

Liver Resection Improves Survival in Colorectal Cancer Patients

Causal-effects From Population-level Instrumental Variable Analysis

Mustafa Raouf, MD, MS,* Sidra Haye, MS,† Philip H. G. Ituarte, PhD,* and Yuman Fong, MD*✉

Objective: The aim of this study was to estimate population-level causal effects of liver resection on survival of patients with colorectal cancer liver metastases (CRC-LM).

Background: A randomized trial to prove that liver resection improves survival in patients with CRC-LM is neither feasible nor ethical. Here, we test this assertion using instrumental variable (IV) analysis that allows for causal-inference by controlling for observed and unobserved confounding effects.

Methods: We abstracted data on patients with synchronous CRC-LM using the California Cancer Registry from 2000 to 2012 and linked the records to the Office of Statewide Health Planning and Development Inpatient Database. We used 2 instruments: resection rates in a patient's neighborhood (within 50-mile radius)—NALR rate; and Medical Service Study Area resection rates—MALR rate. IV analysis was performed using the 2SLS method.

Results: A total of 24,828 patients were diagnosed with stage-IV colorectal cancer of which 16,382 (70%) had synchronous CRC-LM. Liver resection was performed in 1635 (9.8%) patients. NALR rates ranged from 8% (lowest-quintile) to 11% (highest-quintile), whereas MALR rates ranged from 3% (lowest quintile) to 19% (highest quintile). There was a strong association between instruments and probability of liver resection (F-statistic at median cut-off: NALR 24.8; MALR 266.8; $P < 0.001$). IV analysis using both instruments revealed a 23.6 month gain in survival (robust SE 4.4, $P < 0.001$) with liver resection for patients whose treatment choices were influenced by the rates of resection in their geographic area (marginal patients), after accounting for measured and unmeasured confounders.

Conclusion: Less than 10% of patients with CRC-LM had liver resection. Significant geographic variation in resection rates is attributable to community biases. Liver resection leads to extensive survival benefit, accounting for measured and unmeasured confounders.

Keywords: causal-effects, instrumental variable, liver metastases, liver resection, survival

(*Ann Surg* 2019;270:692–700)

Every year, colorectal cancer comprises 8.1% of new cancers and 8.3% of all cancer deaths in the United States.¹ Approximately 21% patients have distant disease on diagnosis—most (83%)

involving the liver.² The majority (71%) of patients with distant disease die from liver metastases.³

Liver resection has become the main therapy for patients with colorectal cancer liver metastases (CRC-LM). Several landmark observational studies provide evidence of long-term survival after liver resection in highly selected patients with CRC-LM.^{4–7} For instance, complete resection in patients with liver-limited disease has been associated with a median survival of >40 months, a 5-year survival of 30% to 55%.^{7–17} In approximately 20% to 25% of patients who survive 10 years, this defines cure.^{4,8,15,18} Alternatively, all patients with distant disease treated with modern systemic chemotherapy alone have a median survival of 18 to 24 months with rare survivors at 10 years from diagnosis (reviewed here¹⁹). These observations provide a strong rationale for liver resection as life-prolonging and potentially curative therapy.

Even though there is a strong association between liver resection and long-term survival, the possibility that some or all of the improvement in survival after liver resection is attributable to patient selection cannot be ignored.²⁰ Recurrence after liver resection is common—two-thirds recur within 5 years. Even among those who survive 5 years, one-third will still die of their disease.⁴ Therefore, for the vast majority, liver resection is potentially life-prolonging rather than curative. Despite lack of causal evidence from randomized controlled trials, liver resection has now become routine practice in tertiary centers around the globe.^{21–27} At present, a randomized trial to empirically test the benefit of liver resection in CRLM is neither ethical nor feasible.

The objective of the study was to estimate population-level causal-effects of liver resection on overall survival of patients with CRC-LM. It is difficult to estimate the causal effect of liver resection on overall survival using traditional regression methods employed in observation studies.²⁰ These analyses adjust for known confounders but cannot address the unmeasured biases in treatment choice. Estimating the causal effect of liver resection on survival, therefore, requires an exogenous source of variation that affects liver resection and is not correlated with other factors that affect survival. Liver resection rates across geographical areas provide plausibly exogenous variation in the probability that a given patient gets liver resection. For instance, it is likely that the decision to perform liver resection will depend on the prevailing practice in the patient's area of residence. The area resection rates are exogenous because patients typically do not choose residence based on a need for future liver resection. Accordingly, we use instrumental variable (IV) analysis²⁸—a quasi-experimental technique—to exploit this geographic variation in liver resection rates as source of variation in the probability that a patient gets liver resection.^{29–31}

We hypothesize that increasing geographic area liver resection rates will improve long-term overall survival of patients with CRC-LM. The causal inference from this study directly applies to patients in whom treatment choices are influenced by geographic area resection rates (eg, by prevailing physician beliefs, practice patterns, and surgical capability in the geographic area).

From the *Department of Surgery, City of Hope National Medical Center, Duarte, CA; and †Department of Economics, University of California Irvine, Irvine, CA.

✉yfong@coh.org.

Authors' contributions: Conception and design: M.R., Y.F.; Provision of study materials or patients: M.R., P.H.G.I., Y.F.; Collection and assembly of data: M.R., P.H.G.I.; Data analysis and interpretation: all authors; Manuscript writing: M.R., S.H.; Final approval of manuscript: all authors; Accountable for all aspects of the work: all authors.

The authors report no conflict of interests.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/19/27004-0692

DOI: 10.1097/SLA.0000000000003485

METHODS

Databases

The California Cancer Registry (CCR) is one of the most complete registries in the country.³² In California reporting of cancer care is mandatory, yielding a low rate of records missing or lost to follow-up. Patient discharge data (PDD) after inpatient hospitalization were acquired from the California Office of Statewide Health, Planning, and Development (OSHPD). The PDD files contain patient-level data for all general, acute-care, nonfederal hospitals in California. For each admission, the PDD files include principal diagnosis and as many as 24 secondary diagnoses coded using the *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-CM-9)* format, the principal procedure, and as many as 20 secondary procedures. This information enables a more accurate assessment of patient comorbidities and more detailed information on surgical procedures than is currently available from any cancer registry data alone. The study was approved by the institutional review boards of the State of California and City of Hope National Medical Center, Duarte, CA, with a waiver of informed consent.

Database Linkage

Cases identified in the CCR from January 1, 2000, through December 31, 2012, were linked to PDD files from OSHPD by applying a probabilistic linking algorithm based on sex, date of birth, and social security number as described previously.³³ Follow-up data was available through December 31, 2015.

