1. Table 1. Characteristics of study participants (men with hormonally treated prostate cancer)

	Relapsed	Not relapsed
Number of participants	22	28
Age (years)	68.3 (66 – 71)	66.7 (65 – 69)
Pretreatment PSA (ng/mL)	732 (97 – 1368)*	617 (76 – 1159)*
Nadir PSA (ng/mL)	31.9 (8.7 – 55)	4.12 (-2.6 – 11)
Performance status	76.5 (71 – 82)*	83.9 (80 – 88)
Bone scan score	2.80 (2.6 – 3.0)*	2.32 (2.0 – 2.6)

Values are means (95% confidence interval) except number of participants.
*Data on pretreatment PSA, performance status and bone scan score were missing for two observations in the relapsed group (n=20). **Data on pretreatment PSA were missing for five observations in the not relapsed group (n=23).

2. Table 2. Odds ratio for prostate cancer relapse within 24 months by nadir PSA

Nadir PSA modeled as	Odds Ratio (95% CI)	p-value
Continuous	1.034 (0.99, 1.08)	0.1557
Continuous, log transformed	2.363 (1.42, 3.93)	0.0009*
Linear splines**	--	0.0260*

Odds ratios and statistical significance were determined using logistic regression models adjusted for performance status and bone scan score. P-value for linear spline model was calculated by testing if any of the spline psa parameter estimates were different from 0 using a Wald Chi-square with 4 degrees of freedom.
*Statistically significant ($p < 0.05$).
**Knots at 1, 4 and 16 ng/mL.

a. There is not sufficient evidence to support an association between prostate cancer relapse within 24 months after hormonal treatment and nadir PSA when nadir PSA is modeled as a continuous, untransformed variable (OR = 1.034, $CI_{95\%}$= 0.99, 1.08) (Table 2).

b. When nadir PSA is modeled as a continuous, log-transformed variable, odds of relapse within 24 months is estimated to be 2.4 times higher for each log-unit increase in nadir PSA (OR = 2.363, $CI_{95\%}$= 1.42, 3.93) (Table 2).

c. There is evidence to support an association between prostate cancer relapse within 24 months after hormonal treatment and nadir PSA when nadir PSA is modeled as a linear spline with knots at 1, 4 and 16 ng/mL (χ^2 = 11.05, df = 4, p = 0.0260) (Table 2 & Figure 1).

