

Impact of retrospective data verification to prepare the
ICON6 trial for use in a marketing authorization
application

Andrew Embleton, Elizabeth Clark, Stephen Townsend, Laura
Farrelly, Babasola Popoola, Judith Parker, Gordon Rustin,
Matthew Sydes, Mahesh Parmar, Jonathan Ledermann, Richard
Kaplan

Background

The ICON6 trial (ISRCTN68510403) was a phase III academic-led, international, randomized, three arm, double-blind, placebo-controlled trial of the addition of cediranib to chemotherapy in recurrent ovarian cancer. It investigated the use of placebo during chemotherapy and maintenance (arm A), cediranib alongside chemotherapy followed by placebo maintenance (arm B) and cediranib throughout both periods (arm C). Results of the primary comparison showed a meaningful gain in progression-free survival (PFS, time to progression or death from any cause) when comparing arm A (placebo) with arm C (cediranib). As a consequence of the positive results AstraZeneca engaged with the trials unit to discuss regulatory submission using ICON6 as the single pivotal trial.

Methods

A relatively limited level of on-site monitoring, single data entry and investigator's local evaluation of progression were used on trial. In order to submit a license application the trial team and company involved decided that: (i) extensive retrospective Source Data Verification (SDV) of medical records against Case Report Forms should be performed; (ii) conduct further Quality Control (QC) checks of accuracy of data entry, and (iii) that Blinded Independent Central Review (BICR) of images used to define progression should be undertaken. To assess the value of these extra activities we summarize the resulting individual changes and the impact on the reporting of both efficacy and safety outcomes.

Results

The volume of data point changes was minimal; those checked that were key to the primary results had a 0.47% error rate (36/7686), and supporting data points relating to the complete reporting of the trial had a 0.18% error rate (109/59,261). The impact of the SDV and QC processes were analyzed jointly. The conclusion drawn for the primary outcome measure of PFS between arm A and arm C was unchanged. This was shown by the log-rank test p-value changing only at the sixth decimal place, the Hazard Ratio not changing from 0.57 with the exception of a marginal change in its upper bound (0.74 to 0.73), and the median PFS benefit from arm C remaining at 2.4 months. Separately, the BICR of progression scans was performed as a sensitivity analysis. The analysis set was reduced by 4% due to the availability of scans. Estimates and p-values varied slightly, given the change in population, but overall demonstrating a difference in arms consistent with the initial result. Some increases in toxicity were observed, though these were generally minor, with the exception of hypertension where 12% of patients had an increase in their highest grade observed on trial. Though none of these increases were systematically biased towards one arm.

Conclusions

The conduct of this academically-sponsored, pragmatically-conducted trial was sufficient given the robustness of the results, shown by the results remaining largely unchanged following retrospective verification despite not being designed for use in a marketing

authorization. The burden of the retrospective effort required to ensure the results of ICON6 were acceptable to regulators is difficult to justify.

Keywords

data verification; quality; independent review; data entry; monitoring, error rate

Background

The extent of data verification conducted on a trial led by a pharmaceutical company with a marketing authorization in mind and one conducted more pragmatically with a focus principally on the advancement of knowledge of the intervention, can diverge substantially. Usually this is a practical decision driven by its feasibility as extensive verification of data is costly with only limited value to the sponsors.

The ICON6 trial (NCT00532194) was an academic-led, randomized phase III clinical trial investigating the use of cediranib (original tentative marketing name Recentin and later Zemfirza, made by AstraZeneca) in addition to a standard chemotherapy regimen in relapsed ovarian cancer. This UK Medical Research Council-sponsored trial was conducted pragmatically, allowing investigator's-choice chemotherapy and delegating roles and responsibilities for conduct and regulation to each of the four international collaborative groups. These groups were MRC-NCRI (UK Medical Research Council-National Cancer Research Institute), CCTG (Canadian Cancer Trials Group), ANZGOG (Australia New Zealand Gynecological Oncology Group) and GEICO (Grupo Español de Investigación en Cáncer de Ovario). The trial was centrally coordinated by the MRC Clinical Trials Unit at UCL (CTU). Centralized risk-based monitoring was used in a proactive manner throughout the conduct of ICON6 with, for example, consent and eligibility checks, database plausibility algorithms and triggered for-cause site visits (full detail in the Appendix). The trial had enrolled fewer than 390 patients (from a planned 2000) when drug supply availability drove a redesign and premature termination of recruitment.