Study Cohort

Patient Eligibility and Exclusion Criteria

We included patients from CCR with histologically confirmed colorectal cancers based on *ICD-O-3* codes for site (C180, C182-C189, C199, C209) and histology codes for adenocarcinoma (814, 821, 822, 848, 849, and 857). Data on patients with synchronous liver metastases (CRC-LM within 6 months of diagnosis) at diagnosis were obtained from PDD are cross-referenced with CCR to confirm AJCC Stage IV disease. Although PDD files can be used to identify patients with metachronous CRC-LM, this could not be cross-referenced with the CCR because CCR only records stage at diagnosis. Therefore, to ensure robustness of our findings, we limited our analysis to patients with synchronous CRC-LM. We excluded those that were: non-analytic cases; without histologically confirmed diagnoses; with other primary malignancies; <18 years; diagnosed at autopsy; diagnosed in hospice; or without follow-up information. Although presence of extrahepatic metastases was once considered an absolute contraindication to liver resection, recent literature challenges this assertion.³⁴ Therefore, patients with extrahepatic liver metastases were not excluded from analysis. The specific exclusion steps are presented in the Supplementary Information (Figure S1, <http://links.lww.com/SLA/B707>).

Patient's Data Variables and Definitions

Variables included age, race/ethnicity, sex, marital status, Charlson–Deyo score³⁵ for co-morbid conditions, tumor laterality, nodal status, T-stage (in accordance with AJCC 6th edition), extrahepatic metastases, resection of primary, and receipt of chemotherapy. As all liver resections require inpatient hospitalization, performance of liver resection was confirmed with *ICD-9-CM* procedure codes for liver resection in PDD files.

Outcome

Our main outcome of interest was overall survival. Follow-up was measured from the time of diagnosis to date of death or last

contact. Survival was defined as proportion of patients alive at 1 year, 2 years, 3 years, 4 years, or 5 years after diagnosis. We also modeled survival as a continuous variable. Entire life-span data (diagnosis to death) were available for 91% of the patients.

Instrument Variables

We used 2 geographic area-based instrument variables: neighborhood (50-mile radius) area liver resection (NALR) rate³⁶; or Medical Service Study Area liver resection (MALR) rate. NALR rate was calculated for each patient and was defined as the proportion of other patients (excluding the index patient) with CRC-LM undergoing liver resection within a 50-mile radius of the index patient's residence. We used a 50-mile threshold because a majority of patients undergo liver resection within 50 miles of their residence (Figure S2, <http://links.lww.com/SLA/B707>). Similarly, MALR rate was calculated as proportion of patients with CRC-LM undergoing liver resection within a geographically defined boundary of the Medical Service Study Area (MSSA). MSSAs are sub-city and sub-county geographic units as defined by the US Census Bureau 2010 Census.³⁷ MALR and NALR rates were calculated during the entire 13-year period of the study. During the course of this study there is a possibility that a geographic area might have changed from a low to high liver resection rate region or vice versa. For approximately two-thirds of the patients the assignment does not change over time (Figure S3, <http://links.lww.com/SLA/B707>). Conversely for approximately one-third of the patients the assignment changes over time. Despite this variation, the 13-year aggregate rate was chosen as the instrument because the resection rate estimates get more precise with higher number of patients. We believe that this trade-off was an acceptable compromise. As we show later in the results, the instrument captured treatment philosophy that the patient encountered when the treatment decision was made.

Statistical Analysis

Data Reporting

Continuous variables are presented as median (interquartile range) and categorical variables as frequencies with percentages. Distribution of characteristics and potential bias was estimated using absolute standardized differences (ASDs). ASDs are increasingly used to describe or compare groups in clinical trials and observational studies, in preference over *P* values. ASD of ≥ 0.10 indicates that covariates are imbalanced between groups.³⁸ Magnitude of ASD corresponds to the degree of imbalance.

Empirical Model

IV analysis is shown in Figure 1. We used a nonparametric 2-stage least squares (2SLS) variant of IV estimation as described before.³⁶ We chose not to use the 2-stage residual inclusion (2SRI) method commonly used to generate hazard ratio estimates in a Cox proportional hazard model because of concerns for bias.³⁹ As opposed to other methods that rely on error term distributional assumptions, 2SLS yields consistent estimates regardless of the underlying error distributions.⁴⁰ Our main equation of interest is:

$$survival_i = \alpha_0 + \alpha_1 Liver_resection_i + \alpha_2 X_i + u_i \quad (1)$$

However, as discussed above, the estimates from this equation would be potentially biased because of omitted variable bias.²⁰ Therefore, we estimate the effect of liver resection on survival using 2SLS method. In the “first stage” of the 2SLS approach we used

