APPENDIX

Top 25 Diagnoses among Decedents

Twenty five most prevalent principal diagnoses among decedents over the age of 65, Pennsylvania Cost Containment Council Data, 2001-2005, by Clinical Classification System category

Code	Diagnosis
2	Septicemia (except in labor)
100	Acute myocardial infarction
122	Pneumonia (except that caused by tuberculosis or sexually transmitted disease)
108	Congestive heart failure; nonhypertensive
109	Acute cerebrovascular disease
131	Respiratory failure; insufficiency; arrest (adult)
129	Aspiration pneumonitis; food/vomitus
157	Acute and unspecified renal failure
42	Secondary malignancies
127	Chronic obstructive pulmonary disease and bronchiectasis
19	Cancer of bronchus; lung
55	Fluid and electrolyte disorders
153	Gastrointestinal hemorrhage
226	Fracture of neck of femur (hip)
145	Intestinal obstruction without hernia
114	Peripheral and visceral atherosclerosis
115	Aortic; peripheral; and visceral artery aneurysms
106	Cardiac dysrhythmias
233	Intracranial injury
237	Complication of device; implant or graft
159	Urinary tract infections
101	Coronary atherosclerosis and other heart disease
238	Complications of surgical procedures or medical care
103	Pulmonary heart disease
107	Cardiac arrest and ventricular fibrillation

STATISTICAL METHODS

In an observational study or nonrandomized trial, the method of inverse probability treatment weight (IPTW) is usually used to remove sampling bias. The basic concept of the IPTW method is to assign a weight to each observation as the inverse of the probability the observation belongs to a certain treatment, conditional on a set of covariates. An observation with a higher probability of being in a certain treatment is over-representative for that treatment, and thus a lower weight is assigned. Conversely, an observation with a lower probability of that treatment receives a higher weight. After weighting, the treatment effect is no longer related to the confounders. Note that robust standard errors must be used to derive confidence intervals of the estimator of interest and the P values in the inference test.(21)

Most often, the treatment is binary (treatment vs. control), or otherwise categorical, and the level of observation is one individual. A logistic regression model is used to estimate the probability of being in the treatment group, conditional on a set of covariates. For a patient in the treatment group, this probability is interpreted as the probability of being in the treatment group, given the characteristics of that patient. For a patient in the control group, one minus the estimated probability gives the probability of being in the control group, conditional on their specific characteristics. The inverse of this estimated probability (or one minus the probability) is then the IPTW. This is reminiscent of a typical propensity score modeling.

We extend this in two ways. First, our "treatment" is the continuous measure of hospital intensity: it is not categorical. It is natural to think of characteristics being "balanced" across two treatment groups, as in a randomized controlled trial. The typical binary IPTW case replicates this balance in the weighted sample. In the context of continuous treatment, the analogous "balance" is achieved in the weighted sample when the continuous treatment has a similar distribution given any characteristics.(18)

In the typical analysis each patient receives one treatment assignment only for the study (for example, a drug therapy or surgery), and thus one IPTW for their entire observational period. However, in our analysis patients may be admitted to several hospitals over the observational period, and thus are exposed to varying treatment. Hence, the second extension is that patients have varying weights over time. These time-varying weights extend the "balance" over the observational study, to ensure that the continuous treatment has a similar distribution, given any characteristics, throughout the study. As patients change treatment (go to variously more or less intense hospitals), their assigned weights change to reflect over- or under-representation at that time. At each time the patient's history is accounted for, both treatment intensity history and past prognostic factors such as illness and mortality risk, as this history may affect current treatment.(17,22)

The steps to accomplish this weighting proceed similarly to implementations of marginal structural models, simply altered to allow for a

continuous treatment. This is accomplished using linear models to obtain a predicted cumulative distribution function for the continuous hospital intensity measure, the "treatment". This allows us to compute the necessary treatment probabilities. This is analogous to the predicted probabilities obtained from logistic regression models in the binary case. Further, we compute stabilized weights, which are superior as they have smaller variance and yield narrower confidence intervals.(17,22) These weights are the ratio of two probabilities, as opposed to the inverse of a single estimate. The denominator in either case is the same. We obtain the denominator probability from a linear model that controls for patient characteristics, past treatment and history, and current prognostic information. However, in the stabilized weight the numerator is not one, but rather a probability obtained from a similar model which considers past treatment history only, not current nor past prognostic information.