Patients were randomized to receive one of three arms: chemotherapy with placebo throughout (arm A); chemotherapy with concurrent cediranib, followed by placebo during a maintenance period (arm B); chemotherapy with concurrent cediranib, followed by maintenance cediranib (arm C). This trial design is shown in Figure 1. The original pre-specified primary outcome measure was a comparison of the length of overall survival (OS) between arms A, B and C. Secondary outcome measures included progression-free survival (PFS, time to progression or death from any cause), toxicity and quality of life.

While ICON6 was still accruing patients AstraZeneca (AZ) discontinued manufacture of the drug, following evidence from other randomized clinical trials that cediranib was not effective in several tumor types (colorectal cancer, non-small-cell lung carcinoma and glioblastoma).¹⁻³ This meant a reduction in the target sample size of ICON6 to approximately 470 patients owing to the very limited remaining supply of the drug, and a reconfiguration of the primary comparison to maintain as much power as possible with the much smaller sample size. The primary outcome was swapped from OS to the more frequent, and earlier occurring, measure of PFS (OS is discussed in detail in the Appendix). The primary comparison was also revised from arms A vs. B vs. C to only compare arms A vs. C which was considered the most appropriate comparison to target in view of emerging data for concurrent and maintenance bevacizumab. It is important to note that this change of primary outcome was made before any interim analysis was performed of efficacy outcome measures.

Results of the pre-specified primary analysis showed clear evidence of an improvement in progression-free survival with a Hazard Ratio 0.57 (95% CI 0.45—0.74; p=0.0000097) and a 2.3 month improvement in the difference between the median time before the cancer returned (arm C versus arm A). This improvement, from 8.7 to 11.1 months in the median time to progression or death from any cause would be considered by many to be a clinically meaningful improvement in relapsed ovarian cancer, a setting where patients receive diminishing absolute benefit from each subsequent treatment regimen post-progression and the duration of the response is short. Although this benefit came with added toxic effects. These trial results were initially presented at the Presidential Session of the European Cancer Congress in 2013 and later, after the work described here, published in *The Lancet* in 2016.^{4,5}

As a consequence of the results of ICON6, AZ restarted drug manufacturing and initiated preparation of a marketing authorization application, initially in the European Union with the European Medicines Agency (EMA), and also potentially also in the USA with a submission to the Food and Drug Administration (FDA).⁶ The lack of benefit seen in other tumor types meant that, despite its limited size, ICON6 would be the single pivotal randomized phase III trial in the submission. It was to be supported by NCI8348, a randomized phase II trial in 90 patients demonstrating a PFS improvement for cediranib plus olaparib over olaparib alone in the same setting of relapsed ovarian cancer.^{7,8} ICON6 had been conducted using a long-established risk-based monitoring approach that is common in clinical trials sponsored by our academic CTU and is consistent with recent

approaches advocated by EMA/FDA.⁹ However, this is very different to typical company-sponsored monitoring practices and in any case, plans for on-site monitoring in ICON6 had been weighted towards triggered for-cause visits in the middle to later stages of the trial, so had been infrequent at the time recruitment was prematurely terminated. Consequently, the trial team in conjunction with AZ instigated a retrospective data verification process. Retrospective data verification was undertaken within three broad themes of Source Data Verification (SDV), Quality Control checks (QC) and Blinded Independent Central Review of scans (BICR) as it was judged that the regulator would require such processes. In this manuscript we report the impact of these verification processes on the key efficacy and safety results.