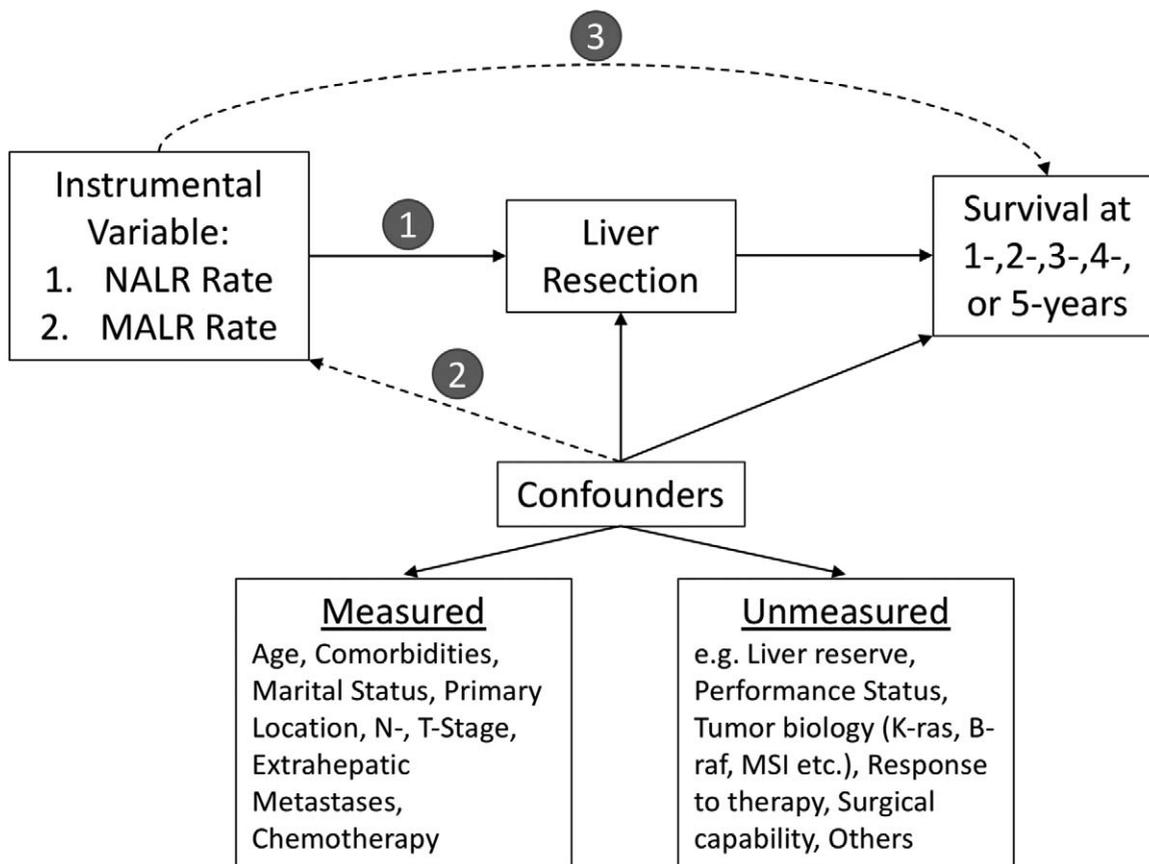


FIGURE 1. Conceptual framework of instrumental variable analysis. An instrumental variable is valid if it is associated with outcome (survival) only through exposure (liver resection). If instrumental variable is associated with exposure (liver resection), first assumption of IV analysis is valid; if instrumental variable is not independent of confounders, second assumption of IV analysis is violated; if instrumental variable has a direct effect on survival, this assumption of IV analysis is violated.

ordinary least squares (OLS) to estimate the likelihood of liver resection:

$$Liver_Resection_i = \gamma_0 + \gamma_1 Area_Rates_i + \gamma_2 X_i + v_i \quad (2)$$

Where Liver_resection is the surgical choice for patient i (1=liver resection, 0 = no liver resection), X_i includes confounding variables (age, comorbidities, year of diagnosis, socioeconomic status, marital status, tumor laterality, nodal status, T-stage, extrahepatic metastases, and chemotherapy). $Area_Rates_i$ is a binary variable that defines patient group based on patient’s instrument value. The distribution of each instrument did not reveal an obvious cut-off; therefore, we analyzed results using multiple cut-offs at 25th percentile, median, and 75th percentile. The strength of the instrument was estimated using the F -statistic.⁴¹ This tests the assumption that the IV ($Area_Rates_i$) describes a significant portion of the variation in liver_resection _{i} , that is, whether the instruments affect choice of liver resection. In general, an instrument with an F -statistic >10 is considered to be a strong instrument.

Next, we estimated survival using 5 different survival measures (1 year, 2 years, 3 years, 4 years, and 5 years). Each survival model was specified as follows:

$$S_i = \beta_0 + \beta_1 Liver_resection_i + \beta_2 X_i + \varepsilon_i \quad (3)$$

where S_i is a binary variable equal to 1 if patient “ i ” survives beyond a certain time interval past their diagnosis (1 year, 2 years, 3 years, 4 years, or 5 years), and ε_i is the set of unmeasured factors that affect patient survival and not surgery choice. β_1 measures the local average treatment effect (LATE) of liver resection on survival for a patient that gets the liver resection because of high liver resection rates in the area. Alternatively, we modeled survival as a continuous variable limiting the analysis to 91% of the patients for whom actual survival (time from diagnosis to death) was available. We accounted for temporal trends and socioeconomic disparity using year of diagnosis and county-level Yost-Yang socioeconomic index,⁴² respectively. Robust standard errors are reported.

Software: For all statistical analyses we used STATA/MP software (version 14.1; StataCorp LLP, College Station, TX) with assumption of 2-sided tests and a criterion for statistical significance set at $\alpha < 0.05$ unless otherwise indicated. Patients within 50-mile radius of index patient were identified based on patient residence coordinates using Quantum Geographic Information System (version 3.6).

RESULTS

Differences in Measured Confounders by Instrument Variable Groupings

Two geography-based instrument variables were used for this study: NALR rates and MALR rates. Figure 2 shows that NALR rates

