecostasia in each blinded treatment arm), at its scheduled meetings, the DSMB monitored overall enrollment, the aggregate (both treatment groups combined) event rate, and country-specific data. Throughout the trial, the DSMB elected not to disclose the observed aggregate event rate to the Executive/Steering Committee, but to report to them any important departure from the pre-trial assumption expected rate that might affect statistical power. The interim analysis plan included a DSMB review of unblinded primary outcome data by treatment arm at 33%, 50%, and 75% of the expected number of primary endpoints, using events confirmed by the clinical events committee and respective efficacy conditional power futility boundaries of $\leq 10\%$, $\leq 15\%$,

ABBREVIATIONS AND ACRONYMS

BNP = B-type natriuretic peptide

CI = confidence interval

DSMB = Data Safety and Monitoring Board

HF = heart failure

HFpEF = heart failure preserved ejection fraction

HR = hazard ratio

NT-proBNP = N-terminal pro-B-type natriuretic peptide

and $\leq 20\%$. The trial reached completion, with 671 subjects having a primary event (8). The hazard ratios (HR) and 95% confidence intervals (CI) for the primary endpoint was 0.89 (95% CI: 0.77 to 1.04), $p = 0.14$ (8) and 0.83 (95% CI: 0.69 to 0.99), $p = 0.04$ for time to HF hospitalization as a single endpoint (8).

The TOPCAT protocol defined HFpEF as follows: 1) having at least 1 HF symptom present during screening; 2) 1 HF sign present during the last 12 months; and 3) meeting criteria for 1 of 2 design strata: at least 1 hospitalization in the last 12 months “for which heart failure had to be a major component of the hospitalization,” or an elevated B-type natriuretic peptide (BNP) or N-terminal pro-B-type natriuretic peptide (NT-proBNP) level sampled in the last 60 days. To exclude heart failure with reduced ejection fraction, the left ventricular ejection fraction measured during the previous 6 months had to be $\geq 45\%$. The dual hospitalization or NP qualification scheme for enrollment was necessary because at the beginning of the trial BNP or NT-proBNP assays were not universally available. For the final cohort 28% of patients were enrolled using NP criteria (8), 45% of patients in North and South America and 11% in Eastern Europe. The pre-specified subgroup analysis of HF qualifying criteria yielded a significant interaction p value of 0.01, with an NP subgroup HR of 0.65 (95% CI: 0.49 to 0.87, $p = 0.003$) and the 5 HF hospitalization subgroup having no evidence of a treatment effect (HR: 1.01; 95% CI: 0.84 to 1.21; $p = 0.92$) (8).

Of the 981 patients enrolled by NP criteria, 81% were from North or South America. Additional evidence of geographic differences in outcomes by country and region were noted (8,11). Patients enrolled in Russia and Georgia had lower event rates in the placebo arm, and in the spironolactone arm, they showed no evidence of hyperkalemia or renal dysfunction despite having an increased incidence of gynecomastia plus no evidence of a treatment effect (8,11). The HR of the primary endpoint for patients enrolled in the Americas was 0.82 (95% CI: 0.69 to 0.98, $p = 0.026$) compared with 1.10 (95% CI: 0.79 to 1.51, $p = 0.58$) in Russia and Georgia ($p = 0.12$ for interaction between treatment and region) (8,11).

