Lessons learned from EVOLVE for the planning of future global randomized trials in chronic kidney disease

Parfrey PS*, Block GA[†], Correa – Rotter R[ǂ], Drueke TB[§], Floege J[‖], Herzog CA[¶], London GM[**], Mahaffey KW[††], Moe SM[§§], Wheeler DC[‖‖], Chertow GM[††]

Patrick S. Parfrey*	Department of Medicine, Health Sciences Center, St. John's, NL, Canada
Geoffrey A. Block[†]	Denver Nephrology, Denver, Colorado
Ricardo Correa-Rotter[ǂ]	Department of Nephrology and Mineral Metabolism, Salvador Zubirán National Institute of Health Sciences and Nutrition, Mexico City, Mexico
Tilman B. Drüeke[§]	French Institute of Health and Medical Research Unit 1088, Faculty of Medicine and Pharmacy, University of Picardie, Amiens, France
Jürgen Floege[‖]	Department of Nephrology, RWTH Aachen University Hospital, Aachen, Germany
Charles A. Herzog[¶]	Department of Internal Medicine, Hennepin County Medical Centre, University of Minnesota, Minneapolis, Minnesota
Gerard M. London**	Service of Nephrology, Manhès Hospital, Paris, France
Kenneth W. Mahaffey[††]	Department of Medicine, Stanford University School of Medicine, Palo Alto, California

Sharon M. Moe[§§] Department of Medicine, Indiana University School of Medicine and Roudebush Veterans Administration Medical Center, Indianapolis, Indiana

David C. Wheeler[‖] Centre for Nephrology, University College London, United Kingdom

Glenn M. Chertow[††] Department of Medicine, Stanford University School of Medicine, Palo Alto, California

Abstract: 193

Text: 3379

Table: 5

Fig: 2

References: 33

Abstract

The effect of the calcimimetic cinacalcet on cardiovascular disease in patients undergoing hemodialysis with secondary hyperparathyroidism (sHPT) was evaluated in the EVOLVE trial. This was the largest (in size) and longest (in duration) randomized controlled clinical trial undertaken in this population. During planning, execution, analysis and reporting of the trial many lessons were learned, including those related to the use of a composite cardiovascular primary endpoint, definition of endpoints (particularly heart failure and severe unremitting HPT), importance of age for optimal stratification at randomization, use of unadjusted and adjusted intention-to-treat analysis for the primary outcome, how to respond to a lower than predicted event rate during the trial, development of a pre-specified analytic plan that accounted for non-adherence and for co-interventions that diminished the power of the trial to observe a treatment effect, determination of the credibility of a subgroup effect, use of adverse effects database to investigate rare diseases, collection of blood for biomarker measurement not designated prior to trial initiation, and interpretation of the benefits to harms ratio for individual patients. It is likely that many of these issues will arise in planning of future trials in chronic kidney disease.

Introduction

The Evaluation of Cinacalcet Hydrochloride Therapy to Lower Cardiovascular Events (EVOLVE) trial was the largest (N = 3883) and longest (median follow-up 50.5 months) randomized controlled trial (RCT) undertaken in patients receiving hemodialysis (1). The hypothesis tested was that the calcimimetic cinacalcet, compared to placebo, would reduce the risk of death or nonfatal cardiovascular events among patients with sHPT (intact parathyroid hormone (iPTH) ≥ 300 pg/mL) who were undergoing hemodialysis. In the primary unadjusted intention-to-treat (ITT) analysis, cinacalcet did not significantly reduce the risk of the composite cardiovascular (CV) primary outcome, but there were a number of design and analysis issues that should be considered when interpreting the results. In this review we discuss study challenges faced, decisions made and lessons learned from EVOLVE, and how they may be relevant in the design of future trials in dialysis and chronic kidney disease (CKD).

Choice, definition and analysis of primary events.

***Composite Primary Outcome*.** The primary endpoint was a composite comprised of all-cause mortality or non-fatal CV events (myocardial infarction, hospitalization for unstable angina, heart failure or peripheral vascular disease). These CV events may occur via different pathophysiological pathways: myocardial infarction and unstable angina are principally caused by atherosclerotic

vascular disease, while heart failure results from cardiomyopathy (with systolic or diastolic dysfunction) for which arteriosclerosis and uremia are predisposing factors (2). Peripheral vascular disease is often atherosclerotic, but distal occlusive disease, particularly of the lower extremities, is often accompanied by dense calcific arteriosclerosis. The CV events were chosen because we hypothesized that cinacalcet would (A) reduce heart failure and death by reducing medial calcification of conduit arteries (thus improving vascular compliance and decreasing LV hypertrophy) and by decreasing the potential cardiotoxic effect of PTH, (B) reduce atherosclerotic events by decreasing intimal calcification of atherosclerotic stenoses (3). If the endpoint had been mortality alone a much larger sample size would have been necessary.

Inclusion of overall mortality in the composite outcome is an issue because CV mortality would likely be modified by cinacalcet, in contrast to non-CV mortality. We included overall mortality in the primary outcome because we thought it necessary to demonstrate that the intervention did not cause unanticipated serious adverse effects on non -CV outcomes (such as cancer or death after fracture).