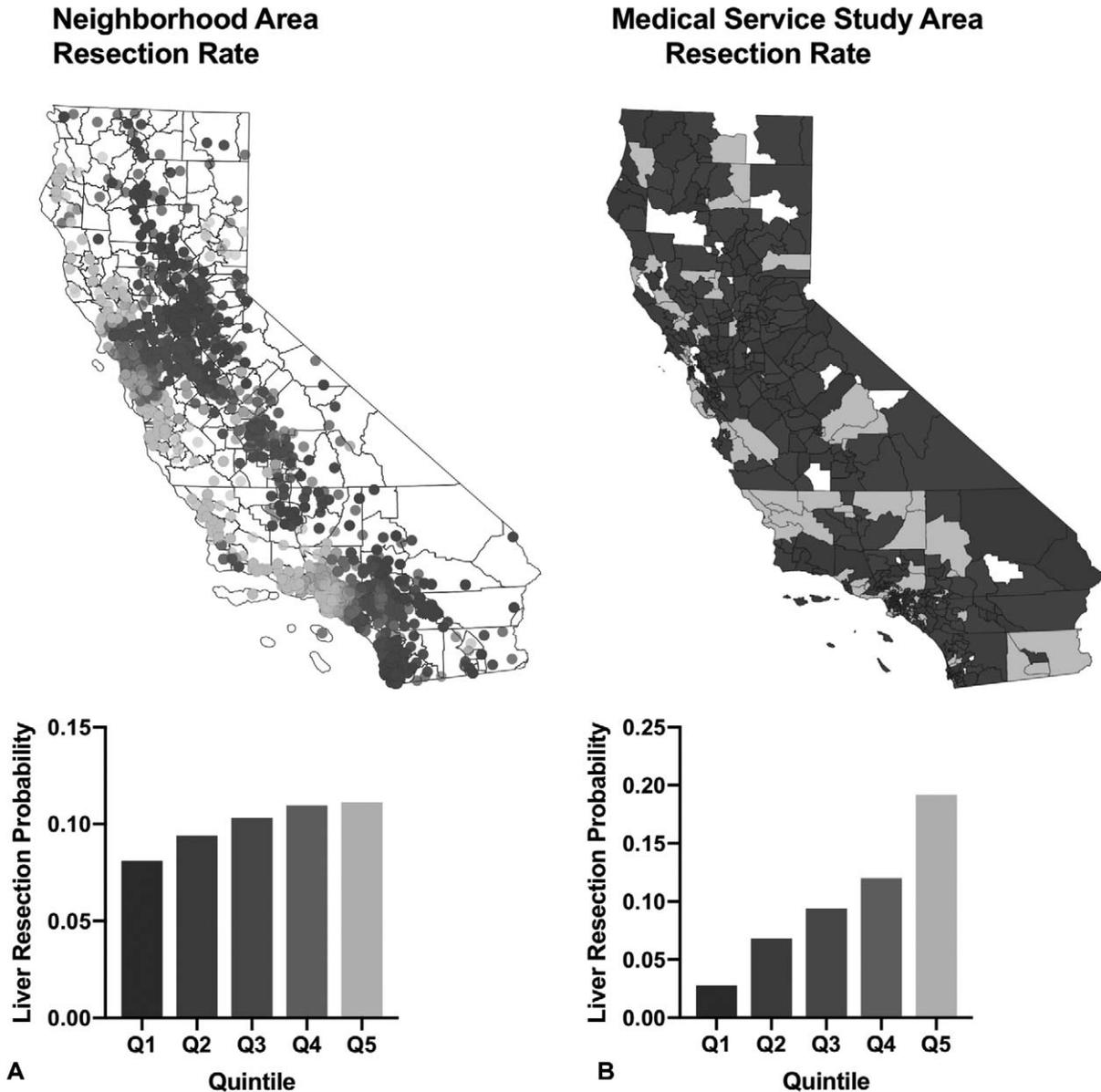


FIGURE 2. Geospatial distribution of instrumental variable quintiles (top panel) and the probability of liver resection in each of the instrumental variable quintiles (bottom panel). (A) Neighborhood-area liver resection rate. (B) Medical Service Study Area liver resection rate.

varied from 8.1% in the lowest quintile to 11.1% in the highest quintile, and MALR rates varied from 2.7% in the lowest quintile to 19.2% in the highest quintile.

First, to check that our instruments affect patient survival only through liver resection, we compared baseline patient and disease characteristics grouped by instrument variables. These results are shown in Table 1. For NALR rate categories (cut-off at median), the liver resection rate was 9% in the low group and 11% in the high group. We note that covariate characteristics were well balanced across the NALR categories. However, there were a fewer proportion of non-Hispanic white patients (51% vs 66%, ASD 0.342) and patients with T3-T4 tumors (44% vs 47%, ASD 0.121) in NALR high group. For MALR rate categories (cut-off at median), the liver resection rate was 6% in the MALR-low group and 14% in the

MALR-high group. The covariate characteristics were well balanced across MALR categories. Compared to MALR-low, the MALR-high group had a higher proportion of non-Hispanic whites (63% vs 55%, ASD 0.248). In addition, MALR-high group has a slightly higher propensity to get chemotherapy (59% vs 53%, ASD 0.137). Although the distribution of race/ethnicity was unbalanced across the categories of instruments, the groups did not differ systematically. These analyses demonstrate that the measured confounders were similarly distributed across IV categories and support the assumption that unmeasured confounders are likely to be balanced across IV groupings as well.

Effect of Area Resection Rates on Liver Resection

Next, to check the validity of the “first stage,” we regressed liver resection on NALR rates and MALR rates in 2 separate

TABLE 1. Comparison of Patient Characteristics Grouped by Treatment and Instruments, 2000–2012

Characteristics	Categories	Full Cohort	Liver Resection			Neighborhood Liver Resection Rate*			MSSA Liver Resection Rate		
			No	Yes	ASD	Low [†]	High [†]	ASD	Low [†]	High [†]	ASD
No. of patients		16,382	14,747	1635		8093	8092		8231	8151	
Liver resection (%)		10	0	100		9	11		6	14	
Age, y (%)	18–44	7	6.4	11.7	0.52	6.8	7	0.027	6.4	7.5	0.068
	45–59	27.2	25.8	40.2		26.9	27.4		26.2	28.2	
	60–80	48.2	48.9	42.7		48.9	47.6		49.5	47	
	>80	5.4	18.9	5.4		17.3	18		17.9	7.2	
Sex (%)	Male	53.1	53.1	53.2	0.002	53.9	52.2	0.034	53.4	52.9	0.011
	Female	46.9	46.9	46.8		46.1	47.8		46.6	47.1	
Charlson-Deyo Score [‡] (%)	0	75.1	74	84.8	0.305	75.1	74.9	0.037	73.8	76.3	0.056
	1	17.7	18.2	12.8		18.1	17.4		18.6	16.8	
	2+	7.2	7.8	2.4		6.8	7.7		7.5	6.9	
Race/ethnicity (%)	White (non-hispanic)	59	58.8	61	0.177	66.3	51.4	0.342	55	63	0.248
	Black	9.7	10.1	6		7.2	12.3		12.9	6.5	
	hispanic	16.6	16.7	15.7		15.6	7.7		17.7	5.5	
	Asian/PI	12.1	11.9	13.6		9	15.5		12.4	11.9	
	Middle Eastern	1.9	1.8	3.1		1.2	2.7		1.4	2.5	
Marital status (%)	Not married	46.6	48	34	0.292	44.1	48.8	0.096	48.9	44.2	0.096
	Married	53.4	52	66		55.9	51.1		51.1	55.8	
Primary location (%)	Left	59.9	59.3	65.2	0.122	60.1	59.7	0.008	39.2	41	0.038
	Right	40.1	40.7	34.8		39.9	40.3		60.8	59	
Node positive (%)	No	49.4	51	35.6	0.314	48.5	50.5	0.04	50.7	48.2	0.051
	Yes	50.6	49	64.4		51.5	49.5		49.3	51.8	
T-stage (%)	T1–2	39.5	40.1	32.9	0.664	40.3	39	0.121	39.5	39.6	0.044
	T3–4	45.7	43.6	65		47.1	44.2		45	46.4	
	Tx	14.7	16.1	2.1		12.6	16.9		15.5	14	
Extrahepatic Metastases (%)	No	44.1	43.2	52.7	0.586	42.4	45.9	0.069	43.7	44.6	0.0188
	Yes	55.9	56.8	47.3		57.6	54.1		56.3	55.4	
Primary resected (%)	No	33.3	36.6	3.6	0.19	32.5	34.1	0.034	35.5	31	0.095
	Yes	66.7	63.4	96.4		67.5	65.9		64.5	69	
Chemotherapy (%)	No	42	44.8	16.8	0.906	42.4	44.9	0.05	47	40.3	0.137
	Yes	54.2	51.3	80.8		57.6	55.1		53	59.7	

Values >0.10 suggest imbalance between the groups.

*Neighborhood resection rate could not be calculated for 197 patients who did not have a neighbor undergoing liver resection.

†Low—below median; High—above median.

‡Charlson-Deyo Score: This score is the Deyo modification of Charlson comorbidity Index. Scores range from 0 to 25, with higher scores indicating more comorbidities.

regressions. The results in Table 2 indicate that there is a strong association between area resection rates and the probability of liver resection. That is, conditional on the controls, patients with CRC-LM living in high resection areas are likely to get a liver resection themselves. In particular, the *F*-statistics for NALR rates regressions vary between 18.6 and 25.2, and the *F*-statistics for MALR rates regressions vary between 199.9 and 297.9 depending on the cut-off. All *F*-statistics are well above the standard cut-off of 10 (*P* < 0.001).