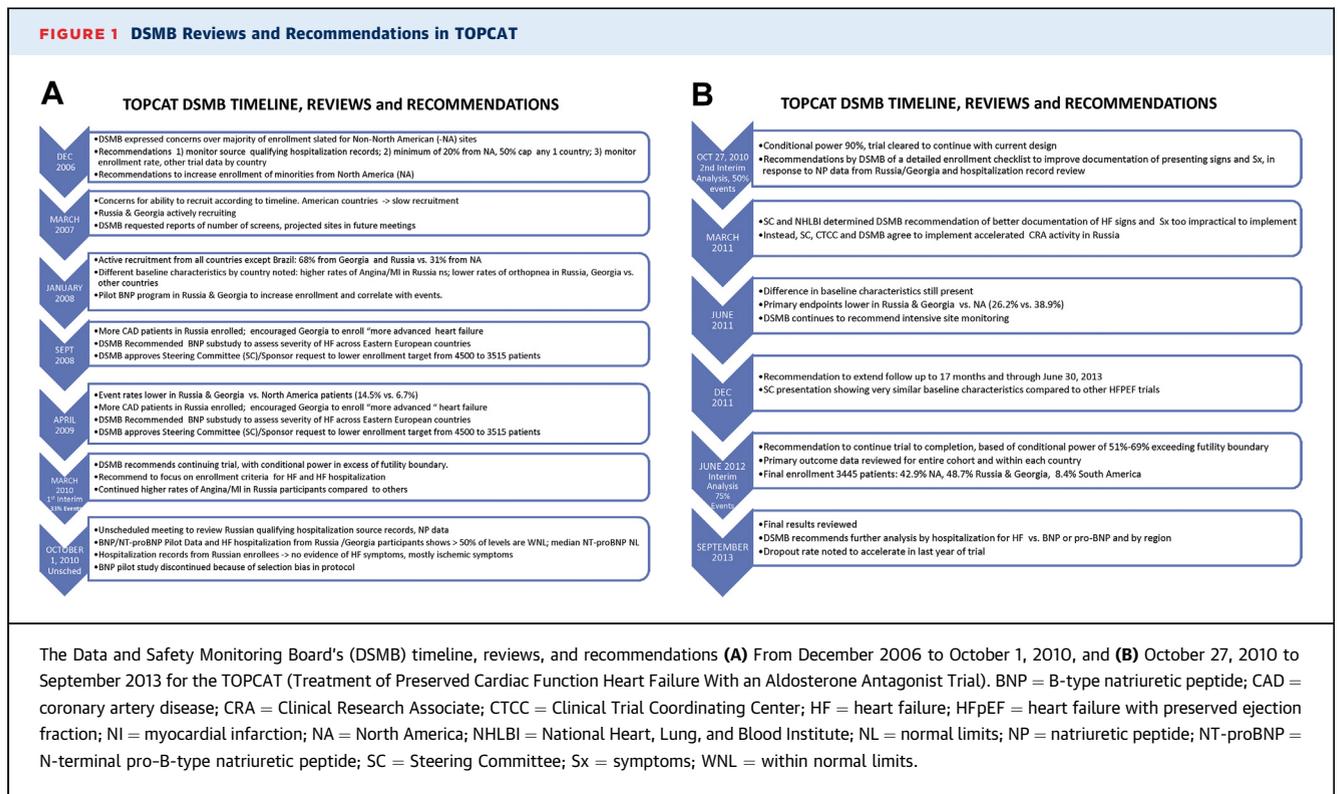
ENROLLMENT. Pre-trial plans and projections. DSMB also served as the protocol review committee, with the final version of the protocol approved in January 2005. The protocol contained language on the logistical and clinical research organizational support for recruiting patients from North America, South America, Western Europe, and Eastern Europe. DSMB became aware of the potential for non-North American sites to influence the overall outcome of the trial

when the plan to enroll a large number of patients from Russia and the Republic of Georgia was reported by trial leadership and the sponsor at the first DSMB meeting in December 2006 (Figure 1). Support for an Eastern European strategy as a means of facilitating recruitment was offered by the TOPCAT Steering Committee leadership, who reported that the CHARM-Preserved (Candesartan Cilexetil in Heart Failure Assessment of Reduction in Mortality and Morbidity: Clinical Study of Candesartan in Patients With Heart Failure and Preserved Left Ventricular Systolic Function) trial (9) “had not observed any interaction between patient prognosis and country,” including Russia. Nevertheless, the DSMB expressed concern over enrolling a large number of patients with a diagnosis of HFpEF in non-North American sites. The DSMB-recommended strategy to deal with this issue included the following: 1) monitoring of source documents to check that the qualifying hospitalization met the HF criterion, for which translation of source records would be needed; 2) requesting that at a minimum, 20% of all patients enrolled should come from North America, and the enrollment from any 1 country should not exceed 50%; and 3) asking that enrollment rates and other trial data be monitored by country.

The projected enrollment by region, presented by Steering Committee leadership at the January 2008 DSMB meeting (Figure 1), is given in Figure 2 juxtaposed against the actual final enrollment on January 31, 2012. In the pre-trial projections, 51.1% of patients were expected to be enrolled in Eastern Europe (42.2% in Russia and 8.9% in Georgia), 27.8% in North America (20.0% in the United States and 7.8% in Canada), and 21.1% in South America (11.1% in Argentina and 10.0% in Brazil). The final Russia-Georgia enrollment was close to projections, whereas North American enrollment was higher and South American was lower (Figure 2). The trend for enrollment to favor more North American participation over time was at the request of the DSMB, to counter the early dominance of Eastern European enrollment.

Recruitment dynamics. Enrollment milestones and important trial developments are summarized in Figure 1 and discussed in more detail in the Supplemental Appendix. The first patient was enrolled in August 2006, approximately 1 year behind schedule due primarily to delays in manufacture and supply of study medication. In March 2007, the DSMB was provided with the first data available for review, for 165 patients randomized in Russia, Georgia, and the United States. Figure 3 shows the dominance of Russia and Georgia in the early phases of trial enrollment, followed by the steady ascendancy of

FIGURE 1 DSMB Reviews and Recommendations in TOPCAT



North American randomization. At the completion of enrollment in January 2012, the Russia and Georgia contribution had fallen below 50% of the total, but still constituted a plurality: North America, n = 1,477 (42.9%); Russia and Georgia, n = 1,678 (48.7%); and South America, n = 290 (8.4%) (Figure 3).