Methods

After presentation of the primary results, the CTU and AZ agreed how ICON6 might be prepared to be the pivotal trial within a regulatory submission. Key areas, and our proposed actions, were:

- a. Source Data Verification – checking of the completed paper case report forms (CRFs) against the source patient notes at site
- b. Quality Control checks – verification of the data entry of the paper CRFs onto the trials unit's database
- c. Blinded Independent Central Review – independent reassessment of the patient's locally-held tumor scans to adjudicate the time point of progressive disease.

All analyses retained the same data freeze date of 19-Apr-2013, with only data prior to this cut-off contributing to the primary analysis. This revised database was used to assess changes in the primary outcome PFS, and the four key toxicities emphasized in the initial presentation due to the nature of this class of drug.⁴

Source Data Verification

Centralized, monitoring based on set risk factors was planned to be used throughout the trial, but supplemented with on-site visits of all UK sites at least once during the course of the trial in addition to any triggered visit. However as the trial was shortened due to drug supply this was not possible. Details of the monitoring plan, and updates to it, is given in the Appendix.

Checking visits for retrospective Source Data Verification (SDV) were fulfilled by a third-party Contract Research Organization (CRO) arranged by AZ across all countries over the space of seven months. This typically involved two Clinical Research Associates (CRAs) visiting each site for 1—2 days and totaling 594 visits. This required a large amount of coordination both at CTU and individual sites, because sites had to retrieve the relevant patient notes and the completed trial CRFs prior to each visit. During SDV the CRF used was compared against the patient notes. Any discrepancies between the source patient hospital notes and the CRF were documented. These discrepancies were then resolved by the site via corrective action on a Data Clarification Form or the completion of a new CRF, which was then sent to the CTU for data entry.

This process of comparing the source notes and the CRF was completed for data critical to the reporting of the trial's key outcome measures. These data included select fields of the Screening, Final Safety Visit, Follow Up and progression forms, and all data on the Tumor Assessment, Adverse Event (AE), Serious Adverse Event (SAE), End of Trial Drug Summary, and Death forms. Further detail is given in the Appendix.

Quality Control checks

Quality Control (QC) checks were performed to ensure that the data entered on to the CTU's database accurately represented the completed CRF sent from the site. This process was deemed prudent as the CTU used trained single data entry as per local

policy. The CTU's policy of trained single data entry is based on the conclusion that when automatic database validation checks and statistical error checking are used, there is only a small marginal gain in accuracy from doubling of the data entry time and cost.^{10,11} Whereas for many company-sponsored studies double data entry would be used.

This second aspect of checking was initiated once the SDV was completed with all queries resolved and closed, and any new CRFs entered onto the database. The data extracted was compared to a scanned copy of the paper CRF and any related documentation. A QC error was defined as any difference between the CRF and the study database where there was no supporting documentation for the inconsistency. The error rate was calculated using the total number of valid, newly reported discrepancies divided by the total number of data points checked.

The process of checking the CRF against the database comprised of two distinct components, performed in parallel:

- a. *Key data* – data points critical to the primary efficacy analysis were examined for errors in all 486 patients.
- b. *Supporting data* – data points contributing to the detailed reporting of the trial were examined for errors in 25 of 486 patients (5% sample).

Data points deemed to be *key data* by the Trial Management Team and so checked for all patients were:

- Screening form – date of randomization

- Follow Up form – survival status, date last seen alive
- Progression form – date of progression
- Death form – date of death
- SAE form – main symptom, grade, date of onset, causal relationship to SAE, expectedness, action taken due to SAE

Data points deemed to be *supporting data* were more extensive and the same as reviewed in the SDV, these are described further in the Appendix. The 5% sample of patients consisted of: two patients from each of the five highest-recruiting sites in the UK; plus 12 patients from the remainder of the UK sites (22 of 380 UK recruited patients); and one patient randomly selected from each of the non-UK collaborating nations, Canada, Australia and Spain, who recruited respectively 87, 17 and 2 patients. An initial 5% sample of patients, expanded if necessary, was done to ease the burden on sites given the age of the trial and the retrospective nature of the verification.

