



Published in final edited form as:

Liver Transpl. 2020 August ; 26(8): 977–988. doi:10.1002/lt.25787.

Risk Factors and Center-Level Variation in Hepatocellular Carcinoma Understaging for Liver Transplantation

Nadim Mahmud, MD MS MPH MSCE^{1,2}, Maarouf A. Hoteit, MD¹, David S. Goldberg, MD MSCE³

¹Division of Gastroenterology, Hospital of the University of Pennsylvania, Philadelphia, PA

²Leonard David Institute of Health Economics, University of Pennsylvania, Philadelphia, PA

³Division of Digestive Health and Liver Diseases, University of Miami Miller School of Medicine, Miami, FL

Abstract

Background & Aims: Liver transplantation (LT) is curative for most patients with hepatocellular carcinoma (HCC), however 10–15% of patients experience HCC recurrence. Patients who are reported as within Milan criteria by imaging are frequently found to be outside criteria on explant. This understaging of HCC worsens post-LT outcomes, however risk factors for understaging have not been elucidated. Furthermore, it is not known if there is regional or center-level variation in understaging.

Methods: We conducted a retrospective analysis of adult patients transplanted for HCC in the United Network for Organ Sharing (UNOS) database between 2012 and 2016. Understaging was determined on the basis of comparing pre-LT imaging to explant findings. Kaplan-Meier methods and Cox regression were used to evaluate the impact of understaging on HCC recurrence and post-LT survival. Mixed-effects logistic regression was used to identify risk factors for understaging, and to study regional and center-level variation in adjusted analyses.

Results: A total 5,424 patients were included in the cohort, of whom 24.9% (n=1,353) were understaged. Post-LT HCC recurrence and death were significantly associated with understaging (each $p < 0.001$). In adjusted analyses, independent predictors of understaging included age (OR 1.13 per 10 years, 95% CI 1.03 – 1.25), male sex (OR 1.61, 95% CI 1.36 – 1.89), downstaging (OR 4.03, 95% CI 2.65 – 6.11), and pre-LT AFP ($p < 0.001$). There was also significant variation in understaging between UNOS regions and among transplant centers, ranging from 14.8% to 38.1%.

Conclusions: We report novel risk factors for HCC understaging, which worsens post-LT outcomes. Significant center-level and regional variation in understaging highlights the need for standards that achieve greater uniformity in staging.

Corresponding Author: Nadim Mahmud MD MS MPH MSCE, nadim@pennmedicine.upenn.edu, Phone: 215-349-8222, Fax: 215-349-5915, 3400 Civic Center Boulevard, 4th Floor, South Pavilion, Philadelphia, PA 19104.

Author Contributions:

Intellectual genesis – NM, MH, DG

Data cleaning and analysis – NM, DG

Manuscript writing – NM

Critical manuscript review – NM, MH, DG

Keywords

United Network for Organ Sharing (UNOS); mixed-effects logistic regression; pathology; explant; Milan criteria

Introduction

Hepatocellular carcinoma (HCC) is a primary hepatic malignancy that typically occurs in the setting of cirrhosis.¹ For selected patients liver transplantation (LT) can be curative, however severity of patient illness has historically been poorly-reflected by the model for end-stage liver disease (MELD) score.² As such, when MELD-based allocation was adopted in 2002, the United Network for Organ Sharing (UNOS) also created a standardized exception points pathway for qualifying HCC patients. This pathway is principally based on the Milan criteria, specifying that a patient must have one tumor ≤ 5 cm or no more than three tumors, each ≤ 3 cm in the maximum diameter.³ In the seminal work behind these criteria, Mazzaferro et al found that patients transplanted within Milan criteria had significantly improved post-LT survival relative to those outside of the criteria.⁴ Because tumor progression is expected to occur while awaiting LT, locoregional therapies (LRT) such as transarterial chemoembolization or radiofrequency ablation are commonly used as bridging therapies to maintain patients within criteria. However, LRT may also be used to downstage patients who are initially outside Milan criteria, such that they may ultimately become eligible for LT.⁵

While HCC has become a major indication for LT, the rates of HCC recurrence are as high as 8–20% at five years.^{6, 7} Once recurrence is diagnosed, median survival is less than 1 year.⁸ There are numerous predictors of post-LT HCC recurrence and poor survival, including elevated pre-LT alpha-fetoprotein (AFP), shorter time on the waiting list, and etiology of liver disease.^{9–11} In order to receive MELD exception points for patients with HCC, centers must submit data on tumor size, number, and imaging criteria to ensure a patient meets MELD exception criteria. Over time, this process has become more standardized, and centers are subject to audits. However, the determination of the exact tumor size (e.g., 2.8 vs 3.1cm) and whether the tumor meets UNOS criteria for a definite HCC are left to each transplant center.