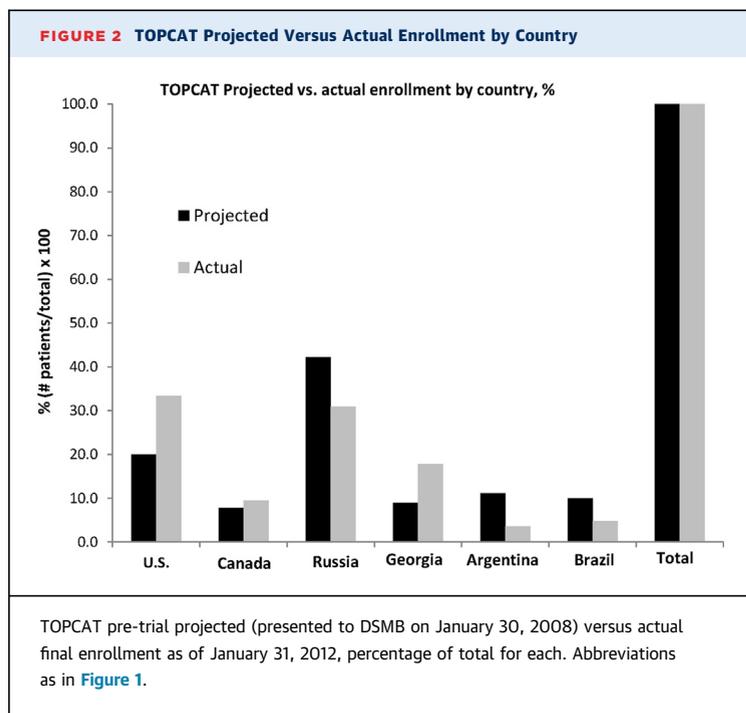
GEOGRAPHIC DISCREPANCIES IN RECRUITMENT AND CLINICAL COURSE OF PATIENT POPULATIONS. Details of the DSMB's detection of and management recommendations for geographic discrepancies in the TOPCAT patient population are given in the Supplemental Appendix, with highlights summarized in the next section and in Figure 1.

Emergence and detection; tactical responses. The first evidence suggesting that patients from Russia or Georgia were potentially different from other TOPCAT regions emerged from country-specific baseline characteristics for the first 794 enrolled patients, presented at the January 2008 DSMB scheduled review (Figure 1). The data indicated that a history of myocardial infarction or angina was more prevalent in patients enrolled in Russia, and that Russian and Georgian patients had less orthopnea by history than did patients from other countries. The significance of these early findings was not clear at the time, but based on subsequent developments, these data likely reflected a predominance of

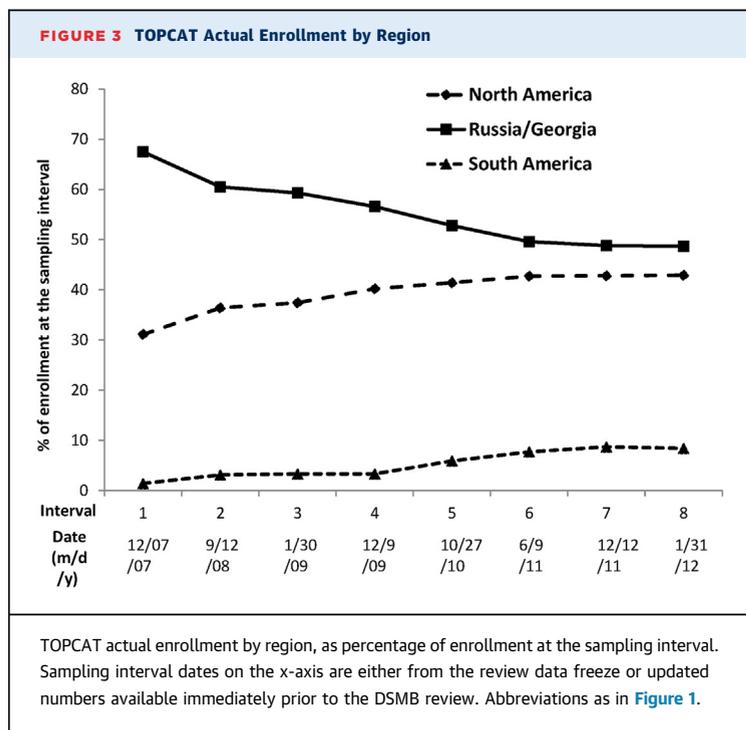
coronary artery disease pathophysiology in Russia and a lesser degree or absence of HF in Russian and Georgian patients compared with patients from other countries.