It was decided that for *key data* errors, an error rate of >0.5% (a common threshold in company-sponsored trials¹²) would signal issues with the trial conduct, and for *supporting data* errors would warrant further investigation with checking promptly expanded into 100% of trial participants. All errors identified in the course of both QC processes were corrected.

Blinded Independent Central Review of progression scans

Blinded Independent Central Review (BICR) was coordinated by AZ using a specialist CRO to assess the study imaging, and to determine an overall cancer tumor assessment at each visit using RECIST criteria.^{13,14} Slight differences were anticipated as reviewing scans is subjective to an extent, and within the trial investigators could also declare progression based on ‘general deterioration’ alone.

Unlike SDV and QC, the BICR was a sensitivity analysis of the primary efficacy outcome (PFS) rather than a refinement. The BICR analysis subset, limited to patients who had at least one scan or radiological report provided for review, was used to assess ascertainment bias and comprised 96% of the all patients randomized intention-to-treat population.

There was a minimum of six independent, qualified, board-eligible radiologists assigned to the evaluation. These reviewers assessed the study imaging to determine overall tumor assessment at each time point using RECIST criteria. Each patient’s data was reviewed blinded to treatment allocation by two reviewers using two methods:

1. Primary review of each patients tumor assessment at each time point
2. Global review of all time points combined

If the two independent reviewers disagreed a third reviewer adjudicated.

Results

SDV/QC data point changes

The site visits for SDV took place over seven months, March—September 2014. 594 individual CRA visit days were completed in total across all 63 recruiting sites. SDV was completed in full for 444 of 456 (97%) patients on all three arms. The remaining 11 patients (3%) notes were irretrievable, but almost all could be partially monitored using annotations, laboratory reports and scan records. 3253 queries were raised in total which is an average of 6.7 per patient. It took on average 1.2 days to complete the required checks for each of the 486 patients recruited to the trial. Individual changes in data points for the SDV were not impractical to isolate, though were recorded clearly for the QC and so these changes are presented.

Changes arising from the QC were:

- a. *Headline-critical* – 7686 data points were QC'd and 36 errors were identified, which is an error rate of 0.47%.
- b. *Reporting-critical* – 25 patients of the total 486 were checked. 59,261 data points were QC'd and 109 errors were identified, which is an error rate of 0.18%.

The *key data* QC, in 100% of patients, discovered 30 errors related to SAE reporting (5% of 571), five to follow up and one to death (0.4% of 225). The death-related error was a data entry error where the date of death was recorded as the date of CRF completion. The five errors related to the follow up form were a variety of data entry errors, mainly a one

digit error in the date last seen entry. The most common error of the 30 in the SAE forms were 10 errors in the action taken in regards to trial treatment in response to the SAE, and eight related to the causal relationship of the trial treatment to the SAE. The others related to expectedness (four), date of onset (three), SAE grade (three) and SAE name (two). No expedited reporting of Serious Adverse Reactions to sponsor or regulators was affected.

In terms of the *supporting data* QC, in a 5% sample of patients, 109 errors were detected in 59,261 checked data points (0.18%). The vast majority of these errors (78%, 85/109) were related to reporting of AEs. 29% of the errors were misreported grade of severity. These errors are concerning given the critical nature of safety reporting, though with up to 71 grades to be entered on a single extensive AE form; over numerous visits at baseline, chemotherapy and follow up phases; and with 25 patients selected for review this error rate is extremely low in totality.

SDV/QC efficacy impact

It is not feasible to isolate and assess the efficacy impact of the SDV and the QC separately; consequently the results of these two processes will be presented together. Full statistics are provided in Table 1.

After SDV and QC, two previously unreported progression events were identified: one on each arm (A and C), resulting in a 1% increase in event rate in each arm. This actually represented the observation of three more progression events prior to the date of data cut-

off, but one progression event was recorded in error pre-SDV because the patient had asked for no further follow up on the trial. Overall, 34 changes to event time were made, 11 in arm A and 23 in arm B (2:3 randomization ratio), with a median change of -7 days for arm A and -3 days for arm C.