It has been shown that patients transplanted for HCC may have a greater number of tumors on explant than what was visualized on radiologic imaging as computed tomography (CT) and magnetic resonance imaging (MRI) are not 100% sensitive or specific.¹² Our group has shown that as many as 40% of LT recipients with HCC may have ‘occult’ lesions on explant pathology.¹² It is unknown, however, if the presence of occult lesions on explant is a random process, or whether there are specific patient, clinical, demographic, or tumor-related variables (e.g., size, treatment type) associated with understaging (having a greater tumor burden on explant than what was evident on radiologic imaging). Furthermore, there are no data on whether understaging varies across different geographic regions of the US, and more specifically, among centers located in the same geographic area. To address these questions, we performed a large analysis with national LT registry data of patients transplanted for HCC with accompanying explant data.

Methods

Study Design and Cohort Creation

We performed a retrospective cohort study using UNOS national LT registry data from April 2012 to September 2016. We chose these dates because beginning in April 2012, UNOS began to systematically collect patient explant data (detailed below), and we wanted to allow several years of follow-up for secondary outcomes ascertainment. We included all waitlisted patients age ≥ 18 who were transplanted for a primary indication of HCC with standardized T2 MELD exceptions points. We excluded patients who were known to be outside Milan criteria at the time of LT, did not have any pre-transplant AFP values, or were missing all data related to tumor number and size.

Exposure Variable Collection

In the UNOS database, we collected complete information on patient demographics (age at LT, sex, race), body mass index (BMI) at listing, MELD at listing, MELD at transplant, pre-transplant AFP, transplant center code, and UNOS region of transplant. We chose to use the most recent AFP prior to transplant, as opposed to maximum pre-LT AFP, because these have been shown to perform similarly in numerous studies addressing post-LT HCC recurrence and outcomes.^{9, 13, 14} Etiology of liver disease was classified as hepatitis C, hepatitis B, alcoholic liver disease, non-alcoholic steatohepatitis, autoimmune, or other. On most recent imaging prior to LT, we collected tumor diameter (in cm) and tumor number data. We also categorized the imaging modality (CT or MRI) and time from most recent imaging to LT (in days). LRT coding included embolization, ablation, or radiation-based treatments. Patients were considered to be downstaged if they were initially outside of Milan criteria, received LRT, and were subsequently within Milan criteria on updated imaging. Patients who otherwise received LRT were considered to have received bridging therapy. Finally, regarding explant data, we collected data on tumor number and diameter, tumor differentiation, vascular invasion, lymph node involvement, and extrahepatic spread.

Outcomes Ascertainment

The primary outcome was tumor understaging. This was defined as being within Milan criteria on the most recently submitted pre-LT imaging data (based on tumor number and diameter), but outside Milan criteria on explant, including if macrovascular invasion, positive lymph nodes, or extrahepatic spread was noted. We included explant HCC tumors that were viable or necrotic, but excluded explant tumors <1 cm in size, as these cannot be called HCC lesions by current imaging criteria. Secondary outcomes included post-LT HCC recurrence and patient death. We ascertained HCC recurrence using previously-validated methods.¹⁵ In brief, this included post-LT classifications of “recurrence of pre-transplant malignancy” or “death from HCC or metastatic malignancy.” Importantly, in a sensitivity analysis isolated to patients with reported explant tumors <1 cm, there were no significant differences in post-transplant HCC recurrence ($p=0.48$) or mortality ($p=0.58$) between understaged and correctly staged patients, suggesting that these lesions do not meaningfully contribute to risk in these groups (data not shown).