The first indication that the aggregate (combination of both treatment arms) event rate was lower in Georgia or Russia was revealed in the September 2008 closed session review (Figure 1), where based on 1,461 total randomized patients, Georgia's was 2.0% compared with 5.4% overall and 9.0% in the United States. However, the number of events was low (only 6 in Georgia and 38 in the United States). These data were insufficient for drawing firm conclusions, but in response, the DSMB recommended that the trial leadership "encourage Georgian Investigators to enroll patients as expected per the study protocol."

At the September 2008 meeting, the DSMB approved a plan to reduce the total target enrollment from 4,500 to 3,515 patients with 2 years of minimum follow-up, based on the aggregate event rate being as expected and estimated length of median follow-up being longer, plus accepting a power calculation of >80% as opposed to 90%. In this event-driven trial, the new sample size calculations were based on 80% and 85% power for 551 and 630 primary events, respectively, assuming a 20% relative difference



between the 2 treatment arms (8). In addition, the DSMB requested a substudy of BNP in patients entering the trial via a history of HF hospitalization to assess the severity of HF at baseline among the different countries, and that the trial obtain a BNP or NT-proBNP on every patient enrolled in Georgia and Russia. Finally, the DSMB requested that adverse



events, serious adverse events, and primary endpoint event rates be presented by country at all future meetings.

At the next DSMB review in April 2009 (Figure 1), the country-specific unadjudicated aggregate primary event rate patterns first noted at the September 2008 meeting persisted and are given in Table 1. Georgia was an outlier, with an event rate 75% lower than the composite of all other countries ($p < 0.0001$) and 83% lower than the United States ($p < 0.0001$). The event rate in Russia was 35% lower than the average of other countries ($p = 0.017$), 56% lower than the United States ($p < 0.0001$), and 2.6-fold higher than Georgia ($p = 0.010$). The statistical analysis of the event rates in Table 1 is post hoc and was not performed at the time of data review. However, the lower event rate in Georgia was noted at the DSMB review, which prompted it to recommend an increase in enrollment of patients in the United States and Canada in an attempt to address the circumstances in Russia and Georgia. In March, 2010, the first unblinded interim analysis was performed at 33% of the expected primary events (Figure 1). Based on the conditional power being above the futility threshold, the recommendation was to continue the trial with the current design.

In early October 2010, the NHLBI Program divisional leadership managing the trial held an unscheduled meeting with the DSMB chair (Figure 1) to review results of the requested BNP pilot project in Russia and to discuss a review of source records for the qualifying hospitalization that had been reviewed by an NHLBI Program staff member fluent in Russian. In the vast majority (19 or 20) of the 22 reviewed hospitalizations, ischemic symptoms rather than HF appeared to predominate. It was noted by NHLBI Program staff that this was consistent with ischemic heart disease prevalence data from the I-PRESERVE (Irbesartan in Heart Failure With Preserved Systolic Function) (10) and EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure: Outcome Study with Tolvaptan) (12) trials. The BNP and NT-proBNP data from Russia were compared with data from other countries (Supplemental Table 1 and associated discussion in the Supplemental Appendix). The majority of patients who had been enrolled via the HF hospitalization criteria, most of whom had NP samples drawn after enrollment and not during the hospitalization, had values within the normal range. In contrast, very few (9%) U.S. or Canadian patients, all of whom had NP draws done during the index hospitalization, had values within the normal range. The NT-proBNP median value for all patients entering the trial via the hospitalization criterion in Russia or

Georgia was within the normal range (respectively, 233 pg/ml and 164 pg/ml), whereas it was markedly elevated (887 pg/ml) in the United States and Canadian patients. Concerns over these data were expressed in written communication to the NHLBI, where the DSMB chair outlined a strategy recommending that the Steering Committee institute closer monitoring of enrollment criteria for patients in both Russia and Georgia (Supplemental Appendix).

These recommendations were accepted by the full DSMB at the scheduled second interim analysis meeting later in October (Figure 1), where unblinded outcome data on 2,732 patients easily exceeded the conditional power futility boundary (Supplemental Appendix). During this meeting, the trial leadership and the Clinical Trial Coordinating Center pointed out that the BNP pilot data were subject to selection bias because most samples in North America were obtained during a HF hospitalization, as opposed to in Russia and Georgia where NP were drawn post-randomization after the index hospitalization. Nevertheless, the low (208 pg/ml combined) median values of Russian and Georgian patients were known to be associated with low mortality or cardiovascular hospitalization event rates in heart failure with reduced ejection fraction (13) and were subsequently shown to be associated with a low composite cardiovascular mortality and HF hospitalization rate in HFpEF (14). Thus, although the NP data from the pilot study did not allow a direct comparison of Eastern European enrolled patients to those enrolled in the Americas, they were consistent with the low primary event rates in Russia and Georgia. After extensive discussion of this issue, it was decided to terminate the NP pilot study and to again emphasize to Russian and Georgian investigators to ensure that patients hospitalized for apparent HF met the eligibility criteria for the trial.