Post-hoc evaluation of the treatment effects on CV events using multi-variable adjusted ITT analysis showed a relative hazard of 0.88 (95% confidence intervals (CI): 0.76 to 1.01) for time to first atherosclerotic event, 0.79 (95% CI 0.66 to

0.96) for heart failure, and 0.79 (0.64 to 0.98) for sudden death, the latter 2 events considered to be outcomes of non-atherosclerotic disease (4). The 95% CI of the atherosclerotic and non-atherosclerotic outcomes overlap, but the magnitude of the relative hazard for heart failure and sudden death are consistent with the hypothesis that cinacalcet may act through the non-atherosclerotic pathway. As medial calcification and intimal calcification may well be different entities in CKD (5), this suggests that the dominant effect of cinacalcet may be through the inhibition of medial calcification.

***Definition of outcomes*.** Clinically relevant, precisely defined, reproducible end points with central adjudication are necessary in RCTs that aim to change clinical practice (6). Death and major atherosclerotic CV events fulfill these criteria, but the definition of heart failure in CKD has been problematic because of the difficulty in differentiating salt and water overload from impaired left ventricular function. This issue was recently crystalized by the proposal for a functional classification system of heart failure in patients with end-stage renal disease (ESRD) (7). This system may be limited by the cost of echocardiography and lack of specificity of dyspnea relief by dialysis for diagnosis of heart failure.

In patients receiving dialysis, heart failure has been defined as dyspnea with at least two of the following four manifestations: bilateral basilar rales on physical exam, raised jugular venous pressure, interstitial edema on chest x-ray, and

increased upper pulmonary vessel diameter on chest x-ray (8). The presence or history of heart failure was associated with an almost 2-fold increase in mortality (8). In EVOLVE, we used a similar definition to that of Harnett et al. (8) but added an additional clause to improve specificity: the patient also was required to have received mechanical ultrafiltration or hemodialysis. Nonetheless, the event adjudication committee confirmed the diagnosis of heart failure in only about half of heart failure events submitted by the local investigators, a result lower than that observed in high risk patients with hypertension (9). This lack of certainty in the diagnosis of heart failure may limit study power because the event rate may be lower than anticipated, and misclassification may limit the capacity to identify a treatment effect. A similar misreporting of myocardial endpoints occurred in the PARAGON-B and PURSUIT trials where investigators misreported myocardial infarction and points, but most later agreed with clinical events committee assessments (10, 11).

Endpoint with selection bias. Certain end points for mineral and bone disease in CKD are also problematic. For example, the endpoint for severe HPT could be parathyroidectomy (PTX) but severe PTH may or may not be treated with PTX. PTX was a secondary endpoint in EVOLVE but prior to starting the study it was evident that the criteria by which physicians selected PTX for the treatment of sHPT varied widely across the world, and no criteria for PTX were provided by the

trial protocol. Factors associated with PTX during the trial indicated selection effects and included younger age, female sex, higher body mass index, markers of co-morbidity (no history of valvular disease or angina or peripheral vascular disease), higher serum calcium and higher PTH; the use of PTX also varied widely by country of origin (lowest rates in USA) (12). In addition, PTX was performed at an advanced phase of sHPT:mean PTH was 1872 pg/mL and serum calcium 10.3 mg/dL (2.58 mmol/L) prior to PTX (12). As there were likely to be additional patients who were not selected for PTX who might have undergone PTX but did not because of various reasons unrelated to disease severity (e.g., surgeon or patient reluctance, no trained surgeon available), we defined an outcome before the trial was completed which we termed severe unremitting HPT, which included PTX or the presence of severe HPT with hypercalcemia (iPTH > 1000 pg/mL with serum calcium > 10.5 mg/dL (2.61 mmol/L) on two consecutive occasions or iPTH > 1000 pg/mL with serum calcium > 10.5 mg/dL (2.61 mmol/d) on one occasion and subsequent use of commercial (off-protocol) cinacalcet within two months of laboratory assessment). Although hypercalcemia is a biochemical endpoint its occurrence implies an advanced phase of HPT, and clinical practice supports PTX in this scenario. In the placebo group, severe unremitting HPT occurred in 24.3% (n=470), of whom 59% (n=278) had PTX. The relative hazard (cinacalcet *versus* placebo) for severe unremitting HPT was 0.43 (95% CI 0.37 to 0.50). This relative

hazard was similar whether baseline iPTH was mildly or markedly elevated but the number of events prevented increased as iPTH increased (12).

Lessons from EVOLVE concerning choice and definitions of endpoints are presented in **Table 1**.

Imbalance in baseline clinical characteristics. Despite enrolling 3883 patients there was a 0.8 years difference in mean age at baseline and a 1 year difference in median age (55 *versus* 54 years), an occurrence that confounded the primary outcome (1). This chance imbalance in a major prognostic factor for CV events necessitates covariate adjustment (13). In the pre-specified primary analysis using an unadjusted intention-to-treat (ITT) analysis, the relative hazard was 0.93 (95% confidence interval (95% CI): 0.85 to 1.02), and the pre-specified multivariable adjusted ITT relative hazard was 0.88 (0.80 to 0.98). It is likely that the observed age imbalance occurred by chance as the likelihood of imbalance is dictated by the population standard deviation (SD) for age and sample size. The probability of an age difference of >0.8 years occurring in EVOLVE was 0.08, as the SD was 14 and sample size 3883. Owing to more restrictive inclusion criteria and older patients enrolled, the SD for age was narrower in SHARP (12 years) (14) and TREAT (10 years) (15), where the probability of observing an age difference (by chance) of >0.8 years between groups was 0.04 and 0.01, respectively.

