

*Question 4): One of your stated areas of expertise is clinical epidemiology. Thus, please consider one of the two articles attached to this exam (S Bent et al, Saw palmetto for benign prostatic hyperplasia. New England Journal of Medicine 2006;354:557-66 OR JA Blumenthal et al., Effects of exercise and stress management on markers of cardiovascular risk in patients with ischemic heart disease: a randomized controlled trial. JAMA 2005;293:1626-34. Do not read commentaries that accompanied or followed the publication of the study.*

*Is the study convincing? Is the methodology appropriate? Are there problems in the study design? Are these problems significant enough to affect the conclusions? Are the statistical methods appropriate? What alternative statistical methods could have been used?*

I will consider the Blumenthal et al. (2005) study on the effects of exercise and stress management on markers of cardiovascular risk in patient-citizens with IHD (ischemic heart disease). This recently published study is convincing to me for several reasons. The improvements in several intermediate cardiovascular risk markers and psychosocial distress/depression measures for the exercise and stress management groups over usual medical care group are convincing because they are sizable and statistically significant in several cases, internally consistent across multiple cardiovascular endpoints, and biologically plausible. A demonstration of “dose-response” would make the benefits of the interventions even more convincing; presumably more rigorous exercise and stress management/reduction regimens would lead to further improvements in cardiovascular function. The fact that the three groups did not have comparable percentages of individuals with annual incomes >\$50,000 or “at least some college” (p. 1630, Table 1) was worrisome given that such measures correlate well with increased psychosocial stress and adverse health outcomes. Greater attention should have been paid in making sure the randomization process made these measures more uniform across treatment groups.

The general methodology of this randomized controlled trial is appropriate, though there are a few small problems which mainly have to do with ambiguity. The study participants were recruited in an appropriate way with sensible inclusion and exclusion

criteria (p. 1627, col. 1). However, while the authors specify what they mean by the inclusion criterion of “documented IHD”, they are ambiguous about what constitutes “evidence of exercise-induced myocardial ischemia within the past year.” Is this evidence from clinical treadmill testing or from self-report of exercised induced angina, or both? Clarification would be helpful for a better understanding of the study participants’ baseline health status. The study uses appropriate methods to generate exercise-induced and mental stress-induced myocardial ischemia. It is good that the authors noted that such methods were found to be effective in previous studies. With regards to subject preparation, the authors write: “Unless medically contraindicated, patients discontinued anti-ischemic medications...at least 48 hours prior to testing” (p. 1627, col. 2). Given that all the patient-citizens being tested have stable IHD, one could make the case that discontinuing medication in all of their cases would be contraindicated. The authors should specify how such decisions of contraindication were made. One of the mental stress induction techniques—judged public speaking on a controversial topic—is innovative and interesting but not all people respond with apprehension in public speaking and not all topics are understood by all to be controversial. Were the topics randomly assigned? Or by some other method? The method of radionuclide ventriculography for determination of the presence of myocardial ischemia is appropriate and was appropriately performed at multiple time points during the testing. Measurement methodology of LVEF (left ventricle ejection fraction) was also appropriate. The method for measurement of segmental wall motion abnormalities (WMA) was ambiguous. “Wall motion for each of the 4 segments was rated by consensus of at least 2 experienced physicians.” Why “at least”? A more appropriate methodology to reduce bias would be