By the June 2011 review (Figure 1), 3,080 patients had been enrolled, and for the first time, U.S. enrollment (n = 1,024) exceeded any other country (Figure 3). The previously noted country-specific differences in baseline characteristics and event rates persisted at this and the subsequent review in December 2011 (Figure 1), which was conducted in 3,317 patients. At the June 2012 DSMB review (Figure 1), full enrollment had been reached 5 months earlier, with the United States leading recruitment at 33.4% of the total (Table 1). Relative to the projected enrollment, the actual final enrollment by country was United States 167% over target, Canada 122%, Russia 73%, Georgia 200%, Argentina 32%, and Brazil 48% (Figure 2). However, despite the overenrollment in the United States and Canada, the 2 Eastern European countries enrolled 48.7% of the total, compared with 42.9% in North America and 8.4% in South America.

The final interim analysis, at 75% of the projected number of primary events, was conducted at the June 2012 review (Figure 1). Conditional power again was well above the futility boundary (20%) and was 51% using the observed event rates in each arm modeled forward to the completion of follow-up, or 69% using the observed placebo event rate but the pre-trial/expected 20% difference in crude event rates in the remaining patients active in the trial who had not had a primary event. The country-specific HR and number of events are shown in Table 2. Based on 382 patients with confirmed primary events, the overall HR was 0.792 (p = 0.020, efficacy boundary for stopping = 0.001). At this interim, an additional 161 patients had unconfirmed events that were pending adjudication, and when they were included, the HR p value was 0.118. Country-specific HR were available for the first time and are shown in Table 2. For the confirmed events, there was no evidence of a treatment effect in

TABLE 1 Enrollment and Event Rate in Data Reviewed on April 13, 2009, for 1,707 Randomized Patients (From Data Freeze of January 30, 2009)

	Final Number of Patients Randomized (% of Total)	Number of Patients Randomized on January 30, 2009 (%*) {%†}	Number of Primary Events (%*)	Event Rate (%)	p Value vs. Georgia	p Value vs. Russia
United States	1,151 (33.4)	501 (29.4) {82}	66 (50.0)	13.2	<0.0001	<0.0001
Canada	326 (9.5)	137 (8.0) {60}	13 (9.8)	9.5	0.0004	0.11
Russia	1,066 (30.9)	657 (38.5) {51}	38 (28.8)	5.8	0.010	–
Republic of Georgia	612 (17.8)	355 (20.8) {119}	8 (6.1)	2.2‡	–	0.010
Argentina	123 (3.6)	57 (3.3) {14}	7 (5.3)	12.3	0.0002	0.053
Brazil	167 (4.8)	0 (0) {0}	0 (0)	0	–	–
Total	3,445 (100)	1,707 (100) {56}	132 (100)	7.7	–	–

Dashes indicate no possible comparison or data were not available. *Of the total trial enrollment or events on September 16, 2008. †Of target allocation. ‡Chi-square, with p = 0.0125 the Bonferroni critical value.

Georgia, but the number of primary events ($n = 14$) was extremely small. There was also little evidence of a treatment effect in Russia (HR 0.95) despite 62 primary events being observed. These country-specific HR are consistent with the final trial outcome in North and South America combined versus Russia and Georgia (8,11).

TOPCAT trial follow-up was completed on June 30, 2013, and top line results were presented to the DSMB and NHLBI on September 18, 2013 (Figure 1). The total number of primary events was 671 (8), which conferred >85% power to detect a relative difference in event rates of 20% between the 2 treatment arms.

Potential causes of geographic discrepancies.

In TOPCAT, 4 of the 6 countries exhibited relatively similar patient characteristics and event rates, as well as favorable effects of spironolactone treatment. Data from the 2 Eastern European countries differed from the 4 countries in the Americas, but also from each other. The Russian cohort was dominated by clinical presentations of ischemic heart disease, a common cause of HFpEF, and patients may have been symptomatic (dyspneic) due to ischemia rather than HF. An increased prevalence of an ischemic etiology in Eastern European HFpEF patients that was known unofficially during the TOPCAT trial was eventually published for the I-PRESERVE and CHARM-Preserved patient populations (15). It is therefore possible that ineffectiveness of spironolactone as an anti-ischemic agent, rather than lack of efficacy against HFpEF, might have contributed to the TOPCAT results in Russia.

A predominance of ischemic heart disease symptoms, however, was apparently not the issue in patients enrolled in the Republic of Georgia, who from the beginning of the trial had a very low event rate and little evidence for HF based on signs and symptoms or random NP measurements. The lack of any treatment effect in these patients may well have been due to the fact that HF was either absent or much milder than in those enrolled in the Americas.