We confirmed that the instrument variable approach was necessary (as opposed to multivariable regression), by Wooldridge’s robust score test and a robust regression-based test for endogeneity that yielded statistically nonsignificant results.⁴³ This finding indicates that standard multivariable regression resulted in biased estimates compared with the IV model.

Impact of Liver Resection on Survival

Survival estimates using different statistical approaches are summarized in Table 2. First, we used unadjusted univariate regression analysis to evaluate the effect of liver resection on survival. There was a highly significant association between proportion of patients surviving at 1 year, 2 years, 3 years, 4 years, or 5-year time points and liver resection. The strength of this association decreased from 0.52 at 1 year to 0.28 at the 5-year mark. After adjusting for known covariates that impact survival, the magnitude of association

decreased but remained significant (coefficient, ranging from 0.26 at 5 years to 0.46 at 2 years). Race/ethnicity was not independently associated with survival in the multivariable models and had minimal effect on estimates (Table S2). This is likely because sociodemographic differences attributable to race/ethnicity were captured by the fixed-effects socioeconomic index variable. Adjusted OLS regression analysis does not address unmeasured confounders such as those related to patient selection and patient choice for surgery. As discussed earlier, this association is not necessarily causal.

Next, we use the 2 instruments separately and then combined to estimate the causal effect of liver resection on patient survival. The estimates in Panel C of Table 2 indicate that regardless of the cut-off used for the instrument variable, liver resection significantly improves survival at the 1-year mark. In particular, liver resection increases probability of 1-year survival by 0.80 to 0.92 depending on the cutoff used. Moreover, using the 75th percentile cutoff indicates that survival improves at all 1- to 5-year time points. This translates to a 34.5 month (SE 11.9, *P* < 0.001, shown in Table 2 Panel C) gain in survival for marginal patients. We then run similar regressions using MALR rates as the instrument in Panel D of Table 2. Again, the results indicate significant improvements in 1- to 5-year survival (regardless of the cutoff used). These estimates translate to a statistically significant survival gain of 14.4 to 30.5 months (shown in Table 2, Panel D). Although the estimates in Panels C and D are

TABLE 2. Instrument Variable and OLS Estimates of Effectiveness of Liver Resection on Survival.

Analysis Method	Panel	Instrument	Instrument		Avg. Treatment Effect of Liver Resection on Survival					
			Cut-off	F-statistic	ΔSurvival	mo	1 y	2 y	3 y	4 y
Unadjusted OLS	A	na		na	25.35* (0.772)	0.52* (0.008)	0.57* (0.011)	0.51* (0.013)	0.39* (0.013)	0.28* (0.013)
Adjusted OLS	B	na		na	20.29* (0.751)	0.35* (0.009)	0.46* (0.011)	0.45* (0.013)	0.35* (0.013)	0.26* (0.013)
Instrumental variable	C	Neighborhood resection rate	25th Percentile	18.6*	27.41 (21.697)	0.90† (0.387)	0.53 (0.339)	0.31 (0.296)	0.09 (0.243)	0.37‡ (0.209)
			Median	24.8*	24.23 (16.930)	0.92† (0.318)	0.69† (0.288)	0.28 (0.240)	0.12 (0.219)	0.18 (0.208)
D	MSSA resection rate	75th Percentile	25.2*	34.52* (11.905)	0.80* (0.306)	0.66† (0.272)	0.51† (0.208)	0.38† (0.175)	0.40† (0.160)	
		25th Percentile	297.9*	30.53* (4.939)	0.63* (0.111)	0.65* (0.098)	0.49* (0.081)	0.29* (0.068)	0.24* (0.058)	
E	Neighborhood and MSSA resection rate	Median	266.8*	23.39* (4.334)	0.33* (0.092)	0.47* (0.084)	0.45* (0.071)	0.27* (0.063)	0.18* (0.054)	
		75th Percentile	199.9*	14.43* (4.244)	0.33* (0.088)	0.44* (0.083)	0.44* (0.072)	0.25* (0.065)	0.16* (0.057)	
				Both 25th percentile	30.88* (5.003)	0.65* (0.112)	0.66* (0.098)	0.48* (0.082)	0.28* (0.069)	0.24* (0.059)
				Both median	23.65* (4.364)	0.43* (0.091)	0.49* (0.083)	0.45* (0.071)	0.26* (0.063)	0.17* (0.055)
				Both 75th percentile	16.64* (4.133)	0.35* (0.088)	0.46* (0.082)	0.45* (0.071)	0.27* (0.064)	0.19* (0.057)

Results are presented as coefficients (robust standard errors). Adjusted OLS and Instrumental variable analysis controlled for age, comorbidities, marital status, tumor laterality, node-status, T-stage, extra-hepatic metastases, and chemotherapy. In addition, year of diagnosis and county-level socioeconomic index were included as fixed effects in the models.
 *Statistically significant at 0.99 confidence.
 †Statistically significant at 0.95 confidence.
 ‡Statistically significant at 0.90 confidence.

slightly different, we note that overall the estimates indicate a gain in survival after liver resection and are mostly statistically significant. The estimates vary depending on the instrument used suggesting that each instrument affects a different set of marginal patients. Lastly, we repeat the analysis by including both MALR and NALR rates as instruments in Panel E Table 2. Not surprisingly, the estimates are more aligned to the stronger instrument in the model and more accurately depict the magnitude of survival change for the marginal patients.

At median cut-offs for both instruments, we find a 23.6 months' improvement in survival (SE 4.4, $P < 0.001$, shown Table 2 Panel E) with liver resection. The coefficients are percentage point increase in the probability of liver resection. At median cut-offs for both instruments (Panel E), the coefficients are interpreted to represent an increase in survival of 105% at 1 year, 223% at 2 years, 375% at 3 years, 356% at 4 years, and 370% at 5 years attributable to liver resection.

DISCUSSION

Liver resection is an important component of multimodality therapy for patients with CRC-LM at expert centers. Most evidence for its use comes from institutional series, multi-institutional retrospective analyses, and population-based cohort analyses.^{2-18,21-27,34} These studies are limited by an inherent bias owing to rigorous patient selection. A randomized trial can minimize bias and provide accurate estimates of the benefit of liver resection. However, such a trial is unlikely to accrue patients as liver resection for CRC-LM is now well engrained in clinical practice. Similarly, at the population-level, it remains unknown whether increasing liver resection rates will improve the survival of patients with CRC-LM. This finding would have important health policy implications.