CONSORT recommends reporting both unadjusted and adjusted analyses and stating whether the adjusted analysis was planned (16). Although substantial variation exists in handling of baseline covariates in RCTs (17), covariate adjustment improves treatment effect estimation accuracy and statistical power, and hence should be performed when strong prognostic factors are observed or anticipated (11, 18).

Lessons concerning imbalance in baseline covariates are presented in **Table 2**.

Threats to statistical power

The major causes of reduction in statistical power in EVOLVE were a lower than predicted primary composite event rate, discontinuation of study drug, particularly due to adverse effects in the cinacalcet group and failure to achieve control of sHPT in the placebo group, and PTH-lowering co-interventions that were applied disproportionately in the placebo group (i.e., parathyroidectomy, kidney transplantation, and use of commercial cinacalcet) (Figure 1).

***Event rate*.** We calculated the sample size on the basis of the following assumptions: an annual rate of the primary composite endpoint of 23.2% in the placebo group, a 20% treatment effect, a 1.5-year enrollment period, a 4-year total study duration, an annual rate of loss to follow-up of 1%, an annual rate of dropout (withdrawal from active treatment before a primary event) of 10% in the cinacalcet

group, and a rate of drop-in (use of commercially available cinacalcet before a primary event) of 10% in the placebo group. On the basis of a two-sided log-rank test for equality of survival functions, accounting for planned interim analyses with an overall alpha level of 0.05, we determined that a primary event would need to occur in 1882 patients in order to ensure a power of approximately 90% (1).

We estimated event rates by integrating data from phase 3 short- and medium-term RCTs comparing cinacalcet to placebo, as well as the observational data linking expected changes in PTH, calcium, and phosphorus within components of the primary composite endpoint. After it became apparent that the overall (blinded) event rate was below 20.8%, we extended the trial by 16 months to allow for accrual of the requisite number of events. The actual annual event rate in the placebo group was 15.5% (19).

The event rate in the placebo arm of RCTs is frequently lower than predicted (20). When it became apparent that the primary composite event rate in EVOLVE was below that anticipated, the choice of intervention was to increase the number of enrolled subjects or to extend the duration of trial follow-up. The first choice would have entailed re-engaging enrollment teams in dialysis units already extended by their contribution to EVOLVE and would have been more costly. The decision was made to extend follow-up. While additional events were expected to accrue over time, adherence with the intervention waned and more PTH-lowering

interventions were instituted (i.e., kidney transplantation, PTX and use of commercial cinacalcet) particularly in the placebo group. High attrition of subjects from and drop-in to the intervention group eroded the original planned power of EVOLVE in such a manner that extension of the trial was of little benefit in identifying a treatment effect.

Lessons:

1. Event rate estimates in the sample size calculation should be conservative.
2. If the event rate is lower than predicted and non-adherence during the trial (especially over time) is likely, then enrollment of more patients is generally preferable to extending follow-up.

***Non-adherence*.** In the cinacalcet group, median time in the trial was 50.6 months and time on drug was 21.2 months. Corresponding times for the placebo group were 50.4 and 17.5 months. Time to first discontinuation of study drug for protocol-specified reasons was similar in both cinacalcet and placebo groups, but for non-protocol specified reasons it was significantly higher in the placebo group, driven by the fear of severe unremitting HPT (1). In the placebo group, drug was discontinued for adverse events in 11.8%, compared with 15.8% in the cinacalcet group. Some of this discontinuation was anticipated. The "drop-in" rate was

expected to be 10%/year and the observed rate was 7.4%/year. The "drop-out" rate was derived from discontinuation rates observed in prior cinacalcet RCTs and was higher than anticipated: expected 10% and observed 27.3% per year. This may have occurred because of the long-term nature of the trial amplifying the impact of kidney transplantation, PTX and use of commercial cinacalcet, and of comorbidity predisposing to study drug discontinuation because of adverse events and fear of severe HPT.

Drug discontinuation was a more important factor in diminishing the power of the study than the lower -than- predicted event rates .The increase in sample size necessary in the presence of x% discontinuation can be estimated as the reciprocal of (1-x) squared .Hence, with 40% discontinuation at the trial mid-point, the sample size necessary to maintain study power at the level planned was about three times bigger than planned.

An analytic plan to take account of non-adherence in the estimates of the treatment effect was pre-specified. It included lag censoring, iterative parameter estimate (IPE) and inverse probability of censoring weights (IPCW). Lag censoring analysis uses data censored at a pre-specified time point. In EVOLVE we had pre-specified the lag time to be 6 months after cinacalcet was discontinued to account for the possibility that the drug had persistent effects. Although lag censoring

preserves randomization, there may be informative bias if non-adherent patients (compared to adherent patients) have different prognostic characteristics associated with primary endpoint (21). This methodologic weakness also occurs with the IPE method. However the IPCW is not prone to informative bias (22).

The IPCW approach censors data when non-adherence occurs. For patients who were adherent and had similar characteristics to those who were not, IPCW assigns larger weights to these patients to "re-create" the population that would have been observed (22) **(Figure 2)**. Weights are calculated based on the inverse of the probability that patients remain adherent using a logistic regression model. The final relative hazard is derived from a weighted regression model. In EVOLVE age, sex, race, country, diabetes history, randomized treatment group, time dependent covariates of PTH, and the adverse events of hypocalcemia and nausea/vomiting were used in the logistic regression model to estimate the probability of adherence (22). The IPCW method is sensitive to the number of non-adherent patients, assumes there are no unknown confounders and is computationally difficult. Nonetheless, it is accepted by many health agencies. The relative hazard was 0.81 using IPCW **(Table 3)**.