Although study drug compliance did not appear to account for the lack of treatment effect in Russia and Georgia, as these countries reported using higher doses of both spironolactone and placebo (11), in the spironolactone arm, there was a substantially higher incidence of hyperkalemia and elevations in serum creatinine in the Americas than in Russia and Georgia (11). This was interpreted as a “lack of pharmacologic effect” in Russian and Georgian patients (11), but it could also be due to the absence of actual study drug consumption. However, Russian and Georgian patients had an increase in gynecomastia in the

spironolactone versus the placebo arm (11), indicating they were likely taking study medication. Gynecomastia in the absence of marked hyperkalemia has been observed in another spironolactone versus placebo HFpEF study (16) and could be characteristic of a subcohort of patients. These pharmacodynamic adverse event data were not reviewed by country during the trial by the DSMB, although they were being tracked monthly for the entire trial. Country-specific adverse events and severe adverse events within their standard regulatory organ system groupings were tracked by the DSMB, and no obvious differences between countries or regions were noted in any review.

LESSONS LEARNED AND RECOMMENDATIONS FOR FUTURE TRIALS

LESSONS LEARNED. Whereas the country-specific and regional heterogeneity in TOPCAT could be viewed as expected statistical variation in a large multinational trial (1,3), the differences in patient characteristics, lower event rates, lack of certain drug class-related pharmacodynamic effects (11), and complete lack of treatment effect in Russia and Georgia compared with the other regions strongly suggest that more than the play of chance occurred.

Enrollment issues. Because of the difficulty in identifying the phenotype (17) and other issues, enrollment of patients into HFpEF clinical trials can be challenging. Consequently, in HFpEF trials, the pressure to enroll at a projected rate versus the imperative to confine enrollment to the target population is even more in conflict than usual. TOPCAT started behind schedule, and once begun, there was brisk enrollment from 2 countries where data ultimately proved to be qualitatively problematic. By the time it was appreciated that there were serious issues with patient characteristics and event rates in patients from Russia and Georgia, both countries had enrolled substantial numbers of subjects. The lesson learned here is that during the early as well as the later phases of trial enrollment, recruitment from 1 or 2 regions or sites should not dominate, and patient characteristics should be monitored carefully to identify potential regional irregularities in study subpopulations.

What you see early may be what you get late.

The discrepant patterns of patient characteristics and event rates in Russia or Georgia were present from the first opportunity to observe them and persisted throughout the trial. Although it is well known that treatment effects can vary during a trial, fluctuations in patient characteristics and overall event rates may

TABLE 2 Overall and Country-Specific HR and Total Number of Primary Events Confirmed by Adjudication at the 75% Interim Analysis

	HR	Number of Subjects With Confirmed Primary Event
United States	0.786	230
Canada	0.642	49
Russia	0.950	62
Republic of Georgia	0.993	14
Argentina	0.745	18
Brazil	0.473	9
Overall	0.792	382

HR = hazard ratio.

not exhibit such plasticity. It would seem prudent to place considerable weight on early observations that remain consistent, particularly if they could threaten trial integrity.

Country- and region-specific event rates need to be carefully followed and may demand dissemination beyond the DSMB. At no time during the trial did the aggregate event rate depart from expectations, and this was periodically communicated to the Steering Committee. Country-specific event rates were not specifically disclosed by the DSMB to the Steering Committee, and it is possible that such information would have triggered earlier or more definitive corrective actions. Continuing to enroll in a region where the event rate is inadequate to assess the tested treatment effect is questionable and, at a minimum, should be brought to a steering committee’s attention. In TOPCAT, this information was communicated only indirectly. Once the DSMB was confident there was an issue, direct reports of the actual aggregate event rates to the trial’s leadership might have resulted in different management decisions.

In a large, multicenter clinical trial, change is difficult but not changing may be fatal. Both aspects of this popular management trope apply to large multicenter trials, especially multinational ones. For various reasons that include regulatory compliance and site burden, investigators are understandably reluctant to make substantive changes in clinical protocols during a trial. Yet there are often developments in trials that need major adjustments or changes in overall approach. In TOPCAT, the first opportunity for geographic disparity corrective measures occurred at the review in April 2009, with the confirmation that event rates were extremely low in Georgia and low in Russia. The DSMB recommendation at the time was continued enhanced enrollment

of patients in North America. Thereafter, an ad hoc NP study strongly suggesting much less advanced HF in Russia or Georgia plus persistence of the lower event rates and the trial’s response of accelerated site monitoring did not lead to any apparent change in the clinical characteristics of randomized patients or to curtailment of enrollment. In fact, Georgia finished at 200% of its enrollment target, and after April 2009, another 666 patients were enrolled in Russia and Georgia. If those patients had been enrolled in the Americas, it could be speculated that the trial may have been positive.