In this study, we use IV analysis to obtain population-level causal effects of liver resection on patient survival. This analysis takes advantage of the natural variation in the use of liver resection owing to factors other than those that may impact patient survival. Conceptually, patients with CRC-LM can be categorized into 3 groups: those who will always get liver resection (eg, solitary lesion, no comorbidities, no extrahepatic metastases); those who never get liver resection (eg, poor performance status, technically unresectable liver metastases, multiple extrahepatic sites of metastases); and those who fall in the middle (marginal patients). For marginal patients, the choice of liver resection is directly influenced by the regional practice patterns (ie, area resection rates). Our results show that for marginal patients during 2000 to 2012, increasing rate of liver resection of their residence area (neighborhood or MSSA) would have significantly and positively impacted survival. These observations also demonstrate that during this time period, liver resection was underutilized in the treatment of CRC-LM in California.

There are several limitations to the IV analysis. First, our estimates are valid as long as the assumptions of the instrumental variable analysis outlined in Figure 1 are satisfied. We demonstrate in our analysis that both instruments meet these assumptions. Despite this evidence, there remains a possibility that the IV grouping may be asymmetric in unmeasured confounders and as such this assumption cannot be tested. For instance it is possible that people pick residence or places to retire with higher-quality healthcare. In our analysis, high and low resection rate areas remain balanced in the use of primary tumor resection. This provides an internal (placebo) control that suggests that access to surgical services was equally present in both areas. We also control for factors that might be associated with overall quality of healthcare in our instrumental analyses including those attributable to patient, disease, and geographic characteristics as best possible. Second, a single instrument is often inadequate to capture all the marginal patients.³⁶ For that reason we employ 2

instruments (alone and in combination) in our analysis. Third, using IV analysis to model survival data is challenging. We use a linear probability model for our IV estimates, as their use in IV analysis is well established in the literature. These models may yield erroneous estimates near the extremes, which was not the case in the present study. We also model the outcome in 2 different ways (binary and continuous). The continuous models were limited by approximately 9% missing data. These are the patients who were still living at the end of the follow-up period and had not completed their life span. The binary models circumvent this problem by providing estimates at annual intervals. Fourth, our estimates provide the relative increase in survival of marginal patients. However, it is not possible to delineate who these patients are and what is the baseline survival for the marginal patients who do not undergo liver resection. Nonetheless, our models provide relative estimates of the magnitude of survival increase because of liver resection in these marginal patients, and this may serve as a starting point for quality improvement initiatives. Fifth, our findings may not be generalizable to other states or time periods that have markedly different liver resection rates because characteristics of patients defined as marginal will vary. In particular, newer alternative therapies (biologics, embolization, arterial infusion, and so on) may improve survival and limit generalization to present day clinical practice. However, this study encompasses a period where most of the therapies used today were already in clinical practice.⁴⁴ Finally, our estimates apply exclusively to patients with synchronous CRC-LM. We could not reliably ascertain the timing and occurrence of metachronous CRC-LM in the nonoperative patients and therefore restricted our analyses to patients with synchronous CRC-LM. Based on previous observational studies, we speculate that patients with metachronous CRC-LM may benefit even more from liver resection, but this assertion remains to be tested.

Patient selection remains paramount. These data do not support a nonselective strategy to liver resection. Instead, the interpretation of this analysis is that *given the same patient selection* that was in high resection areas for marginal patients, the survival of low-resection area marginal patients would have improved if they had liver resection. We agree that given the high rate of recurrence after liver resection, patients should be selected based on the biology of the tumor. One way to assess this is to select patients who have a response to systemic therapy. An analysis of how the patients should be selected for liver resection is beyond the scope of this analysis.

The results have implications for directing health policies that increase liver resection rates with the goal of improving survival of patients with CRC-LM. For this purpose, the policymakers must assume that the patients whose liver resection rates will be increased by the policy are similar to those whose treatment choice was most affected by the instruments in the study (ie marginal patients). Contingent on this assumption and provided that similar data can be obtained from other states, our results indicate that national measures to increase liver resection in the United States will significantly improve survival of patients with CRC-LM. Further work is needed to identify barriers to liver resection and access to multidisciplinary care.

REFERENCES

1. National Cancer Institute: Surveillance EaERP. Cancer Stat Facts: Colorectal Cancer 2008-2014. Available at: <https://seer.cancer.gov/statfacts/html/colorect.html>. Accessed March 14, 2019, 2019.
2. Stewart CL, Warner S, Ito K, et al. Cytoreduction for colorectal metastases: liver, lung, peritoneum, lymph nodes, bone, brain. When does it palliate, prolong survival, and potentially cure? *Curr Probl Surg*. 2018;55:330–379.
3. Ito K, Govindarajan A, Ito H, et al. Surgical treatment of hepatic colorectal metastasis: evolving role in the setting of improving systemic therapies and ablative treatments in the 21st century. *Cancer J*. 2010;16:103–110.