A weighted logistic regression is then used to predict 30-day and 180-day mortality, weighted using the stabilized IPTW and incorporated robust standard errors for intrasubject correlation. A fixed-time survival model, as opposed to a time-to-event analysis as in other implementations of marginal structural models, has advantages and disadvantages. First and foremost amongst the advantages is that comparing the two models allows us to easily and interpretably present distinctions between short-term outcomes and longer-term outcomes, and how hospital treatment intensity may be more or less favorable to one or the other. This is an important consideration, and different

clinical and societal pressures may lean toward one or the other. Further, fixed-time models more consistently appear in other mortality models in the literature, allowing for easier comparison of our intensity measure to others for the same outcome.

There are two primary disadvantages to the fixed-time model. First, patients may just "miss" the time cutoff in either direction. A patient that dies in exactly 30 days is marked as deceased, but a patient that dies in 31 days is not. For any reasonable clinical or societal concern, these outcomes would not be considered different. This shortcoming of the fixed-time survival is mitigated by the richness of our data. With over one million patients, we do not expect a large influence on the results from those patients near the border.

A second drawback to our model is that a patient may have several hospitalizations within the last 30 (or 180) days of life. In our logistic regression model, these will all be marked as deceased. However, each person can clearly only die once. This has two consequences. First, if higher intensity hospitals tended to readmit patients much more or much less often than those of lower intensity, this would introduce bias. In practice this is not an issue: neither the number of admissions per person nor the time between admissions is related to intensity. Secondly, the admissions before death may be to different hospitals, such that a single patient experiences several treatment intensity levels in their last 30 (or 180) days of life. However, because the stabilized IPTW explicitly

include patient history, this effect is capture and the person's mortality is not unduly assigned to one or another hospital.

In order to address these concerns, we additionally ran time-to-event analysis using the same weighting methodology. We used a pooled logistic regression model which approximates the Cox model for the ease of implementation in the statistical package.(20) However, in order to estimate the proper logistic regression model, we must observe all patients at the same fixed time interval (e.g., each month). Clearly, hospital discharge data does not fit this pattern. Therefore, we must force each patient be to observed regularly until death. We choose a time interval and then divide the history of all patients into sections of that duration. If the patient was not observed (had no admissions) in a particular interval the information from the previous interval is copied. For example, if we forced chose one month as our unit of observation for all patients, and a particular patient was first admitted January 1, 2003 and died December 31, 2003, that patient would have twelve observations. If that patient was not admitted during the month of July, the information from June would be used. The choice of the fixed time interval is forced by the data itself. If we chose, say, one month as above, we would observe some patients multiple times in one month. It would not be possible to determine which observation (admission) should be used.

The narrowest interval during which no patient had multiple admissions is one day. Thus, we forced each patient to be observed each day from their first

admission to their death, or the end of the study data collection. The daily characteristics are "updated" at each new admission, and each person is marked as deceased only once, on their final day. This makes for a very large data set. The patient in the example above would have 365 observations. With over one million patients, it was not feasible to perform estimation. To overcome this, we used 500 bootstrap samples, each 1% of the original data. We were able to estimate the daily pooled logistic model on this sample, and combine the results. Even using this bootstrapping method, implementation of this model is extremely computationally intensive, and may not be practical for analysis of hospital claims data.

ADDITIONAL REFERENCES (not in manuscript)

21. **Azoulay P, Ding W, Stuart T.** The effect of academic patenting on the rate, quality, and direction of (public) research output. 2006. Report No.: w11917.

