Supplemental Appendix: Evaluation of Instrumental Variable

1. Association of the Instrument with the Exposure

Instrumental variable analysis depends on 2 critical assumptions. The first assumption is that the instrument is strongly associated with the exposure of interest. In an RCT, treatment assignment should be strongly associated with exposure to the treatment of interest. Compliance with treatment should be measured, even when the main analysis utilizes an intent-to-treat approach, because compliance with treatment assignment is a necessary component of the pathway to influencing the outcome of interest. The strength of an instrument can be directly measured in an instrumental variable analysis following the paradigm of measuring compliance with treatment assignment in a RCT. Summary correlation measures can provide a snapshot of the strength of the instrument. In our analysis the use of PET was strongly associated with calendar year. The correlation coefficient for the overall population (n = 2,977) is 0.42 (P < 0.001). The correlation coefficient among the population of patients who underwent surgery, who are the focus of our study (n = 976), was 0.63 (P < 0.001). Additionally, differences in exposure rates between the levels of the instrument can be measured directly using the Wald estimator, $E[X|Z = 1] - E[X|Z = 0]$.

Here we present differences in use of PET in the early study period compared with the later study period for the overall population and for population subgroups (Supplemental Table 1). Examining patient subgroups allows for the evaluation of potential confounding in the instrument. We also present the Chow F statistic, a measure of the strength of an instrument, with values above 10 considered to be strong instruments. The partial r^2 is presented, which represents the amount of variance explained by the instrument when adjusting for additional patient confounders.

Overall, the instrument—calendar time—was associated with a 34% increase in PET use between the early and later periods. This finding was consistent across all subgroups, with patients with squamous cell carcinoma experiencing the largest difference in PET use across the study instrument. The Chow F statistic was large, with a value of 442.8 for the overall population. A value

of 10 or higher is typically considered a strong instrument. The instrument was also associated with a partial r^2 of nearly 12% in all patient subgroups, indicating that the instrument explains approximately 12% more of the variation in PET use compared with identical models that did not include the instrument. All of these findings suggest that the instrument is strongly associated with use of PET—the exposure of interest.

Supplemental Table 1. Strength of the Instrument Overall and Within Subgroups

	Δ rate of PET associated with levels of the instrument (95% CI)*	*F* statistic (*P*)	Partial r^2
Overall	0.34 (0.31, 0.37)	442.8 (<0.001)	0.116
Age (y)			
<65	0.35 (0.30, 0.37)	204.9 (<0.001)	0.132
65-74	0.38 (0.32, 0.40)	173.1 (<0.001)	0.129
≥75	0.28 (0.21, 0.34)	71.2 (<0.001)	0.076
Marital status			
Not currently married	0.30 (0.26, 0.35)	206.8 (<0.001)	0.098
Currently married	0.39 (0.34, 0.45)	238.9 (<0.001)	0.142
Preexisting comorbidity			
No conditions	0.33 (0.27, 0.39)	125.4 (<0.001)	0.102
1 condition	0.31 (0.25, 0.37)	138.2 (<0.001)	0.119
2–3 conditions	0.37 (0.30, 0.44)	111.2 (<0.001)	0.117
≥4 conditions	0.41 (0.30, 0.51)	59.5 (<0.001)	0.154
Histology			
Adenocarcinoma	0.34 (0.28, 0.41)	101.0 (<0.001)	0.108
Squamous	0.45 (0.39, 0.52)	202.1 (<0.001)	0.173
Non–small cell, not otherwise specified	0.25 (0.20, 0.30)	108.1 (<0.001)	0.073
Bronchioalveolar/ neuroendocrine/other	0.34 (0.20, 0.47)	24.8 (<0.001)	0.104

*Wald estimator.

2. **Independence of the Instrument on Outcomes**

The second critical factor of instrumental variable analysis is that the instrument should only influence outcomes through its effect on the exposure (here the use of PET) and should not be independently associated with the outcome through alternative pathways. Our time-based

instrument assumes that any confounding temporal changes, such as changes in characteristics of patients who are being treated over time (e.g., general health of patients over time), or improvements in chemotherapy agents or other treatment changes, are minimal. This assumption is challenging to test directly among the study population. The face validity of this assumption is explored by examining patient characteristics over time, which are presented in Table 1 in the article. This analysis did identify some differences in patient characteristics over time such as the mix of older and younger patients. However, there was no distinct pattern suggesting that patient characteristics are improving or declining over time.

An additional approach to evaluating this assumption is to examine temporal trends in a related patient population that was not eligible for the exposure being studied but may experience outcomes similar to the outcome of interest. Although our main outcome—unnecessary surgery—is relevant only to those subjects who underwent surgery, we are able to examine the association of calendar time with lung cancer survival generally. Here, we selected patients with clinical stage IV disease diagnosed between 1998 and 2008 who were treated with alternative therapies other than surgery (n = 948). We also explored the association between calendar time and death within 12 mo, testing the hypothesis that improvements in this lung cancer outcome are not associated with calendar time. Among this sample, 85.3% of patients died within 12 mo of diagnosis, with no correlation between probability of death and calendar time (P = 0.158).

The results of this evaluation are presented in Supplemental Table 2. These findings suggest that calendar time was not associated with improvements in progression or survival for patients initially diagnosed with stage IV disease. The instrument was not associated with any difference in this outcome for the overall population or for any patient subgroup. Although stage IV patients are quite different from patients who underwent resection, if there were temporal changes in the characteristics of patients diagnosed with lung cancer, improvements in chemotherapy, or other advances in lung cancer care that could potentially confound our main outcome of interest—unnecessary surgery—it is likely any confound temporal improvement would also influence this

alternative population of patients. Thus, this analysis supports the instrumental variable analysis suggesting that the effect of the instrument is not independently associated with improved outcomes but that the effect is through its influence on increased use of PET.

Supplemental Table 2. Association of the Instrument with the Related Lung Cancer Outcome of Rapid Recurrence/Death Within 12 Months Among Patients with Clinical Stage IV Disease at Diagnosis (*n* = 948)

	Δ outcome associated with levels of the instrument (95% CI)*	*F* statistic (*P*)	Partial r^2
Overall	0.03 (–0.02, 0.07)	1.28 (0.258)	<0.001
Age (y)			
<65	0.01 (–0.08, 0.07)	0.01 (0.924)	<0.001
65-74	0.05 (–0.03, 0.13)	0.04 (0.524)	0.001
≥75	0.06 (–0.03, 0.14)	3.15 (0.077)	0.014
Marital status			
Not currently married	0.01 (–0.05, 0.07)	0.37 (0.543)	<0.001
Currently married	0.05 (–0.02, 0.13)	1.09 (0.297)	0.003
Histology			
Adenocarcinoma	0.07 (–0.02, 0.15)	2.87 (0.092)	0.011
Squamous	0.01 (–0.10, 0.12)	0.05 (0.819)	<0.001
Non–small cell, not otherwise specified	0 (–0.07, 0.07)	0.02 (0.882)	<0.001
Bronchioalveolar/ neuroendocrine/other	0.11 (–0.14, 0.35)	0.31 (0.582)	0.005

*Wald estimator.