Descriptive Analysis and Post-transplant Outcomes

Descriptive statistics were presented using medians and interquartile ranges (IQRs) for continuous variables and stratified by understaging status. Chi-squared and Wilcoxon rank-sum tests were used to compare categorical and continuous variables, respectively, using a $p < 0.05$ threshold for statistical significance. In comparisons of understaging by UNOS regions, we created an accompanying geographic heat map. To identify unadjusted associations between understaging and post-LT HCC recurrence or death, we first compared proportions using chi-squared tests and presented Kaplan-Meier survival estimates at 12, 24, and 36 months. We then performed multivariable Cox regression analysis adjusting *a priori* for age, sex, pre-LT AFP, waitlist time, vascular invasion, maximum explant tumor size, and LRT/downstaging, based on prior literature.^{7, 9, 11, 14} Hazard ratios (HRs) with 95% confidence intervals (CIs) were computed, along with plots of adjusted survival curves.

Primary Statistical Analysis

Because a number of poorly-quantified center or regional factors likely impact understaging risk (e.g., listing practices, center-specific selection practices, differences in imaging/interpretation), we used mixed-effects logistic regression to identify variables associated with HCC understaging. In particular, we designated transplant centers and UNOS regions as random intercepts, as many of the above factors vary at these levels. Importantly, this approach accounts for correlation of outcomes related to clustering of observations. For example, patients transplanted at the same center may have similar outcomes due to center-specific practices or differences in case mix. To confirm that significant clustering occurred at the level of the transplant center and UNOS regions, we performed likelihood ratio tests to compare a fixed effects model (i.e. no random intercepts) to single-level (clustering about UNOS regions) and two-level (clustering about UNOS regions and transplant centers) models, using an $\alpha = 0.05$ threshold.

To identify factors associated with understaging, we began with univariate analysis. Each continuous variable was plotted against the outcome using locally weighted scatterplot smoothing curves to evaluate the linearity assumption for subsequent regression. Through this analysis, we used restricted cubic splines to model a non-linear relationship between pre-LT AFP and understaging. Univariable regression was then performed for each potential predictor, using a $p < 0.10$ threshold for testing in multivariable analysis. Of note, in light of prior evidence that imaging to LT duration of >90 days increases the likelihood of staging errors,¹⁶ we modeled imaging to LT as both a continuous variable and as a dichotomous variable with a cutpoint of 90 days. We used backwards stepwise selection to identify candidate models, using a $p < 0.05$ threshold for statistical significance. This was followed by several clinician-driven models, where a minimized Bayesian Information Criterion value was used to select a final model. This included a model testing an *a priori* interaction term between pre-LT AFP and LRT bridging/downstaging status, which was not statistically significant. Odds ratios (ORs) and associated 95% CIs were reported for each predictor in the final model. To visualize the relationship between pre-LT AFP and the probability of understaging, we plotted median spline curves in unadjusted and adjusted regression frameworks. We also plotted this relationship as stratified by LRT bridging/downstaging status.

Sensitivity Analyses

Given the possibility that LRT could induce a necrosis zone on explant larger than the original tumor size on imaging, and thus lead to misclassification of understaging, we performed two additional sensitivity analysis. First, we excluded all patients who received LRT (n=4,706), and repeated analyses of post-transplant outcomes as stratified by patients who were understaged. Second, we excluded all patients with complete tumor necrosis on imaging (n=1,059), and repeated mixed-effects regression modeling to identify predictors of understaging. Similar results to those of the primary analysis would argue against significant misclassification of understaging. Finally, given that some “downstaged” patients might be guaranteed to be understaged by the definition used in this study—in particular those who have received LRT to more than 3 lesions—we performed a final sensitivity analysis selectively excluding patients who were downstaged prior to LT.

LRT Subgroup Analysis

In order to investigate the potential impact of different LRT modalities on understaging, we performed a subgroup analysis of patients who received LRT prior to LT. Because some patients receive multiple LRT modalities, we limited this analysis to patients who received only a *single modality* of LRT, classified as transarterial chemoembolization (TACE), ablation, or transarterial radioembolization (TARE). To compare these approaches, we constructed a mixed-effects logistic regression model adjusting for the relevant covariates identified through the primary analysis. We also used multivariable Cox regression to evaluate post-LT outcomes and test for an interaction between understaging and LRT modality.

