Reviewer's report

Title: Coronary Artery Disease Evaluation in Rheumatoid Arthritis (CADERA). Quantitative MRI to define mechanisms of cardiovascular co-morbidity in patients with early Rheumatoid Arthritis and to measure the effect of biological therapy: study protocol for a randomized controlled trial.

Version: 2  Date: 16 July 2014

Reviewer: Marissa Lassere

Reviewer's report:

This paper is difficult to follow because (i) it mixes two studies, the 1-year initial versus delayed TNF randomized trial and the matched case control study of subjects with rheumatoid arthritis versus non-rheumatoid subjects, (ii) it does not clearly put each study in clearly in perspective regarding type/level of study, and (iii) it does not clearly discuss the issues underlying their design.

1. The status of the case control study of 30 pairs of patients. Is this a pilot study? The “seminal text of Lancaster” (ref 66) is a discussion of pilot studies and the authors note that “a major reason for such studies is to determine data for the primary outcome measure in order to perform a sample size calculation for a larger trial.” Can the authors clarify what analyses will be used to identify imbalances in covariates other than age and BP and the power of those analyses, and estimate the power of those analyses. There are concerns both that such analyses will be underpowered, and that their use will increase the risk of false positives given the multiple possible risk factors being assessed for balance in the small number, 30, of case-control pairs. There are many CV MRI measures being compared and by chance some will differ significantly – increase in Type 1 error. These other cardiac MRI measures are likely to be influenced by strong CVD risk factors – diabetes, hyperlipidemia and smoking.

2. Please state the justification for choosing aortic distensibility as the primary MRI endpoint? If CADERA is exploratory, multiple endpoints can be dealt with. The issue is more important for VEDERA. Is the VEDERA MRI component also considered exploratory?

3. Are the VEDERA patients stratified by prior/present use of steroids and non-steroidals (traditional and coxibs), given the possibility that these treatments may also alter CV risk?

4. Could the VEDERA design result be confounded if TNF treatment increases CV risk? Will detection of this confounding, if present, be reduced by the crossover design (control patient target failure at 24wk switched to TNF)?

5. Are all 120 VEDERA patients recruited into CADERA and cardiac MRIs available for analysis, or is CADERA just a single time point comparison of CV MRI? Or is CADERA just the first 30 VEDERA patients matched to 30
non-VEDERA controls.

6. Could the authors provide in greater detail where and how are controls recruited given that this aspect of the study is not randomised? Can the reliability of the cardiac MRI measures be provided (see p14) using other measures of reliability rather than the coefficient of variation? I note that later in the text the smallest detectable difference is provided. Therefore, could the intra-observer and inter study reliability for aortic distensibility, circumferential strain, LV twist, and quantitative perfusion analysis be provided in the units of analysis as well as ICCs.

7. Can the mean difference and standard deviation be provided for the primary outcome measure - aortic distensibility - in the relevant units in the text in addition to providing the effect size for both studies? What is the power for Aim 1. What is the power for Aim 2.

2. IS THE WRITING ACCEPTABLE?

On an initial glance it appears acceptable but once one attempts to evaluate design and statistical issues regarding replication it is difficult to do because of the way the protocol details are provided and because it has a case-control and a RCT component. It may be clear to the authors but it is somewhat confusing for readers.

Restructure the paper and separately describe the following features for each study: hypothesis, enrolment criteria, primary outcome (measure and timepoint) and the justification for its choice, sample size calculation including previous patient source, measure used, treatment effect desired, magnitude of difference shown, SD of difference shown, and statistical test used. Please make the descriptions, e.g., of powering, self-contained, not requiring consulting further references for understanding, and with enough detail to enable the reader to at least understand if not reproduce the analysis. As I read the manuscript CADERA seems to consist of the one-time point matched control analysis to see if CV MRI abnormalities are different in RA patients vs non-RA matched controls. VEDERA is a standard RCT comparing early versus delayed TNF treatment in RA using standard clinical measures and using baseline vs 1-year CV MRI.

General comment regarding surrogate endpoints:

The background section loosely uses the term “surrogate” and even describes arterial stiffness as a “strong surrogate marker for CVD risk”. There are two issues here: one is that an accepted surrogate requires multiple RCT data showing a change in the surrogate associated with treatment translates into an improved outcome, and the only such accepted surrogate endpoints to date in CVD are blood pressure and LDL-cholesterol. The article referenced in support of arterial stiffness, ref 10, is a 2 ½ hear cohort study of 141 geriatric patients from 2001 that noted that their result “needed to be confirmed in an intervention trial”. The second issue is that CV MRI is being investigated as a predictor of CV risk status, so it is a risk factor predicting another (set of) risk factors. The term surrogate marker has been replaced by the term surrogate endpoint in a number
of reports from USA National Institutes for Health, Institute of Medicine and the
FDA. Neither arterial stiffness nor Carotid Intimal-Media Thickness is considered
a surrogate endpoint. However, both are, respectively, physiological and
structural biomarkers that are associated with an increased risk of CV events.