The log-rank test for equality of the curves provided very strong evidence of a difference between arm A and arm C both initially and after SDV/QC, with a change only at the sixth decimal place.

For the Hazard Ratio (HR) there was only a minor adjustment on one side of the confidence interval (CI). The restricted mean survival time (RMST) estimated gain remained at 3.1 months with small adjustments in the estimates of the two arms. In the Statistical Analysis Plan it was stipulated that if there was evidence of non-proportionality the RMST would be used as the primary point estimate of the treatment effect, and the HR if there was no evidence, though both estimates would be reported as supportive evidence.¹⁵ The Grambsch-Therneau test for non-proportionality p-value change did result in an adjustment in interpretation, and so consequently the primary description of the difference became the HR over the RMST.

Median progression-free survival times shortened slightly in arm C but remained with a 2.4 month gain in time to progression or death.

The conclusion drawn for the primary outcome measure of PFS and in the primary comparison of arm A vs. arm C was therefore unchanged from the initial presentation.

In Figure 2 the primary visual comparison of arm A vs. C as a Kaplan-Meier plot is presented, clearly illustrating the marginal effect on the primary efficacy endpoint following the SDV and QC processes.

Blinded Independent Central Review (BICR) of scans

The number of patients included in the analysis set in the Blinded Independent Central Review (BICR) was reduced by 4% in each arm in comparison to the local evaluation (LE), due to only including those patients with at least one scan or radiological report provided for review. This meant the proportion of events was reduced in absolute terms by 5% in arm A and by 3% in arm C. In arm A one patient (1%) had a newly identified progression and eight patients (7%) were classed as not having progressed – contrary to the primary analysis. The arm C was rather similar in that two patients (1%) had progressions added and 13 (8%) had theirs omitted. Overall the concordance rate in ICON6 between the LE and the BICR-based analyses was 19%, which is comparable to other contemporary trials that have reported their concordance such as AGO-OVAR 16 (15%), GOG-0218 (23%), OCEANS (26%) and AURELIA (31%).¹⁶⁻¹⁹ Again, all statistics are presented in Table 1.

The strength of evidence for a PFS difference between the treatment arms using the log-rank test in the BICR analysis decreased slightly, though still provided very strong evidence of a difference and should not be a surprise given the decrease in both population size and event rate. This difference in survival curves is presented graphically in Figure 3. The drop in the number of events in both arms, and consequently the lift in the curves in each arm is clear, but the overall consistency of the difference in treatments over time is comparable (described in appendix). The test for non-proportionality p-value was different to the initial presentation and very similar to the post-SDV/QC data, consequently the HR was used as the primary estimate of the difference. The size of the HR benefit of arm C was similar in both the LE and BICR, but marginally stronger in the LE. The RMST for the BICR-based analysis followed the same pattern of a very slight reduction in the estimated gain.

Toxicity reporting impact

Toxicity on the trial was collected using Common Terminology Criteria for Adverse Events (CTCAE) from the National Cancer Institute (NCI).²⁰ Grades within this system broadly correspond to the following: Grade 1 ‘Mild’; Grade 2 ‘Moderate’; Grade 3 ‘Severe or medically significant but not immediately life-threatening’; Grade 4 ‘Life-threatening consequences’; and Grade 5 ‘Death related to AE’.

During the conduct of ICON6, the NCI revised the CTCAE scales, updating from version 3 to version 4. This occurred on 28-May-2009 by which time 71 of the eventual 486

patients had been recruited. The guidance for sites participating in the trial was that patients should continue with the same version of the CRF (and consequently same version of CTCAE) with which they started on the trial to enable comparison with baseline grades given the change in some key toxicities definitions. This meant 15% (71/486) should have assessed patients using version 3, with the rest using version 4.