22. **Hernan MA, Brumback BA, Robins JM**. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med*. 2002;21(12):1689-709.

Sensitivity Analyses

Just as indicated in Figure A1, the adjusted odds ratio for death at 30 days (A, B) and 180 days (C, D) postadmission, given treatment in a hospital with a particular end-of-life treatment intensity compared to if the patient had been admitted instead to a hospital with the average end-of-life treatment intensity. Covariables included patient demographics and clinical characteristics, hospitalization history, and hospital characteristics. Panels A and C depict a patient with an average predicted probability of death (PPD) upon admission (black line with 95% confidence interval in shaded grey), panels B and D depict patients at lowest (5th percentile, blue line with 95% confidence interval in shaded blue) and highest (95th percentile, red line with 95% confidence interval in shaded red) PPD upon admission (confidence interval overlap in shaded purple) (note: as the length of follow-up increases, the odds ratio cannot be interpreted as the risk ratio because the event rate exceeds 5%).

FIGURE LEGENDS

Figure 2 Sensitivity Analyses. As in Figure 2, figures represent the adjusted odds ratio for death at 30 days (A, B) and 180 days (C, D) post-admission, given treatment in a hospital with a particular end-of-life treatment intensity compared to if the patient had been admitted instead to a hospital with the average end-of-life treatment intensity intensity. Covariables included patient demographics and clinical characteristics, hospitalization history, and hospital characteristics. Panels A and C depict a patient with an average predicted probability of death (PPD) upon admission (4.6% PPD; black line with 95% confidence interval in shaded grey), panels B and D depict patients at lowest (0% PPD; blue line with 95% confidence interval in shaded blue) and highest (41% PPD; red line with 95% confidence interval in shaded red) PPD upon admission (confidence interval overlap in shaded purple). (Note: as the length of follow-up increases, the odds ratio cannot be interpreted as the risk ratio because the event rate exceeds 5%).