4. Tomlinson JS, Jarnagin WR, DeMatteo RP, et al. Actual 10-year survival after resection of colorectal liver metastases defines cure. *J Clin Oncol*. 2007;25:4575–4580.
5. Nikfarjam M, Shereef S, Kimchi ET, et al. Survival outcomes of patients with colorectal liver metastases following hepatic resection or ablation in the era of effective chemotherapy. *Ann Surg Oncol*. 2009;16:1860–1867.
6. House MG, Ito H, Gonen M, et al. Survival after hepatic resection for metastatic colorectal cancer: trends in outcomes for 1,600 patients during two decades at a single institution. *J Am Coll Surg*. 2010;210:744–752. 752-5.
7. Fong Y, Fortner J, Sun RL, et al. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. *Ann Surg*. 1999;230:309–318. discussion 318-21.
8. Scheele J, Stang R, Altendorf-Hofmann A, et al. Resection of colorectal liver metastases. *World J Surg*. 1995;19:59–71.
9. Nordlinger B, Guiguet M, Vaillant JC, et al. Surgical resection of colorectal carcinoma metastases to the liver. A prognostic scoring system to improve case selection, based on 1568 patients. *Association Francaise de Chirurgie Cancer*. 1996;77:1254–1262.
10. Fernandez FG, Drebin JA, Linehan DC, et al. Five-year survival after resection of hepatic metastases from colorectal cancer in patients screened by positron emission tomography with F-18 fluorodeoxyglucose (FDG-PET). *Ann Surg*. 2004;240:438–447. discussion 447-50.
11. Adam R, Avisar E, Ariche A, et al. Five-year survival following hepatic resection after neoadjuvant therapy for nonresectable colorectal. *Ann Surg Oncol*. 2001;8:347–353.
12. Kato T, Yasui K, Hirai T, et al. Therapeutic results for hepatic metastasis of colorectal cancer with special reference to effectiveness of hepatectomy: analysis of prognostic factors for 763 cases recorded at 18 institutions. *Dis Colon Rectum*. 2003;46(10 suppl):S22–S31.
13. Choti MA, Sitzmann JV, Tiburi MF, et al. Trends in long-term survival following liver resection for hepatic colorectal metastases. *Ann Surg*. 2002;235:759–766.
14. Minagawa M, Makuuchi M, Torzilli G, et al. Extension of the frontiers of surgical indications in the treatment of liver metastases from colorectal cancer: long-term results. *Ann Surg*. 2000;231:487–499.
15. Jamison RL, Donohue JH, Nagorney DM, et al. Hepatic resection for metastatic colorectal cancer results in cure for some patients. *Arch Surg*. 1997;132:505–510. discussion 511.
16. Gayowski TJ, Iwatsuki S, Madariaga JR, et al. Experience in hepatic resection for metastatic colorectal cancer: analysis of clinical and pathologic risk factors. *Surgery*. 1994;116:703–710. discussion 710-1.
17. Hughes KS, Simon R, Songhorabodi S, et al. Resection of the liver for colorectal carcinoma metastases: a multi-institutional study of patterns of recurrence. *Surgery*. 1986;100:278–284.
18. Fortner JG, Fong Y. Twenty-five-year follow-up for liver resection: the personal series of Dr. Joseph G. Fortner. *Ann Surg*. 2009;250:908–913.
19. Fakhri MG. Metastatic colorectal cancer: current state and future directions. *J Clin Oncol*. 2015;33:1809–1824.
20. Giordano SH, Kuo YF, Duan Z, et al. Limits of observational data in determining outcomes from cancer therapy. *Cancer*. 2008;112:2456–2466.
21. Nanji S, Mackillop WJ, Wei X, et al. Simultaneous resection of primary colorectal cancer and synchronous liver metastases: a population-based study. *Can J Surg*. 2017;60:122–128.
22. Angelsen JH, Horn A, Sorbye H, et al. Population-based study on resection rates and survival in patients with colorectal liver metastasis in Norway. *Br J Surg*. 2017;104:580–589.
23. Booth CM, Nanji S, Wei X, et al. Surgical resection and peri-operative chemotherapy for colorectal cancer liver metastases: A population-based study. *Eur J Surg Oncol*. 2016;42:281–287.
24. Sultana A, Meng R, Piantadosi C, et al. Liver resection for colorectal cancer metastases: a comparison of outcomes over time in South Australia. *HPB (Oxford)*. 2018;20:340–346.
25. Booth CM, Nanji S, Wei X, et al. Management and outcome of colorectal cancer liver metastases in elderly patients: a population-based study. *JAMA Oncol*. 2015;1:1111–1119.
26. t Lam-Boer J, Al Ali C, Verhoeven RH, et al. Large variation in the utilization of liver resections in stage IV colorectal cancer patients with metastases confined to the liver. *Eur J Surg Oncol*. 2015;41:1217–1225.
27. Hackl C, Neumann P, Gerken M, et al. Treatment of colorectal liver metastases in Germany: a ten-year population-based analysis of 5772 cases of primary colorectal adenocarcinoma. *BMC Cancer*. 2014;14:810.
28. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2001.

29. Sheetz KH, Norton EC, Regenbogen SE, et al. An instrumental variable analysis comparing medicare expenditures for laparoscopic vs open colectomy. *JAMA Surg.* 2017;152:921–929.
30. McDowell BD, Chapman CG, Smith BJ, et al. Pancreatectomy predicts improved survival for pancreatic adenocarcinoma: results of an instrumental variable analysis. *Ann Surg.* 2015;261:740–745.
31. Hamada T, Tsujino T, Isayama H, et al. Pseudorandomization using an instrumental variable: a strong tool to break through selection bias. *Gastrointest Endosc.* 2012;76:1079–1080.
32. Registries NAACC. Certified Registries 2019. Available at: <https://www.naacr.org/certified-registries/>. Accessed March 14, 2019.
33. Zingmond DS, Ye Z, Ettner SL, et al. Linking hospital discharge and death records—accuracy and sources of bias. *J Clin Epidemiol.* 2004;57:21–29.
34. Leung U, Gonen M, Allen PJ, et al. Colorectal cancer liver metastases and concurrent extrahepatic disease treated with resection. *Ann Surg.* 2017;265:158–165.
35. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992;45:613–619.
36. Brooks JM, Chrischilles EA, Scott SD, et al. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res.* 2003;38(6 pt 1):1385–1402.
37. Development COoSHPa. Medical Service Study Area 2010. Available at: <https://oshpd.maps.arcgis.com/home/item.html?id=a201100c4bf374bd081bb49b82cbaaac3>. Accessed March 14, 2019, 2019.
38. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med.* 2009;28:3083–3107.
39. Martinez-Cambor P, Mackenzie T, Staiger DO, et al. Adjusting for bias introduced by instrumental variable estimation in the Cox proportional hazards model. *Biostatistics.* 2019;20:80–96.
40. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *J Business Econ Stat.* 2001;19:2–16.
41. Chow G. Tests of equality between sets of coefficients in two linear models. *Econometrica.* 1960;28:591–605.
42. Yost K, Perkins C, Cohen R, et al. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control.* 2001;12:703–711.
43. Wooldridge J. *Score diagnostics for linear models estimated by two stage least squares.* Oxford: Blackwell; 1995.
44. Kopetz S, Chang GJ, Overman MJ, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *J Clin Oncol.* 2009;27:3677–3683.

DISCUSSANTS

H. Chen (on behalf of Dr. Douglas Tyler (Galveston, TX))

I rise to read Dr. Doug Tyler's questions. He could not attend because of an urgent family medical issue.

One of the interesting and unusual aspects of this article is that it utilizes geographic variation in surgical procedure rates to support using a procedure, that is, liver surgery in this instance, in patients with synchronous colorectal liver metastasis. Many studies that look at geographic variation in utilization of surgical procedures usually find no significant improvement in targeted outcomes and question the high use of the procedure.

However, there should be some caution in the use of geography as an unbiased instrument in this analysis based on the assumption that where people live is independent of targeted health care needs. Increasingly, people do pick places to live and retire that do have easy access to higher-quality health care even without knowing what specific types of health care they may need down the road which I think can confound the current analysis.

Dr. Tyler has 5 questions for you, and I'll try to be brief.

Number 1, how do individuals compare from a given location who have surgery within their geographically defined residence as compared to those who have surgery outside their geographic residence as defined by your 2 instruments?

Number 2, were there differences in survival by geographic location within high-volume liver resection regions?

Number 3, during the 12 years of your study, how does this analysis account for the possibility that geographic area might have changed from a low- to a high-volume liver resection region or vice versa.

Number 4, was there a way to use this type of analysis to evaluate the effect of timing of the liver resection or the use of chemotherapy to identify optimal pathways for patients?