The four key toxicities associated with VEGF treatments and highlighted in the primary presentation are given in Figure 4.⁴ This figure displays the highest grade of each toxicity experienced on trial, separately for the two phases (during chemotherapy and on maintenance). As shown by the left hand four plots, as expected, patients report very high levels of all four toxicities during the chemotherapy phase. Comparing the 'original' and 'revised' bars, after SDV/QC there were only very small uplifts in maximum grade reported in diarrhea, nausea and fatigue – but some important differences within hypertension. This pattern of a lack of difference for diarrhea, nausea and fatigue but increase in hypertension was also observed in the maintenance phase, as shown by the right hand four plots, and so warrants further discussion. Within both the chemotherapy and maintenance phases 12% of patients had their maximum hypertension grade altered and all of these changes were uplifts in grade. The majority of the hypertension changes were due to Blood Pressure (BP) readings in the source notes (70%) or the start of anti-hypertensive treatment (17%) which had not been reported on the CRFs for the trial. There was no evidence of a systematic bias by arm.

We had informal discussions with investigators around the reasons for BP readings not being recorded on CRFs. The conclusion being that this was a specific circumstance where the definition had changed and staff were more familiar with the older CTCAE version 3 classification, substantially based on need for changing therapeutic regimens, whereas the updated version 4 mandated a change in grade based on firm numerical guidelines. Many clinicians would have been comfortable with the older system and graded according to it. For example, a patient observed in clinic with a one-off BP reading of systolic 140 or diastolic 90 previously would have been ungraded, but in the revised version would have been assigned grade 2 hypertension. The consequence being that some elevated BP readings that met version 4 criteria were not always recorded as hypertension AEs, but CRAs completing the SDV identified these as additional AEs. This may have been particularly so when the clinician considered the BP reading was likely to be related to stress in clinic.

In the original trial report, AEs were formally compared if they were experienced by >5% of the trial population (described in the preceding paragraph), or if they were high grade ('life-threatening' 4, or 'fatal' 5) of any frequency. Within this high grade AE reporting group, only two occurrences were altered. For one patient their pneumonia was revised from grade 4 to grade 5 (arm A), and one grade 3 pancreatitis revised to grade 5 (arm C). Both had been appropriately reported as fatal SAEs, and given the difficult nature of assigning a definitive cause of death to a specific symptom minor changes are not surprising. There was a downgrading from grade 4 to grade 3 hypomagnesemia for one arm C patient.

Conclusion

We found that the three themes of retrospective data verification (SDV, QC and BICR) made only immaterial changes to the key outcome measures of the trial – and consequently its interpretation.

The two most time-consuming parts, the SDV and QC, had the least impact on the primary efficacy outcome measure. The impact of SDV was consistent with previously reported experiences, one of which states ‘the true effectiveness value of SDV [...] is minimal (0.1%–0.4%)’²¹ and another that ‘discrepancies identified made no impact on the main conclusions of the study’.²² The QC process did discover errors, in the *key data* this almost met the 0.5% threshold we set that would signal concern with trial conduct, but overall compared well to error rates in the literature and again made no impact on the interpretation of the study.²³ The BICR sensitivity analysis, on the other hand, did make a difference to individual assessments, as was expected given the complexity and partially subjective nature of declaring progression in cancer patients, especially retrospectively. This sensitivity analysis supported the primary, local evaluation, result in that the overall test of a difference and the size of the estimate remained overwhelmingly positive. Consequently, the local evaluation conclusion that cediranib increases time to progression or death was reinforced. The literature contains much discussion as to the limited added value of BICR already, and suggests that local evaluation ‘provides a reliable estimate of treatment effect’ in most settings in the study of cancer.^{24,25}

As anticipated there were some changes in the toxicity profile. For the most part, these were minor and made little or no difference to the standard toxicity summary statistics, and does not appear to have been systematically biased towards one treatment arm. Some additional hypertension episodes were reported, though in discussion with treating gynecological oncologists we feel that this is likely due to substantial differences in the method of grading hypertension between versions of the AE system used. This meant that what many interpreted to be one-off, non-clinically significant BP readings were graded as a low grade AE in updated CTCAE guidance – for which a subset of investigators had not realized had been revised. This substantial difference in definition of hypertension between versions is emphasized by Akhtar et al. who, in study NCI 6981, observed the majority of patients (65%) increased by two grades in their retrospective regrading using CTCAE version 4. They concluded the impact of this change as ‘particularly important with VEGF-targeted agents’ such as cediranib.²⁶ It is our belief that these changes would not have altered the overall toxicity assessment as the changes were not biased towards one arm, it was already acknowledged that the toxic effects of VEGF inhibitors can be problematic, and prompt management guidelines were generally successful in controlling further hypertension.