Requirement of an elevated B-type natriuretic peptide measurement in future HFpEF trials? As referenced in the trial overview section, the HR in the 981 patients enrolled with an elevated BNP or NT-proBNP was highly statistically significant at 0.65 compared with the nonsignificant HR in the 2,464 patients enrolled based on a history of HF hospitalization (8,11). In HFpEF an elevated NP level adds precision to the HF diagnosis and provides objective evidence for a certain degree of HF severity (13,14). NP assays are now available in virtually all regions where clinical trial capability exists. However, in I-PRESERVE, a treatment effect of the angiotensin-receptor blocker irbesartan was only observed in patients with randomization NT-proBNP levels below the median of 339 pg/ml (18), a value below the TOPCAT qualifying value of 360 pg/ml. In addition, in TOPCAT, NP levels are not easily separated from the geographic disparity issues. Therefore, further analysis of the TOPCAT data will be required to add credence to any recommendation regarding baseline NP levels and eligibility criteria for future HFpEF trials.

Adaptive increases in trial sample size can be prospectively designed and may increase the probability of success. Despite the country-/region-specific issues in TOPCAT, it is possible that the trial could have been salvaged based on information available at the 75% of total events interim analysis. Specifically, when a late/pre-specified interim analysis has a HR within certain boundaries termed the “promising zone” the sample size may be increased without inflating type 1 error (19,20). In general the promising zone band encompasses conditional powers between 50% and 80% (19,20), which is where the TOPCAT conditional power calculations were at the 75% interim. Operationally, the option to increase sample size by a pre-specified amount based on conditional power or Bayesian predictive probabilities needs to be prospectively defined, and the sponsor must be willing to support the increase in trial budget to account for the sample size increase.

An increase in sample size, by for example 25%, could have been a recommendation by the DSMB if this “adaptive” strategy had been incorporated into the design. However, NHLBI budget limitations, overall trial fatigue, and the downside of extended follow-up of already enrolled patients all would have argued against such an approach. Notably the 75% interim analysis overall HR was substantially lower (0.79) (Table 2) than the final result of 0.89, and so some of these negative factors including an increasingly high study drug discontinuation rate that reached 34% in the spironolactone arm at trial completion (8) (Supplemental Appendix) may have adversely affected outcomes between the 75% interim and completion of the trial. If in fact any additional enrolled patients would have had outcomes similar to those recorded between the 75% interim analysis and the end of the trial, the additional increase in sample size would not have rendered the trial positive.

RECOMMENDATIONS FOR FUTURE TRIALS. Based on these issues and discussion, the TOPCAT DSMB has the following suggestions for multinational clinical trials:

1. Launch the trial in a variety of geographic jurisdictions, and do not allow 1 or 2 geographic areas to dominate early (or late) enrollment.
2. Follow country- and region-specific patient characteristics and aggregate event rates carefully, beginning early in the trial; if a country or region exhibits event rates that are statistically significantly lower than the composite of other regions and especially if this is reinforced by differences in disease characteristics, bring this to the attention of Steering Committee leadership.
3. Establish detailed plans for trial surveillance in the DSMB charter, and at the initiation of the trial, inform investigators and national leaders of proposed country- and region-specific analyses of patient data and requirements for characteristics of the study population, with the directive that the trial may be subject to geographic constraint if enrolled patients do not fulfill pre-trial assumptions and/or are substantially different from other regions.
4. Incorporate objective measures used to determine disease presence and severity to the greatest extent possible in enrollment criteria, particularly for conditions such as HFpEF where the diagnosis can be challenging.

CONCLUSIONS

In summary, to paraphrase a popular aphorism from American football (21), multicenter, multinational clinical trials are a rough game and often a cruel one. They require extreme cooperation from groups of individuals and institutions with experience and skill, a willingness to adjust to unanticipated circumstances, and the ability to make difficult decisions. Unanticipated developments are to be expected, and provisions can and should be built into trial design to facilitate identifying and managing them.

ACKNOWLEDGMENTS The authors thank Bertram Pitt, MD (chair of the Steering and Executive Committees), as well as Marc A. Pfeffer, MD, PhD, and Sonja McKinlay, PhD (trial co-principal investigators), for their tireless and expert devotion to design and management of the TOPCAT trial. The authors thank Sonja McKinlay and Susan F. Assmann, PhD, for their inputs into the manuscript. The authors are grateful to Susan F. Assmann and Brian J Harty, MA, of the Clinical Trial Coordinating Center (New England Research Institutes) for supplying the DSMB monthly safety reports, interim trial data, and statistical support. The authors also thank Rachel Rosenberg for manuscript editing and handling, and Ben Harnke for sourcing the NFL Films track.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Michael Bristow, University of Colorado, 12700 East 19th Avenue, Campus Box B-139, Aurora, Colorado 80045. E-mail: michael.bristow@ucdenver.edu.