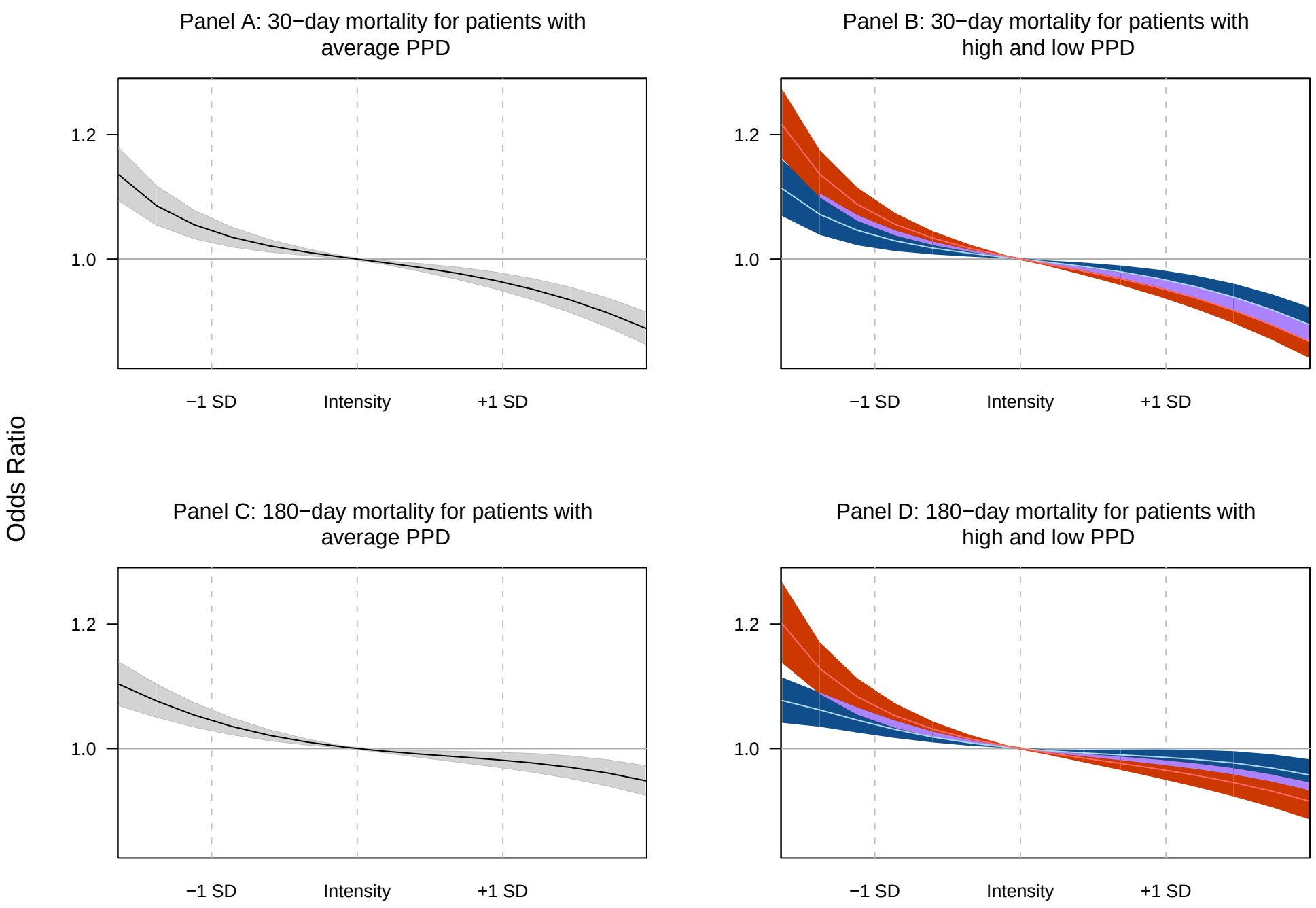
Sensitivity Analysis -- Population (3 columns, each with panels A-D). Adjusted odds ratio 30-day and 180-day post-admission mortality, by hospital treatment intensity among the subset of admissions with a high probability of dying in 3 categories (column 1 – urban [Pittsburgh & Philadelphia] admissions only; column 2 – CHF admissions; column 3 – AMI admissions).