Variation in Understaging by Transplant Center

To evaluate for differences in rates of understaging by transplant center, we restricted analysis to centers that performed at least 20 transplants for HCC during the follow-up period. We computed unadjusted understaging rates based on multivariable logistic regression with fixed effects, and adjusted rates using the final mixed-effects regression model. We plotted the HCC understaging rates for each center, weighted by transplant volume, and stratified by UNOS region. Unadjusted and adjusted rates of understaging were also computed for UNOS regions and overlaid onto this plot. Finally, we plotted understaging rates in descending order by transplant center, again weighted by transplant volume.

Exploratory Analysis

To evaluate the degree to which sizing errors occurred in staging by Milan criteria, we plotted overlaid histograms of maximum tumor size by imaging and by explant for qualitative analysis. We stratified these plots by correctly staged and understaged patients. To remove possible confounding by understaging on the basis of tumor number, we repeated this analysis restricted only to patients with a solitary HCC tumor on pre-LT imaging.

Results

Cohort Characteristics and Post-LT Outcomes

After applying selection criteria (Supplemental Figure 1), we identified 5,424 patients transplanted with exception points for HCC, 24.9% (n=1,353) of whom were understaged. Median post-LT follow-up was 23.7 months (IQR 12.1 – 35.8). Patients who were understaged were more likely to be male (82.9% versus 75.8%, $p<0.001$), more likely to have been downstaged (4.4% versus 1.6%, $p<0.001$), had higher median pre-LT AFP ($p<0.001$), and more viable tumors pre-LT ($p<0.001$; Table 1). There were also significant differences in understaging by UNOS region ($p<0.001$; Figure 1). For example, 20.4% of patients were understaged in region 2 as compared to 30.8% in region 5. On explant pathology, understaged patients were less likely to have complete tumor necrosis, and more likely to have vascular invasion (each $p<0.001$; Supplemental Table 1). In Kaplan-Meier analysis, understaged patients had a higher hazard of post-LT HCC recurrence (HR 3.31, 95% CI 2.60 – 4.20) and death (HR 2.03, 95% CI 1.72 – 2.41; Figure 2A/B). For example, three year post-LT HCC recurrence and mortality in the correctly staged group were 5.01% and 12.06%, respectively, in comparison to 14.91% and 23.31% in the understaged group (Table 2). These differences persisted in multivariable Cox regression analysis (each understaging $p<0.001$; Figure 2C/D, Supplemental Tables 2 and 3). A sensitivity analysis among patients who did not receive LRT yielded similar results in both unadjusted and adjusted models (Supplemental Figure 2, full data not shown), as did a sensitivity analysis selectively excluding patients who were downstaged (Supplemental Figure 3, full data not shown).

Predictors of Understaging

In multivariable models clustering for UNOS region and transplant center, we found that increasing age (OR 1.13 (per 10 years), 95% CI 1.03 – 1.25), male sex (OR 1.61, 95% CI 1.36 – 1.89), and a higher number of viable tumors at transplant ($p<0.001$) were significantly associated with understaging (Table 3). Patients who received LRT as bridging therapy (OR 1.36, 95% CI 1.10 – 1.67) or for downstaging (OR 4.03, 95% CI 2.65 – 6.11) also had notably increased odds of understaging relative to patients who did not receive LRT. Finally, increasing pre-LT AFP was associated with understaging ($p<0.001$), however this relationship was non-linear. Predicted probabilities of understaging as a function of pre-LT AFP (adjusted and unadjusted) are shown in Figure 3A. Patients who were downstaged had uniformly higher probabilities of understaging across the spectrum of pre-LT AFP, with a possible plateau near ~500ng/mL (Figure 3B). In a sensitivity analyses excluding (1) patients with complete tumor necrosis and (2) patients who were downstaged, the same set of predictors were identified in mixed-effects regression analysis, with similar point estimates to the primary analysis (Supplemental Tables 4 and 5).

LRT Subgroup Analysis

A total 2,326 patients received a single type of LRT prior to LT, where TACE was most frequent (63.5%), followed by ablation (33.3%) and then TARE (2.7%). In univariate analysis, TACE was significantly associated with correct staging, whereas ablation was associated with understaging (Supplemental Table 6, each $p<0.001$). These differences

persisted in multivariable mixed-effects regression, where ablation was associated with a 1.85-fold increased odds for understaging relative to TACE (95% CI 1.46 – 2.33, $p < 0.001$; Supplemental Table 7). In adjusted Cox regression models, the HRs for HCC recurrence and post-LT mortality attributable to understaging were similar between patients who received ablation or TACE (each $p > 0.05$), and understaging remained a strong predictor in each models (HCC recurrence HR 2.23, $p < 0.01$; post-LT mortality HR 1.86, $p < 0.01$; full data not shown).