***Impact of co-interventions*.** Before the trial started, it was clear that three co-interventions that lower PTH – kidney transplantation, PTX and use of commercial cinacalcet – could diminish the treatment effect of cinacalcet, particularly if they

were delivered more frequently in the placebo group, as was expected for the latter two. We considered including PTX in a composite endpoint, but were concerned about the lack of pathophysiological concordance for PTX and CV events (see above). We considered censoring follow-up time after kidney transplantation or PTX, but were concerned about introducing bias, in that patients who were too ill would be unlikely to receive either of these two interventions. Instead we pre-specified estimates of the treatment effect once follow-up time after these three PTH-lowering interventions was censored. **Table 3** shows that the relative hazard was 0.90 when censoring occurred at either of these three co-interventions, and 0.84 with censoring of time after any one of these co-interventions (1).

Table 4 outlines lessons concerning the analytic plan for non-adherence and co-interventions.

***Subgroup effects*.** We pre-specified seven subgroup analyses including age ≥65 and <65 years. The relative hazard for the primary endpoint using ITT was 0.74 (0.63 to 0.86) in the older group and 0.99 (95% CI 0.88 to 1.11) in the younger group (19). There was a 27% reduction in mortality in patients > 65 years (p <0.001). The test of treatment x age interaction was significant (p=0.03, using age as a continuous variable, p=0.007 when age was dichotomized at 65 years). The age modification of the treatment effect was partly related to (1) 3-fold higher rates of kidney transplantation and PTX in younger patients, co-interventions that

limited detection of a treatment effect, (2) lower baseline CV disease in younger patients limiting the potential of cinacalcet to decrease CV events rates, and (3) lower CV event rates that decreased the power to observe an effect (19).

Criteria by which to evaluate the credibility of subgroup effects have been published (23). **Table 5** outlines the lessons from EVOLVE concerning the age subgroup effect.

***Capture of rare diseases*.** RCTs collect accurate data on pre-determined endpoints reviewed by an event adjudication committee, but in addition substantial amounts of data are collected during drug exposure by local investigators for assessment of adverse effects. This may facilitate the study of rare diseases, particularly if the disorder has been identified in advance and has a plausible biological rationale. However care must be taken with post hoc analyses of databases with multiple events as statistically "significant" associations may occur by chance. In EVOLVE, calcific uremic arteriolopathy (CUA) occurred infrequently, but cinacalcet reduced its incidence (relative hazard 0.31, 95% CI 0.13 to 0.79) (24). Predictors of CUA included allocation to placebo, female sex, higher body mass index, hypertension, prior PTX and prior tobacco use. In addition, use of vitamin K antagonists at time of CUA was significantly higher than use in patients without CUA.

Lesson.

1. RCTs can provide important information on rare diseases, particularly when a large cohort of patients is followed in the trial at relatively high risk for developing the rare disease (25), and for which a plausible biological rationale exists.

Biomarkers. Collection and storage of blood for subsequent biomarker measurement may contribute to the understanding of disease pathogenesis (26), and in a large RCT the incremental cost of collecting patient samples is relatively low. Because of the long duration of trials (EVOLVE started in August 2006 and finished in January 2012, with primary publication of trial results in November 2012) and the potential for biomarker research to be published during the execution of the trial, we did not designate which biomarkers would be measured until after the trial was completed. From 2007-2011 data had accumulated in patients with advance CKD and ESRD that serum fibroblast growth factor 23 (FGF 23) concentrations were elevated and were associated with mortality, left ventricular hypertrophy and cardiovascular events independent of serum phosphate, PTH and a variety of demographic and other clinical factors (27, 28). Sixty four percent of patients assigned to cinacalcet had ≥ 30% reduction in serum FGF23 concentrations from baseline to week 20 compared to 28% of the placebo group (29). Among patients randomized to cinacalcet ≥ 30% reduction in FGF23

was associated with a significant reduction in the composite CV endpoint (relative hazard 0.82, 95% CI 0.69 to 0.98) (29). This non-randomized analysis generates the hypothesis that FGF23 plays a role in the pathogenesis of CV events in CKD, and the deleterious effects of FGF23 might be ameliorated by cinacalcet.

Lesson.

1. Collection of serum for biomarker investigation in a large RCT should be undertaken because it facilitates rapid testing of hypotheses generated after the RCT started, and creation of new hypotheses.

Interpretation of clinical benefits. Trials should not be judged by the result of a single analysis and a single p- value, but inferences should be made on the totality of the data. The decision to prescribe cinacalcet should be informed by consideration of the benefits to risks ratio in individual patients. Cinacalcet is approved for the treatment of secondary HPT and in EVOLVE it was effective in preventing severe unremitting HPT. Treatment effects of cinacalcet on fracture rates were similar to those on CV events using unadjusted ITT ,multi variable adjusted ITT, and censoring at co-interventions that reduce PTH;age was also an effect modifier(30).A recent review has examined the clinical and practical use of calcimimetics in dialysis patients(31).An economic evaluation of cinacalcet in the United States has been reported(32).

Conclusions

The EVOLVE trial provides multiple lessons in the planning, execution, analysis and interpretation of RCTs in CKD/ESRD. These include deciding upon, defining and analyzing outcomes, responding to lower than predicted event rates, taking account of non-adherence and of co-interventions, subgroup effects, capture of rare diseases, biomarkers, and interpretation of benefits to harms ratio.