Number 5, can you use this analysis akin to the Dartmouth atlas and identify areas that should be doing more or less surgery?

Dr. Tyler appreciates the opportunity to discuss this well-presented article and encourages the authors to further their work into defining the role of optimal timing of liver resection in the management of patients with colorectal liver metastases.

Response From M. Raof

I thank Dr. Chen for narrating Dr. Tyler's comments, and thank him for his valuable insight that really gets at the crux of this analysis. I will address them point by point.

The first point was asking; is the instrument truly exogenous? People plan migration based on the future need for health care. This is a very interesting and important observation. Our IV analysis assumes that the liver resection rate in an individual's place of residence is exogenous to the patient (ie, the patient did not choose to live in an area with a high vs low liver resection rate.) Although plausible, there is not much research on whether people move for medical benefits. For instance, a recent 2016 study in the Journal of Policy and Management found that Medicaid expansion had no impact on across state migration. Certainly, patients with the means to afford private insurance, likely factor in quality of healthcare in their decision-making when choosing future residence.

To the reviewer's point we found a slight imbalance in the use of chemotherapy with the MSSA instrument. However, this was not the case for neighborhood area instrument. Further in our analysis, high and low resection rate areas remain balanced in the use of primary tumor resection. This provides an internal (placebo) control that suggests that access to surgical services was equally present in both areas. Both instruments pointed to similar results.

Finally, we control for these factors in our instrumental analyses including those attributable to patient, disease, and geographic characteristics as best possible. We also used 2 instruments. The neighborhood area instrument includes overlapping geographic areas (patient-centric) and MSSA instrument consists of nonoverlapping predefined geographic areas (area-centric). Two ways of classifying patients mitigate the chance that the observed effects are because of access of "other" health care rather than receipt of liver resection.

In regards to the second question; how do patients compare from a given location who have surgery within versus outside resection area? For neighborhood areas, ~83% had resection in their geographic area (50-mile radius). We compared these patients to those who had resection outside of their neighborhood. In general, those who had surgery outside their neighborhood were younger, had fewer comorbidities, more likely to be non-Hispanic white, and more likely to receive systemic chemotherapy. We will include these results in the manuscript.

The third question was; are their differences in survival by geographic location within high-volume liver resection regions? This is hard to address from these data. As we subsegment the data, it

becomes increasingly hard to draw meaningful conclusions in a heterogeneous patient population from smaller sample sizes.

Fourth, how do we account for variation during the 12 years of the study? We saw some variation in resection rates over time for a given geographical region. For instance, for MSSA resection rate, approximately two-thirds of patients consistently remain in the group assigned based on the 13-year aggregate (ie, high vs. low resection groups at a median cut-off) regardless of the time period. However, one-third of patients could be in a different grouping based on the time period. Despite this variation, the 13-year aggregate rate for the MSSA was chosen as the instrument because the resection rate estimates get more precise with higher number of patients. We believe that this trade off was an acceptable compromise. We show that the first stage of the IV analysis was satisfied, that is, dichotomization of the instrument based on 13-year aggregate resection rate was associated with the receipt of liver resection. This means that our instrument captured the treatment philosophy that the patient encountered when the treatment decision was made. To account for temporal trends, we used year of diagnosis as a fixed effect in all multivariable models.

Fifth question was regarding timing of liver resection with respect to chemotherapy to define optimal pathways. I think that the population-level data are not granular enough to answer the question regarding timing of liver resection or the use of perioperative chemotherapy. These questions can best be addressed in the context of innovative clinical trials.

The final comment was about Dartmouth atlas analysis. Our maps are very similar to those in the Dartmouth atlas. The geographic segmentation differs slightly. However, we believe the benchmarking should be hospital- or hospital network-driven and was not the emphasis of the study but certainly can be addressed in future studies.

J.B. Dimick (Ann Arbor, MI)

Thank you for a really nice presentation. I received the article ahead of time. It was really well done, really airtight from a methodical perspective. Very sophisticated methods, and a clever way to get at what's probably the right answer to this question.

You didn't mention it in your talk, I don't think, but you did find that the non-IV estimates were actually a little higher in terms of survival. When you adjust them, probably the observational data we have are overestimating, so still a very strong, positive effect of survival.

I have one question for you. My question gets to the crux of instrument variables. The big assumption that is untestable, at least directly, empirically, is this notion that the instrument isn't associated with the outcome except through the differences in treatment, in this case liver resection. So you can't study that directly. You have to have a leap of faith there. There is a way to study empirically indirectly, and that is by a placebo test.

My question to you is, did you look for a placebo tracer condition? For example, stage 3 colorectal cancer. Does survival for stage 3 colorectal cancer differ between the areas on one side or the

other of your IV partition or stage 4 lung cancer? Because if the populations are different in a way that drives survival, I noticed that the people that more likely to get liver resection were also more likely to get chemotherapy. If they are more likely to get all kinds of good health care, that could systematically bias, and you can test that with an indirect placebo test. Did you do that, and do you plan to do that?

Response From M. Raouf

That is a great point. In our analysis, high and low resection rate areas remain balanced in the use of primary tumor resection. This provides an internal (placebo) control that suggests that access to surgical services was equally present in both areas and the differences we observe are due to philosophy, facilities, and capabilities pertinent to liver resection. Thank you.

P. Clavien (Zurich, Switzerland)

Congratulations, and thank you for this important message. However, understanding the subtleties of the methodology in such a population-based survey is sometimes difficult. We have drawn similar conclusions in Europe with examples relating to pancreatic cancer in France and Germany, as well as a comparable population-based study for resection of colorectal liver metastases in England, which will be presented at the upcoming ESA meeting in Madrid, and published in the November issue of *Annals of Surgery*.

This brings me to my question: who should we blame for the discrepancies in the availability of complex procedures? It is difficult to accept that patients are not being offered a life-saving procedure because of where they live. Should we investigate why some surgeons are unable to perform such a procedure? Why do the oncologists or gastroenterologist not refer these patients to the relevant centers? Is it the patients, who do not wish to go to such centers? I would like to understand why the search for the most appropriate therapy is currently being thwarted.

I understand that your current data are insufficient to answer these questions with confidence. However, could you please offer some advice as to what action must be taken to correct such discrepancies?

Response From M. Raouf

That is a great question, Dr. Clavien. It's multifactorial. Looking at this data, obviously we can't capture the physician preferences in the area. These physicians include not only the surgical oncologists but also medical oncologists. It is not clear how many patients don't even see a surgeon for this potentially curable problem. There's also a lot of structural basis to this disparity between different regions, and I think studying some of that disparity is possible through population-level data, and that is one of our next steps. We also plan to do surveys that will be given to medical oncologists to determine what their perceptions are about liver resection; how do they view these data that's been out there for >2 decades and has not been picked up readily.