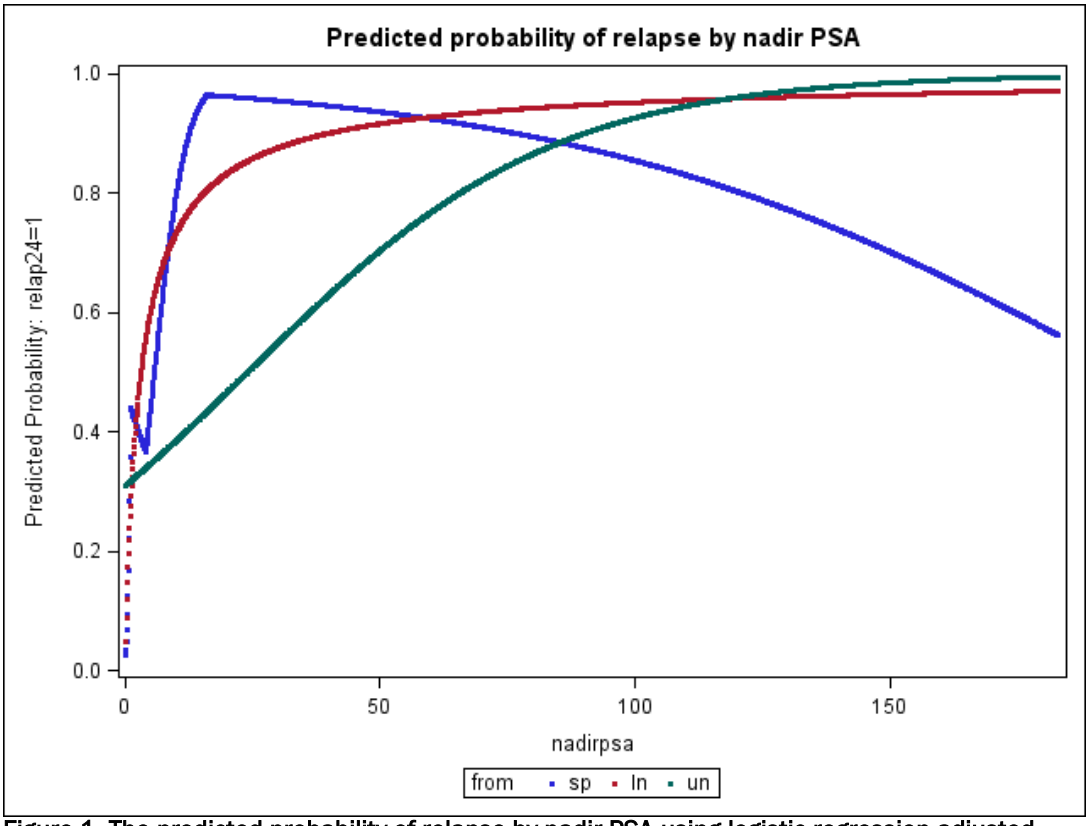

Figure 1. The predicted probability of relapse by nadir PSA using logistic regression adjusted for bone scan score and performance status where nadir PSA is modeled as a continuous, untransformed variable (green), a natural log-transformed variable (red) or linear splines with knots at 1, 4 and 16 ng/mL (blue).

d. For the models using continuous PSA and spline PSA, the estimated odds of relapse within 24 months when nadir PSA, performance status and bone scan score are equal to zero (intercept parameter in logistic regression model) are 2.07 and 0.51, respectively. According to the model using continuous, log-transformed PSA, 3.06 is the estimated odds of relapse within 24 months when nadir PSA is equal to 1 (log nadir PSA is equal to 0) and when performance status and bone scan score are equal to zero. In all cases, this parameter is not highly meaningful because 0 is well outside the range of the data for performance status and 0 is not a possible value for nadir PSA or bone scan score.

3. **Table 3. Difference in mean nadir PSA in men who relapsed compared to those who did not relapse within 24 months adjusted for performance score and bone scan score.**

	Relapse (95% CI)	No relapse (95% CI)	Comparison of means (95% CI)	p-value
Mean nadir PSA	30.58 (10.4, 50.8)	7.061 (-1.73, 15.9)	Difference: 23.51 (2.06, 45.0)	0.0317*
Geometric mean nadir PSA	7.988 (3.53, 18.1)	0.5848 (0.315, 1.09)	Ratio: 13.66 (4.48, 41.6)	<0.0001*

Comparison of means and statistical significance were determined using linear regression with standard errors calculated via the Huber-White sandwich estimator. For comparison of geometric means, log-transformed nadir PSA was used as response in the regression analysis.
* Significantly significant ($p < 0.05$).

a. Mean nadir PSA is 23.5 ng/mL higher in men who relapsed within 24 months compared to those who did not relapse after adjusting for performance score and bone scan score (Difference = 23.51, $CI_{95\%}$= 2.06, 45.0, p=0.0317) (Table 3). This association is statistically significant.

b. Geometric mean nadir PSA is 13.66 times as high in men who relapsed within 24 months as in those who did not relapse after adjusting for performance score and bone scan score (Ratio = 13.66, $CI_{95\%}$= 4.48, 41.6, p<0.0001) (Table 3). This association is statistically significant.

4.

a. Models 2a and 3a yields easily interpretable results and are parsimonious. However, these models may not detect an association in the presence of nonlinearity in the relationship of log odds of relapse and nadir PSA, particularly in the event of a well-balanced U-shaped function. Log-transformation of nadir PSA (Models 2b and 3b) can help improve fit of the models to a nonlinear relationship between relapse and PSA. Because the range of nadir PSA spans more than one order of magnitude (0.1 – 183 ng/mL), log-transformation is likely to be a good way to linearize the relationship. Modeling nadir PSA as linear splines (Model 2c) offers further flexibility in detecting a nonlinear relationship between relapse and PSA.

In thinking about which analysis I would have pre-specified for this study, I first think the scientific question of interest. Since we wish to know "whether the nadir PSA level following hormonal treatment for prostate cancer is prognostic of time in remission", this helps us define the outcome and the predictor. This question presumes that nadir PSA level precedes any relapse. Thus, for these purposes nadir PSA represents the exposure and relapse represents the outcome. For this reason, I would have selected a logistic regression model with relapse as the outcome.

Having decided on predictor/response, the next decision point is then to think about how to model the predictor of interest, nadir PSA. As described above, the spread of nadir PSA suggests that log-transformation may be a good way to account for nonlinearity caused by right skewness in the PSA distribution. Linear splines also offer another acceptable way to capture this non-linearity, but at the expensive of adding parameters to the model and of loss of parsimony. And because our question is not about nonlinearity in the relationship, I would prefer the analysis using log-transformed over the splines.

However, it is worth noting that the study design does provide more data that is relevant to the scientific question which asks about "time in remission." Some sort analysis that includes the variable, obstime, might be preferred to any of the analyses presented here.

b. Based on the documented study design, the temporal sequence is not clear. We don't know whether nadir PSA value occurred before or after relapse. Thus, we don't know which is most likely to be the outcome and which the predictor. In this sense, the data is like cross-sectional data and, therefore, any causal inference based on detected associations should be very cautious.