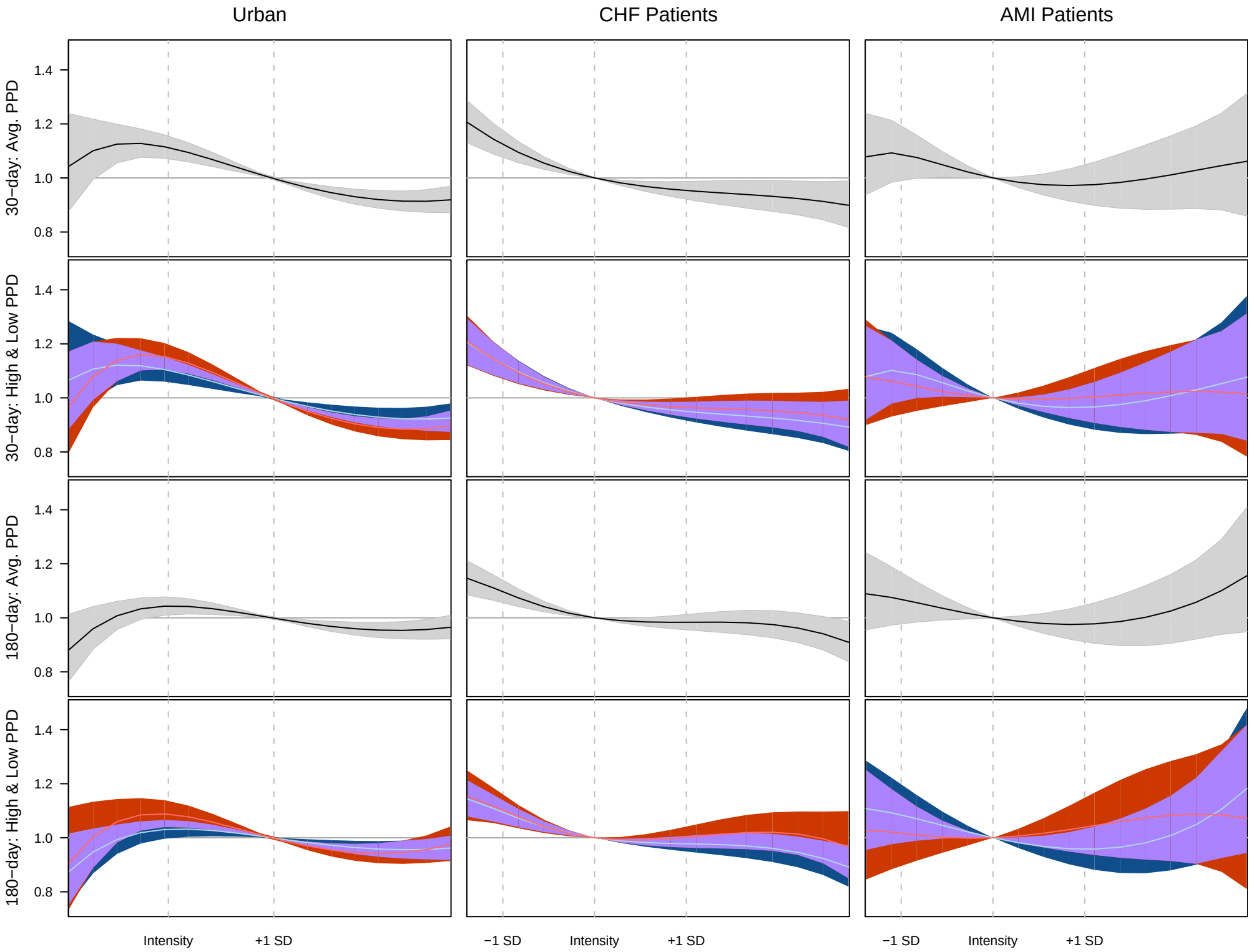
Column 1, Panels A-D: the average treatment intensity of hospitals in Pittsburgh and Philadelphia are approximately 1 SD above the state average. The nonlinear relationship between intensity and survival, which wanes with length of follow up, is retained in this more homogenous urban sample.