A risk-based monitoring approach was used in the ICON6 trial, as is standard at the CTU given studies have demonstrated it to be non-inferior to extensive on-site monitoring.²⁷ The EMA and FDA published their support for alternative approaches such as centralized, risk-based monitoring in 2013 but the practice of 100% SDV was, and is still likely to be, commonplace in company-sponsored trials.^{9,28,29}

This retrospectively-performed work took very substantial amounts of time. Our crude estimate is that it was the primary task for our trial team for two years, equivalent to 4.5 full time staff members per year. External contractors were used for more focused periods but we estimate this to be equivalent to 9 full time staff per year. It is likely that retrospectively conducting these processes would be more time-consuming than if done prospectively.

The conduct of this pragmatic, academic-led trial was sufficient given the robustness of the results, shown by the results remaining largely unchanged following retrospective verification. The burden of the retrospective effort required to ensure the results of ICON6 were acceptable to regulators is difficult to justify.

Acknowledgements

The ICON6 trial team would like to thank patients and their families; Investigators (45 UK sites, 10 Canada, 5 Australia, 2 Spain); collaborating GCIG Groups (CCTG, ANZGOG, GEICO); TMG members and staff at MRC Clinical Trials Unit at UCL; IDMC members; TSC members; and AstraZeneca. Also, Sharon Love for providing a critical review of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article

Grant support

Funding received for ICON6 from Medical Research Council (UK); Cancer Research UK (CRUK/07/025); Canadian Cancer Society Research Institute (015469, 021039); Cancer Australia (APP1006602) and National Gynecological Cancer Centre; and AstraZeneca.

References

1. Goss GD, Arnold A, Shepherd FA, et al. Randomized, Double-Blind Trial of Carboplatin and Paclitaxel With Either Daily Oral Cediranib or Placebo in Advanced Non-Small-Cell Lung Cancer: NCIC Clinical Trials Group BR24 Study. *J Clin Oncol* 2010; 28: 49–55.
2. AstraZeneca. AstraZeneca Announces Results of Recentin HORIZON II Phase III Trial in Metastatic Colorectal Cancer, <https://www.astrazeneca.com/media-centre/press-releases/2010/AstraZeneca-Announces-Results-of-Recentin-HORIZON-II-Phase-III-Trial-in-Metastatic-Colorectal-Cancer-28052010.html> (2010, accessed 24 March 2017).
3. AstraZeneca. *Annual Report*. 2011.
4. Ledermann J, Perren T, Raja F, et al. Randomised double-blind phase III trial of cediranib (AZD 2171) in relapsed platinum sensitive ovarian cancer: Results of the ICON6 trial. *Eur J Cancer* 2013; 49: S5 (LBA10).
5. Ledermann JA, Embleton AC, Raja F, et al. Cediranib in Patients with Relapsed Platinum-Sensitive Ovarian Cancer (ICON6): A Randomised, Double-Blind, Placebo-Controlled Phase 3 Trial. *The Lancet* 2016; 387: 1066–1074.
6. Schmidt C. Cediranib Aims for a Comeback. *J Natl Cancer Inst* 2015; 107: djv068.
7. Liu J, Barry WT, Birrer MJ, et al. A randomized phase 2 trial comparing efficacy of the combination of the PARP inhibitor olaparib and the antiangiogenic cediranib against olaparib alone in recurrent platinum-sensitive ovarian cancer. *J Clin Oncol* 2014; 32: LBA5500.
8. Liu JF, Barry WT, Birrer M, et al. Combination cediranib and olaparib versus olaparib alone for women with recurrent platinum-sensitive ovarian cancer: a randomised phase 2 study. *Lancet Oncol* 2014; 15: 1207–1214.
9. U.S. Department of Health and Human Services Food and Drug Administration. *Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring*. August 2013.
10. Gibson D, Harvey AJ, Everett V, et al. Is double data entry necessary? The CHART trials. CHART Steering Committee. Continuous, Hyperfractionated, Accelerated Radiotherapy. *Control Clin Trials* 1994; 15: 482–488.
11. Day S, Fayers P, Harvey D. Double Data Entry: What Value, What Price? *Control Clin Trials* 1998; 19: 15–24.