Disclosures:

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

The EVOLVE Trial was funded by AMGEN.

References

1. Chertow GM, Block GA, Correa-Rotter R, Drüeke TB, Floege J, Goodman WG, Herzog CA, Kubo Y, London GM, Mahaffey KW, Mix CH, Moe SM, Trotman M-L, Wheeler DC, Parfrey PS.
Effect of Cinacalcet on Cardiovascular Disease in Patients Undergoing Dialysis.
N Engl J Med 367:2482-2494, 2012.

2. Parfrey PS, Barrett BJ. Cardiovascular disease and chronic kidney disease in Chronic Kidney Disease. Edited by PL Kimmel and Rosenberg ME. Pp 181-198. Elsevier. USA. 2015.

3. McCullough PA, Agrawal V, Danielewicz E, Abela GS. Accelerated atherosclerotic calcification and Monckeberg's sclerosis: a continuum of advanced vascular pathology in chronic kidney disease.
Clin J Am Soc Nephrol 2008;**3**(6):1585-98.

4. Wheeler DC, London GM, Parfrey PS, Block GA, Correa-Rotter R, Dehmel B, Drüeke TB, Floege J, Kudo Y, Mahaffey KW, Goodman WG, Moe SM, Trotman M-L, Abdalla S, Chertow GM, Herzog CA.
Effects of cinacalcet on atherosclerotic and non atherosclerotic cardiovascular events in patients receiving hemodialysis. The Evaluation of Cinacalcet HCC Therapy to lower CardioVascular Events (EVOLVE) Trial.
J Am Heart Assoc 2014; 3:001363 doi:10. 1161/JAHA. 114. 001363.

5. Amann K. Media calcification and intima calcification are district entities in chronic kidney disease.
Clin J Am Soc Nephrol 3:1599-605, 2008.

6. Seltzer JH, Turner JR, Geiger MJ, Rosano G, Mahaffey KW, White WB, Sabol MB, Stockbridge N, Sager PT. Centralized adjudication of cardiovascular endpoints in cardiovascular and noncardiovascular pharmacologic trials: A report from the Cardiac Safety Research Consortium.
Am Heart J 169;197-204, 2015.

7. Chawla LS, Herzog CA, Costanzo MR, Tumlin J, Kellum JA, McCullough PA, Ronco C, for the ADQI XI Workshop.
Proposal for a functional classification system of heart failure in patients with end-stage renal disease.
J Am Coll Cardiol 63:1246-52, 2014

8. Harnett JD, Foley RN, Kent GM, Barre PE, Murray D, Parfrey PS. Congestive heart failure in dialysis patients: Prevalence, incidence, prognosis and risk factors.
Kidney Int 47: 884-90, 1995.

9. Piller LB, Davis BR, Cutler JA, Cushman WC, Wright JT, Williamson JD, Leenan FHH, Einhorn PT, Randall OS, Golden JS, Haywood LJ, and the ALLHAT Collaborative Research Group. Validation of heart failure events in the antihypertensive and lipid lowering treatment to prevent heart attack Trial (ALL HAT) participants assigned to doxazosin and chlorthalidone. Curr Control Trials Cardiovasc Med 3:10-19, 2002.10.

10. Mahaffey KW, Roe MT, Dyke CK, Newby LK, Kleiman NS, Connolly PA, Berdan lG, Aparapanil R, Lee KI, Armstrong PW, Topol EJ, Califf RM, Harrington RA for the PARAGON B Investigators. Misreporting of myocardial infarction end points: Results of adjudication by a central clinical events committee in the PARAGON-B trial. Am Heart J 143:242-8, 2002.

11. Mahaffey KW, Harrington RA, Akkerhius M, Kleiman NS, Berdan LG, Crenshaw BS, Tardiff BE, Granger CG, DeJong I, Bhapkar M, Widimsky P, Corbalon R, Lee KL, Deckers J, Simoons ML, Topol EJ, Califf RM for the PURSUIT Investigators. Disagreements between a central clinical events committee and site investigator assessments of myocardial infarction end-points in an international clinical trial: Review of the PURSUIT study. Curr Control Trials Cardiovasc Med 2:187-94, 2001.

12. Parfrey PS, Chertow GM, Block GA, Correa-Rotter R, Drüeke TB, Floege J, Herzog CA, London GM, Mahaffey KW, Moe SM, Wheeler DC, Dehmel

B, Trotman ML, Modafferi DM, Goodman WG: The clinical course of treated hyperparathyroidism among patients receiving hemodialysis and the effect of cinacalcet: The EVOLVE trial.
J Clin Endocrinal Metab 98: 4834-4844, 2013.

13. Chu R, Walter SD, Guyatt GP, Devereaux PJ, Walsh M, Thorlund K, Thabane L. Assessment and implication of prognostic balance in randomized controlled trials with a binary outcome – a simulation study. Plos one, 2012. DOI:10, 1371/journal. pone 0036677

14. Baigent C, Landrey M, Emberson J, Wheeler DC, Tomson C. SHARP Investigators. The effects of lowering LDL cholesterol with simvastatin plus ezetimide in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomized placebo controlled trial.
Lancet 377:2182-92, 2011.

15. Pfeffer M, Burdmann EA, Chen CY, Cooper ME, de Zeeuw D, Eckardt K-E, Feyzi JM, Ivanovich P, Kemelramani R, Levey AS, Lewis EF, McGill JB, McMurray JJ, Parfrey P, Parving HH, Remuzzi G, Singh AK,Solomon SR, Toto R; TREAT investigators.
A trial of darbepoietin alfa in type 2 diabetes mellitus and chronic kidney disease.
N Eng J Med 361:2019-32, 2009.

16. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, for the CONSORT Group, et al.: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med, **134:**663-94, 2001.

17. Austin PC, Manca A, Zwarenstein M, Juurluik DN, Stanbrook MB. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. J Clin Epidemiol 63:142-153, 2010.

18. Hernandez AV, Eijkemans MJC, Steyerberg ED. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power?
Ann Epidemiol 16:41-48, 2006.

19. Parfrey PS, Drüeke TB, Block GA, Correa-Rotter R, Floege J, Herzog CA, London GM, Mahaffey KW, Moe SM, Wheeler DC, Kubo Y, Dehmel B, Goodman WG, Chertow GM;
The Effects of Cinacalcet in Older and Younger Patients on Hemodialysis: The Evaluation of Cinacalcet HCI Therapy to Lower Cardiovascular Events (EVOLVE) Trial. Clin J Am Soc Nephrol:10:791-9, 2015.

20. Walach H, Sadaghiani C, Dehm C, Bierman D. The therapeutic effect of clinical trials: understanding placebo response rates in clinical trials – a secondary analysis. BMC Med Res Methodol 5:26, 2005.

21. Tripepi G, Heinze G, Jager KJ, Stel VS, Dekker FW, Zoccali C. Lag-censoring analysis: lights and shades.
Nephrol Dial Transplant, 30:700-5,, 2015

22. Kubo Y, Sterling LR, Parfrey PS, Gill K, Mahaffey KW, Gioni I, Trotman M-L, Dehmel B, Chertow GM.
Assessing the treatment effect in a randomized trial with extensive non-adherence: the EVOLVE trial.
Pharm Stat 14:242-51, 2015.

23. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, Bala MM, Bassler D, Mertz D, Diaz-Granados N, Vandvik PO, Malaga G, Srinathan SK, Dahm P, Johnston BC, Alonso-Coello P, Hassouneh B, Walter SD, Heels-Ansdell D, Bhatnagar N, Altman DG, Guyatt GH: Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ 344: e1553, 2012.

24. Floege J, Kubo Y, Floege A, Chertow GM, Parfrey PS. The Effect of Cinacalcet on Calcific Uremic Arteriolopathy Events in Patients Receiving Hemodialysis: The EVOLVE Trial.
Clin J Am Soc Nephol; 10:800-7, 2015.

25. Ketteler M, Biggar PH. Evolving calciphylaxis – what randomized, controlled trials can contribute to the capture of rare disease. Clin J Am Soc Nephrol 10:726-8, 2015.

26. Hsu CY, Ballard S, Batlle D, Bonventre JV, Böttinger EP, Feldman Hl, Klein JB, Coresh J, Eckfeldt JH, Inker LA, Kimmel PL, Kusek JW, Liu KD, Mauer M, MifflinTE, Molitch ME, Nelsestuen GL, Rebholz CM, Rovin BH, Sabbisetti VS, Van Eyk JE, Vasan RS, Waikar SS, Whitehead KM, Nelson RD; CKD Biomarkers Consortium.
Cross-Disciplinary Biomarkers Research: Lessons Learned by the CKD Biomarkers consortium.
Clin J Am Soc Nephrol;10:894-902, 2015.

27. Faul C, Amaral AP, Oskouei B, Hu MC, Sloan A, Isakova T, Gutierrez OM, Aguillon-Prada R, Lincoln J, Hare JM, Mundel P, Morales A, Scialla J, Fischer M, Soliman EZ, Chen J, Go AS, Rosas SE, Nessel L, Townsend RR, Feldman HI, St John Sutton M, Ojo A, Gadegbeku C, Di Marco GS, Reuter S, Kentrup D, Tiemann K, Brand M, Hill JA, Moe OW, Kuro OM, Kusek JW, Keane MG, Wolf M. FGP23 induces left ventricular hypertrophy. J Clin Invest 121:4393-4408, 2011.

28. Isakova T, Zie H, Yang W, Xie D, Anderson AH, Scialla J, Wahil P, Gutierrez OM, Steigerwalt S, He J, Schwartz S, Lo J, Ojo A, Sondheimer J, Hsu CY, Lash J, Leonard M, Kusek JW, Feldman HI, Wolf M. Fibroblast growth factor 23 and risks of mortality and end-stage renal disease in patients with chronic kidney disease. JAMA 304:2432-2439, 2011.

29. Moe SM, Chertow GM, Parfrey PS, Kubo Y, Block GA, Correa-Rotter R, Drüeke TB, Herzog CA, London GM, Mahaffey KW, Wheeler DC, Stolina M, Dehmel B, Goodman WG, Floege J. Cinacalcet, FGF23 and

cardiovascular disease in hemodialysis: the EVOLVE trial. Circulation 132:27-39, 2015.

30. Moe SM, Abdalla S, Chertow GM, Parfrey PS, Block GA,Correa-Rotter R, Floege J, Herzog CA, London GM, Mahaffey KW, Wheeler DC, Dehmel B, Goodman WG, Drüeke TB; for the Evaluation of Cinacalcet HCI Therapy to Lower Cardiovascular Events (EVOLVE) Trial Investigators.