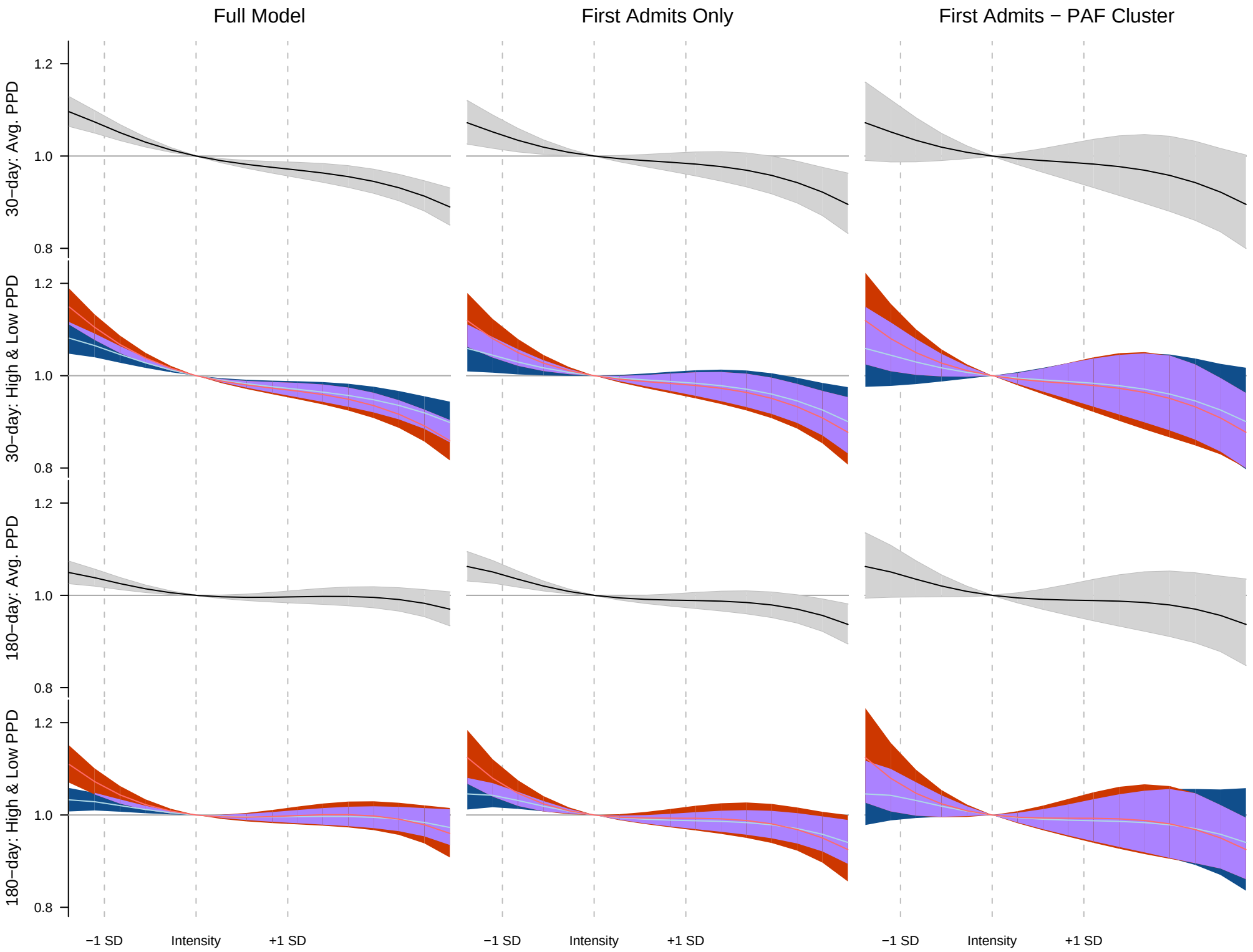
Column 2 and Column 3, Panels A-D: the relationship between treatment intensity and survival is preserved when we restricted analyses to a single chronic condition (CHF), but not when we restricted analyses to a single acute condition (AMI). We hypothesize that post-admission survival after AMI may be less subject to the hospital's tendency to use life-sustaining treatments than to its use of evidence-based treatments. Patients with life-limiting chronic illnesses like CHF, on the other hand, may be rescued – at least in the short-term – by overall greater use of life-prolonging treatments.

Sensitivity Analysis – Dead Measure, Panels A-D. These models are identical to those reported in Figure 2, with the exception that the end-of-life treatment intensity index is based upon treatment patterns among *decedents* rather than among patients "prospectively" identified as being at high risk of dying. The relationship between intensity and survival is more linear, such that returns to intensity increase above "average" treatment intensity, although the relationship still wanes with length of follow-up.

Figure 2, Panels A–D: Dead Cohort




Urban
CHF Patients
AMI Patients
30−day: Avg. PPD
30−day: High & Low PPD
180−day: Avg. PPD
180−day: High & Low PPD
1.4
1.2
1.0
0.8
Intensity
+1 SD
−1 SD
Intensity
+1 SD
−1 SD
Intensity
+1 SD


Full Model
First Admits Only
First Admits − PAF Cluster
30−day: Avg. PPD
30−day: High & Low PPD
180−day: Avg. PPD
180−day: High & Low PPD
1.2
1.0
0.8
−1 SD
Intensity
+1 SD