Variation in Understaging by Transplant Center

Using both unadjusted and adjusted estimates, we found significant differences in transplant center understaging, both between and within UNOS regions (likelihood ratio test $p < 0.001$; Figure 4A). Ordering transplant centers by understaging percentage again revealed significant variation by center (Figure 4B). Higher and lower volume transplant centers were present across the entire range of understaging percentage, from 14.8% to 38.1%.

Exploratory Analysis

When considering maximum tumor sizes between pre-LT imaging and explant for all patients, the general distributions were very similar among those who were correctly staged (Figure 5A). However, when evaluating patients who were understaged, there were significant shifts in the distribution towards increased tumor sizes on explant (Figure 5B), and a relative spike in pre-LT cases with reported 5cm tumors. Similar findings were noted when restricting the analysis cohort to only those patients with a single HCC tumor on pre-LT imaging (Figures 5C/D), with an even more prominent spike in pre-LT tumors measured as 5cm.

Discussion

In this study of 5,424 patients transplanted for HCC in the United States from 2012 to 2016, we found that one quarter of patients transplanted for HCC were understaged, which is consistent with a contemporary study of UNOS data.¹⁷ Interestingly, this is similar to the seminal 1996 Mazzaferro paper, where 27% of patients had discordance between imaging and explant with respect to the Milan criteria.⁴ This suggests that, despite improvements in imaging resolution and modalities, there has been minimal if any improvement in the ability to correctly stage patients with HCC. Understaged patients in our cohort had increased HCC recurrence and higher post-LT mortality, a finding consistent with established literature.^{18, 19} It is therefore critically-important to improve our understanding of factors that predict discordance between findings on imaging and explant. To this end, for the first time, we have identified several novel predictors of understaging. Furthermore, we found significant variation in understaging by transplant center, and argue that some predictors suggest an undue influence of behavioral bias that may ultimately act as a disservice to the broader transplant community.

Few studies have focused on identifying predictors of understaging, and none have done so using national data, as was done here. A single-center study of 118 patients found understaging to be more likely in patients with 2 tumors,¹⁹ a finding which our data

corroborate. A 2006 study by Freeman et al evaluated imaging-related variables associated with radiologic staging accuracy (i.e. understaging or overstaging) in 789 patients transplanted for HCC.¹⁶ Similar to our study, the modality of most recent imaging (CT or MRI) did not impact concordance with explant pathology, and approximately one quarter of patients were understaged. In contrast, however, the authors identified an imaging to LT duration >90 days to be a predictor of inaccurate staging. We attempted to model imaging to LT duration as both a continuous and dichotomous variable (using a 90-day cutoff) but did not find this to be a significant predictor. One possible explanation for this difference is that Freeman et al excluded patients who received LRT prior to LT, whereas the majority of patients in our study received LRT. We were unable to determine the timing of LRT relative to the most recent imaging study. It is therefore possible that some patients received LRT in the interval from imaging to LT, which could mitigate the expected consequences of tumor progression in patients with long intervals from imaging to LT. Finally, Ecker et al evaluated MRI imaging-explant discordance in 318 patients transplanted for HCC, 22% of whom were found to be understaged.¹⁸ The only independent variable associated with understaging in this study was an elevated pre-LT AFP, which was treated as a categorical variable. Our work validates this finding, and offers the advantage of AFP modeled precisely using restricted cubic splines. This allows for visualization of understaging risk based on exact values of pre-LT AFP, which we further stratified by LRT status. In that regard, we found LRT bridging/downstaging status to be a significant predictor of understaging, which has not been previously reported. Furthermore, in a subgroup analysis of patients receiving single-modality LRT, we found that ablation conferred a 1.85-fold increased odds of understaging relative to TACE. Pursuit of LRT may complicate interpretation of subsequent imaging due to tumor necrosis, as suggested by prior studies demonstrating variable interobserver and intermethod LRT tumor response assessments,^{20, 21} and this impact may not be uniform among LRT options. Alternatively, patients who require selected LRT or downstaging may harbor intrinsically more aggressive disease that is more prone to understaging. In sum, our findings may alert clinicians to cases where the risks of understaging (and associated poor post-LT outcomes) are particularly high, and they may also influence the selection of a given LRT modality in a case where TACE and ablation are both reasonable options.