REFERENCES

1. Wedel H, Demets D, Deedwania P, et al., for the MERIT-HF Study Group. Challenges of subgroup analyses in multinational clinical trials: experiences from the MERIT-HF trial. *Am Heart J* 2001;142:502-11.
2. O'Connor CM, Fiuzat M, Swedberg K, et al. Influence of global region on outcomes in large heart failure β -Blocker trials. *J Am Coll Cardiol* 2011;58:915-22.
3. Pocock S, Calvo G, Marrugat J, et al. International differences in treatment effect: do they really exist and why? *Eur Heart J* 2013;34:1846-52.
4. Glickman SW, McHutchison JG, Peterson ED, et al. Ethical and scientific implications of the globalization of clinical research. *N Engl J Med* 2009;360:816-23.
5. Taylor MR, Sun AY, Davis G, Fiuzat M, Liggitt SB, Bristow MR. Race, genetic variation, and therapeutic response disparities in heart failure. *J Am Coll Cardiol HF* 2014;2:561-72.
6. Gravlee CC. How race becomes biology: embodiment of social inequality. *Am J Phys Anthropol* 2009;139:47-57.
7. Desai AS, Lewis EF, Li R, et al. Rationale and design of the treatment of preserved cardiac function heart failure with an aldosterone antagonist trial: a randomized, controlled study of spironolactone in patients with symptomatic heart failure and preserved ejection fraction. *Am Heart J* 2011;162:966-72.e10.

8. Pitt B, Pfeffer MA, Assmann SF, et al., for the TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014;370:1383-92.
9. Yusuf S, Pfeffer MA, Swedberg K, et al., for the CHARM Investigators and Committees. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved trial. *Lancet* 2003;362:777-81.
10. Massie BM, Carson PE, McMurray JJ, et al., for the I-PRESERVE Investigators. Irbesartan in patients with heart failure and preserved ejection fraction. *N Engl J Med* 2008;359:2456-67.
11. Pfeffer MA, Claggett B, Assmann SF, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation* 2015;131:34-42.
12. Blair JE, Zannad F, Konstam MA, et al., EVEREST Investigators. Continental differences in clinical characteristics, management, and outcomes in patients hospitalized with worsening heart failure results from the EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure: Outcome Study with Tolvaptan) program. *J Am Coll Cardiol* 2008;52:1640-8.
13. Kubánek M, Goode KM, Lánská V, Clark AL, Cleland JG. The prognostic value of repeated measurement of N-terminal pro-B-type natriuretic peptide in patients with chronic heart failure due to left ventricular systolic dysfunction. *Eur J Heart Fail* 2009;11:367-77.
14. Kristensen SL, Jhund PS, Køber L, et al. Relative importance of history of heart failure hospitalization and N-terminal pro-B-type natriuretic peptide level as predictors of outcomes in patients with heart failure and preserved ejection fraction. *J Am Coll Cardiol HF* 2015;3:478-86.
15. Kristensen SL, Køber L, Jhund PS, et al. International geographic variation in event rates in trials of heart failure with preserved and reduced ejection fraction. *Circulation* 2015;131:43-53.
16. Edelmann F, Wachter R, Schmidt AG, et al., for the Aldo-DHF Investigators. Effect of spironolactone on diastolic function and exercise capacity in patients with heart failure with preserved ejection fraction: the Aldo-DHF randomized controlled trial. *JAMA* 2013;309:781-91.
17. Cleland JG, Pellicori P. Defining diastolic heart failure and identifying effective therapies. *JAMA* 2013;309:825-6.
18. Anand IS, Rector TS, Cleland JG, et al. Prognostic value of baseline plasma amino-terminal pro-brain natriuretic peptide and its interactions with irbesartan treatment effects in patients with heart failure and preserved ejection fraction: findings from the I-PRESERVE trial. *Circ Heart Fail* 2011;4:569-77.
19. Chen YH, DeMets DL, Lan KK. Increasing the sample size when the unblinded interim result is promising. *Stat Med* 2004;23:1023-38.
20. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med* 2011;30:3267-84.
21. Facenda J. Pain is inevitable, on *The Power and the Glory: The Original Music and Voices of NFL Films*. New York, NY: Tommy Boy Music, 1998; track 25. Available at: <http://www.cduniverse.com/search/xx/music/pid/1020022/a/power+and+the+glory%3A+music+%26+voices+of+nfl+films.htm>. Accessed January 7, 2016.

KEY WORDS clinical trials, data and safety monitoring of clinical trials, geographic inconsistencies in clinical trials, heart failure with preserved ejection fraction, multinational clinical trials, spironolactone

APPENDIX For the DSMB organization, additional narrative as well as a supplemental table, please see the supplemental appendix of this article.