Effects of cinacalcet on fracture events in patients receiving hemodialysis: the EVOLVE trial.
J Am Soc Nephrol. 26:1466-75, 2015

31. Bover J, Urena P, Ruiz-Garcia C, DaSilva I, Lescano P, del Carpio J, Ballarin J, Cozzolino M. Clinical and practical use of Calcimimetics in dialysis patients with secondary hyperparathyroidism. Clin J Am Soc Nephrol. doi: 10.2215/CJN. 01760215, 2015.

32. Belozeroff V, Chertow G M, Graham CN, Dehmel B, Parfrey PS, Briggs AH. Economic evaluation of Cinacalcet in the United States: the EVOLVE trial. Value in Health. 2015 (in press).

33. Suki WN, Zabaneh R, Cangiano JL, Reed J, Fischer D, Garrett L, Ling BN, Chasan-Taber S, Dillon MA, Blair AT, Burke SK: Effects of sevelamer and calcium-based phosphate binders on mortality in hemodialysis patients. Kidney Int 72: 1130-1137, 2007.

Table 1. Lessons from EVOLVE concerning choice and definition of composite endpoints.

1. Primary composite endpoints should include events resulting from pathophysiologic pathways highly likely to be influenced by the intervention.

2. **A**. Heart failure in patients receiving dialysis occurs frequently but is often misdiagnosed by nephrologists even when the definition includes manifestations of pulmonary edema. Sample size should be increased to take account of misclassification.

B. Randomized clinical trials in CKD or ESRD that include an endpoint of heart failure should include focused education to ensure accurate identification of heart failure events.

C. A classification system of heart failure based on patient-reported dyspnea assessed pre- and post-ultrafiltration in conjunction with echocardiography may be useful (6), but requires validation.

3. Selection bias associated with PTX necessitated a more holistic approach to the definition of severe unremitting hyperparathyroidism.

Table 2. Lessons from EVOLVE concerning imbalance in baseline covariates.

1. Randomization is not guaranteed to prevent significant imbalance in baseline clinical characteristics (even for a trial with nearly 4000 subjects).

2. RCTs in CKD/ESRD should employ both multivariable adjusted ITT and unadjusted ITT in the analysis of the primary outcome.

3. If there is risk of imbalance of baseline covariates, multivariable adjusted ITT should be used as the primary analysis on which the trial is judged (18).

4. If participants across a broad age range are enrolled, stratification by age may be advisable, as age is an important determinant of many outcomes. A similar argument could be made for other key covariates (e.g., baseline blood pressure or proteinuria) depending on the outcome(s) of interest.

Table 3. The Treatment Effect on the Primary Composite Cardiovascular Endpoint: Prespecified Sensitivity Analyses (1, 20)

	HR ((%% CI)	p-value
Unadjusted ITT	**0.93 (0.85, 1.02)**	**0.112**
MV adjusted ITT	**0.88 (0.79, 0.97)**	**0.008**
Inverse probability of censoring weights	**0.81 (0.70, 0.92)**	**0.031**
Censor at PTX	**0.90 (0.82,0.99)**	**0.029**
Censor at KT	**0.90 (0.82. 0.99)**	**0.032**
Censor at Commercial Cinacalcet Use	**0.90 (0.82. 0.99)**	**<0.001**
Censor at PTX, Commercial Cinacalcet, or KT	**0.84 (0.76, 0.93)**	**<0.001**
Age < 65 years	**0.99 (0.88, 1.11)**	**Interaction 0.007**
Age ≥ 65 years	**0.74 (0.63, 0.86)**	

HR: hazard ratio; CI: confidence intervals

ITT: intention to treat; MV: multivariable

PTX: parathyroidectomy; KT: Kidney transplant

Table 4. Lessons from EVOLVE concerning the analytic plan for non-adherence and co-interventions.

1. Commercial availability of study drug can make drop-in a serious problem as the use of commercial drug off protocol limits the capacity of the trial to observe a treatment effect. During trial execution major efforts are required to reduce use of commercial drug and an analytic plan needs to take account of drop-in.

2. **A.** Discontinuation of study drug is likely in a long-term trial in patients on dialysis, and generous estimates should be incorporated into power calculations.

B. Longer follow-up time does not necessarily increase study power.

3. An analytic plan to take account of non-adherence must be pre-specified. IPCW maybe the best method to account for non-adherence particularly if determinants of non-adherence are well established.

4. It is necessary to take account of the impact of co-interventions that limit assessment of the trial's main treatment effect. Depending on the nature of the intervention, it may be wise to censor follow-up time after the intervention, or to incorporate the intervention into a composite endpoint (e.g., initiation of dialysis in a trial aiming to slow progression of CKD).

Table 5. Lessons from EVOLVE concerning the credibility of sub-group effects.

A subgroup effect is credible (23) when

(a) it is relatively large and highly statistically significant,

(b) it is one of a small group of prespecified hypotheses tested,

(c) the test of treatment x subgroup interaction is significant,

(d) The sub-group effect is consistent with other reports (33)

(e) The sub-group effect is consistent across related outcomes (19)

(f) the biological rationale is plausible.