to specify an exact number of experienced WMA readers. The methodology for measuring the flow-mediated dilation index of vascular endothelial function, specifically in the brachial artery, was appropriate. However, glyceryl trinitrate response may be less reliable in those patient-citizens who are taking long-acting nitrates. The method used to measure heart rate variability during deep breathing (HRV-DP) was problematic in that only data from patient-citizens in whom R-R interval lengthening was observed was analyzed. This introduces the possibility of researcher bias in the selection of some patient-citizens with slightly greater or less R-R interval lengthening. A fixed R-R interval change should have been specified or no R-R interval lengthening-based discrimination should have been done in the measurement of HRV-DP. The methods for measuring baroreflex sensitivity, cardiorespiratory fitness, and psychosocial functioning were appropriate and meaningful intermediate clinical endpoints in that poor measurements in all three categories are associated with adverse cardiovascular outcomes. The methodology used in the intervention and “usual care” groups was appropriate but sometimes ambiguous. The method for exercise training was appropriate, although authors should have said something about how such high attendance rates were achieved across the consecutive 16 week period. What kinds of incentives/disincentives were offered? The methodology for stress management training, while appropriate, could have been improved significantly by including games/play activities, film/theatrical/art appreciation, therapeutic recreation, suspension therapy, yoga/massage/meditation/manipulation, and therapeutic landscape exposure. The authors were basically vague about the details of the stress management program; such details would be helpful in understanding their methodology, especially given the success of the

intervention. For example, when the authors write that “social support was considered to be a key aspect of the program” (p. 1629 col. 1), particular examples of how this was achieved would have been very helpful. Finally, the “usual care” group was ambiguously defined. Did any stress management or exercise activities occur in these populations? Even though they monitored to ensure that usual care patient-citizens “had not joined any exercise or stress management training program[s]” (p. 1629 col. 2), this does not exclude the possibility that such activities were occurring in the course of their usual medical care. The authors seem to indicate that only medication management occurred in this group. More clarification is needed.

I do not believe that there were any major problems in the study design. The small number of problems in study design did not significantly effect the study conclusions. Study recruitment methods and sample sizes were appropriate given the desire to achieve a study powered to detect ~15% greater improvements in the treatment groups with a 2-sided test and 5% type I error rate. Block and two-stage randomization was appropriate given the authors’ desires to accommodate availability of patient-citizens for study and to begin active patient-citizen participation in the study no more than 4 weeks following baseline evaluations. The patient-citizens were randomized into three groups—usual care, usual care + exercise, and usual care + stress management training—and received baseline and post-treatment assessments. It would have been very interesting if a fourth group that included both stress management training and exercise would have been created in order to see if the combination of both treatments resulted in synergistically improved outcomes. The study design had a small problem in that all patient-citizens did not undergo HRV-DB assessment because these measurements were

initiated only after the trial had begun. Only 47 patient-citizens were assessed. While appropriate statistical adjustments were made so as not to adversely effect study conclusions, the design could have been improved by including all patient-citizens in this assessment.

The study's statistical methods are appropriate. A general linear model was used in which post-treatment measures served as dependent variables, treatment group as the between-subject factor, and baseline measures/demographics as covariates. Other models were used for specific assessments. The models were checked to make sure they met certain assumptions such as additivity, linearity, and distribution of residuals. I assume they sought to ensure a normal distribution of residuals. This all seemed appropriate. The authors compared exercise and stress management versus usual care and exercise training versus stress management. The main data values on Table 3 are expressed as fitted means with standard errors in parentheses and are adjusted for age, sex, prior MI, pretreatment resting LVEF, and pretreatment level of the corresponding outcome under consideration. Significance is calculated presumably with ANOVA testing. This is all done appropriately. Alternatively, they could have also included an analysis of the stress management group versus usual care and the exercise group versus usual care. They did this occasionally but not consistently. The authors appropriately used an intention-to-treat analysis; for drop-outs, baseline measures were carried forward. The authors also describe a propensity score statistical approach. According to Rosenbaum and Rubin (1983) whom they cite, "The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates." They did not specify how they arrived at their propensity score or how pretreatment values were

treated as covariates. They simply reported that the propensity score analysis was “essentially the same as our primary models” (p. 1629 col. 3). Further clarification would have been helpful.

Overall, the study was well-designed, executed relatively well, and analyzed properly. The results are important for stress-reduction, cardiac, and mind-body medicine.

#### REFERENCES:

- Blumenthal JA et al. 2005. Effects of exercise and stress management on markers of cardiovascular risk in patients with ischemic heart disease: a randomized controlled trial. *JAMA* 293:1626-34
- Rosenbaum PR, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 70:41-55.