12. Pomerantseva V, Ilicheva O. Clinical Data Collection, Cleaning and Verification in Anticipation of Database Lock: Practices and Recommendations. *Pharm Med* 2011; 25: 223–233.
13. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000; 92: 205–216.
14. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228–247.
15. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med* 2011; 30: 2409–2421.
16. Floquet A, Vergote I, Colombo N, et al. Progression-free survival by local investigator versus independent central review: Comparative analysis of the AGO-OVAR16 Trial. *Gynecol Oncol* 2015; 136: 37–42.
17. Burger RA, Brady MF, Rhee J, et al. Independent radiologic review of the Gynecologic Oncology Group Study 0218, a phase III trial of bevacizumab in the primary treatment of advanced epithelial ovarian, primary peritoneal, or fallopian tube cancer. *Gynecol Oncol* 2013; 131: 21–26.
18. Aghajanian C, Goff B, Nycum LR, et al. Independent radiologic review: Bevacizumab in combination with gemcitabine and carboplatin in recurrent ovarian cancer. *Gynecol Oncol* 2014; 133: 105–110.
19. Husain A, Wang Y, Hanker LC, et al. Independent radiologic review of AURELIA, a phase 3 trial of bevacizumab plus chemotherapy for platinum-resistant recurrent ovarian cancer. *Gynecol Oncol* 2016; 142: 465–470.
20. National Cancer Institute. *Common Terminology Criteria for Adverse Events v4.0*. NIH publication # 09-7473, NCI, NIH, DHHS, 29 May 2009.
21. Tantsyura V, Dunn IM, Fendt K, et al. Risk-Based Monitoring: A Closer Statistical Look at Source Document Verification, Queries, Study Size Effects, and Data Quality. *Ther Innov Regul Sci* 2015; 49: 903–910.
22. Tudur Smith C, Stocken DD, Dunn J, et al. The Value of Source Data Verification in a Cancer Clinical Trial. *PLoS ONE* 2012; 7: e51623.
23. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. *AMIA Annu Symp Proc* 2008; 242–246.
24. Pignatti F, Hemmings R, Jonsson B. Is it time to abandon complete blinded independent central radiological evaluation of progression in registration trials? *Eur J Cancer* 2011; 47: 1759–1762.

25. Amit O, Mannino F, Stone AM, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis. *Eur J Cancer* 2011; 47: 1772–1778.
26. Akhtar NH, Karpenko I, Singh B, et al. Association of CTCAE v4 grading of hypertension with toxicity in patients with renal cancer receiving vascular endothelial growth factor (VEGF)-targeting agents. *J Clin Oncol* 2013; 31: 447.
27. Brosteanu O, Schwarz G, Houben P, et al. Risk-adapted monitoring is not inferior to extensive on-site monitoring: Results of the ADAMON cluster-randomised study. *Clin Trials J Soc Clin Trials* 2017; 14: 584–596.
28. European Medicines Agency. *Reflection paper on risk based quality management in clinical trials*. EMA/269011/2013, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf (18 November 2013).
29. Eastern Research Group. *Examination of Clinical Trial Costs and Barriers for Drug Development*, https://aspe.hhs.gov/system/files/pdf/77166/rpt_erg.pdf (25 July 2014).