LEGENDS

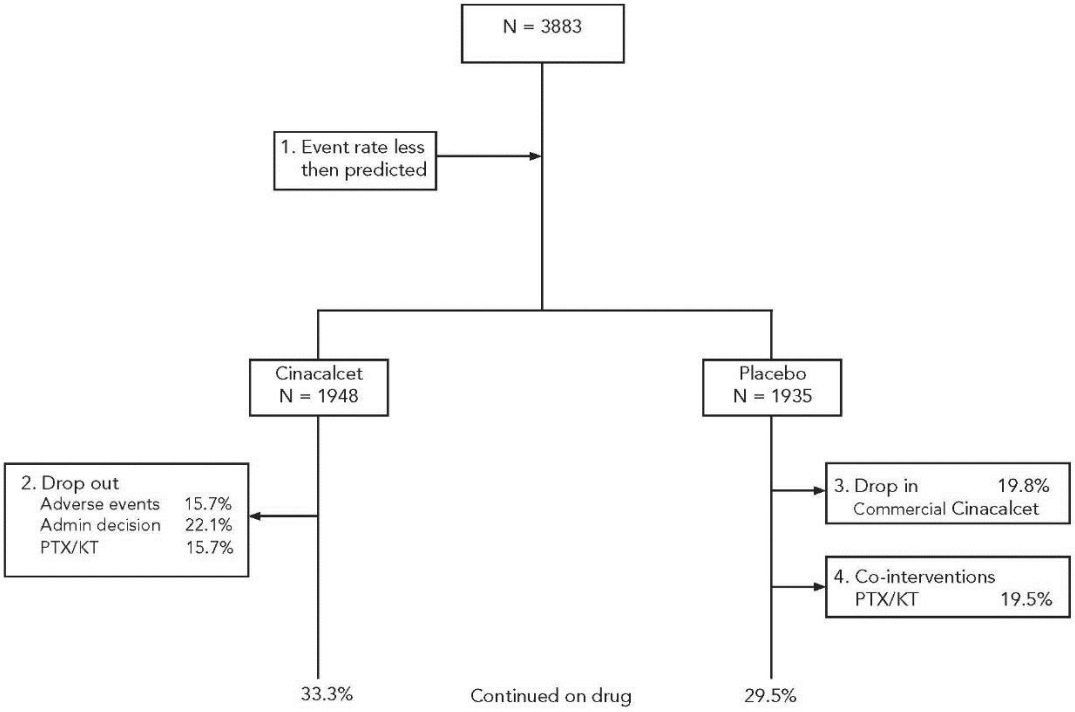
Figure 1. Four main causes of reduction in statistical power that occurred in

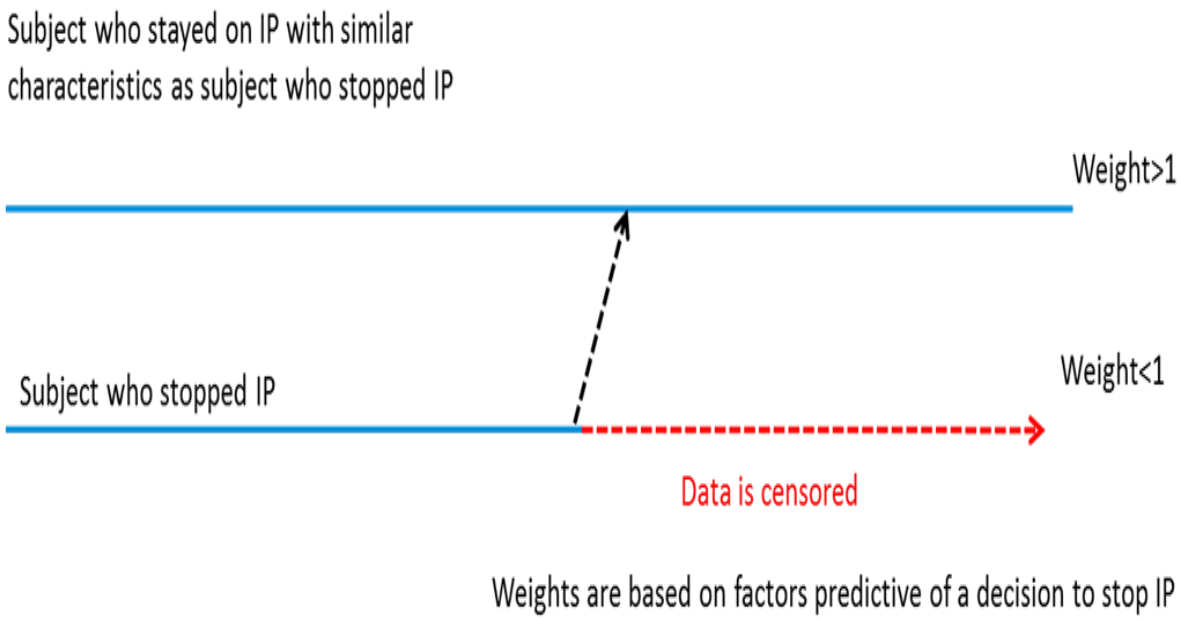
EVOLVE

Figure 2 Inverse Probability of Censoring Weight (IPCW)

- IPCW method creates a scenario of missing follow-up data by censoring the follow-up of each subject at the time of stopping investigational product (IP) (ie, weight=0 for time periods afterwards)

- For subjects with similar characteristics that did not stop IP, IPCW method assigns bigger weights to “to subjects with similar characteristics that dobserved without stopping IP


N = 3883
1. Event rate less then predicted
Cinacalcet N = 1948
Placebo N = 1935
2. Drop out
Adverse events 15.7%
Admin decision 22.1%
PTX/KT 15.7%
3. Drop in 19.8%
Commercial Cinacalcet
4. Co-interventions
PTX/KT 19.5%
33.3%
Continued on drug
29.5%



Weights are based on factors predictive of a decision to stop IP