Another major finding in this study is significant variation in rates of understaging by UNOS region and by individual transplant center. Some centers consistently demonstrated understaging rates >35%, while others were <20%. Importantly, these findings were adjusted for regional and center-level clustering which would therefore account for baseline differences in practice variability and case mix. There are several potential explanations for these differences in understaging rates. First, some UNOS regions with high proportions of understaging are also areas with higher competition for donor organs and higher median MELD scores at LT (e.g., regions 1, 5, 6, and 7). Second, there is center-to-center protocol variation in the thickness of pathology sections that could affect the burden of disease identified on explant. This phenomenon has been described previously,^{22, 23} however in the context of this study, we feel it is a less likely explanation of variation because of a clearly-demonstrated increase in HCC recurrence and post-LT death in patients who are understaged. This suggests real differences in staging as opposed to misclassification based on sectioning technique. Third, differences in imaging protocols, radiologic expertise,

reporting conventions, and HCC diagnostic criteria may explain additional aspects of variation not accounted for in our analysis. It is an ongoing issue that the UNOS criteria for qualifying HCC tumors differs from other commonly reported imaging criteria, such as the Liver Imaging and Reporting Data System (LIRADS) or American Association for the Study of the Liver Disease (AASLD) criteria.²⁴ Indeed, prior work has highlighted that the use of UNOS, LIRADS, or AASLD criteria results in significant disagreement in the classification of indeterminate lesions.^{25, 26}

We report suggestive data that behavioral bias may play a role in certain cases of understaging. In assessing the distributions of maximum tumor sizes between imaging and explant, we found significant discordance between imaging and explant among understaged patients. We also found a greater-than-expected number of tumors measured as 5cm on imaging, the upper limit of single-lesion Milan criteria. These findings were persistent when isolating patients with only single tumors, but were curiously absent when limiting the analysis to patients who were correctly staged. Our data are consistent with recent work by Samoylova et al,²⁷ who performed an in-depth analysis of not only single tumors, but also two or three tumors on imaging, where there was an excess of patients with lesions 2.9cm in size (the upper limit of two-three lesion Milan criteria). In our study, it is also interesting to note that male sex was a significant predictor for understaging in multivariable modeling, implying that men were more likely to receive LT despite ultimately being outside of Milan criteria on explant. Although further work is clearly needed to investigate this finding, bias is an important issue to address in the course transplant evaluation, and transplant centers may consider formal implicit bias training. Additionally, periodic UNOS review of concordance between reported Milan adherence by imaging and actual explant data may provide systems improvement by promoting accountability among transplant centers and identifying centers with consistent deficits, thereby providing a basis for remediation. Ultimately, it is likely that optimal HCC patient selection for transplant would be well served by criteria that are subjected to less variability in interpretation, and to less under-appreciation of tumor burden. We have shown that pre-transplant AFP predicts understaging, but a threshold AFP level where the risk of understaging becomes unacceptably high is unavailable. However, there are several biomarkers for HCC currently in development that may prove useful in this context.^{28–30}

There are several study limitations to highlight. First, as with any large retrospective study, there is possible misclassification of certain exposures and outcomes. This issue is likely mitigated by the standardization of UNOS reporting requirements,³¹ in particular with respect to imaging, as well as our use of validated algorithms where applicable. Furthermore, misclassification of tumor-related factors such as number and size would not be expected to be differential with respect to the understaging outcome, and would therefore only bias results towards the null. Second, we were not able to ascertain the exact timing of LRT relative to imaging or transplant. As noted previously, it is therefore possible that the lack of association between prolonged time from imaging to LT could have been modified by administration of LRT. Timing of LRT may also play a role in the observed association between downstaging and understaging, and may be a useful area of future inquiry. Third, we acknowledge that some patients who receive LRT downstaging may be “knowingly” understaged at the time of transplant, and thus the discussion relating to the risk of

downstaging must be appropriately contextualized. However, this issue represents a minority of cases, and exclusion of downstaged patients from modeling did not alter the observed predictors of understaging or associated post-LT outcomes. Finally, in the analysis of variation in understaging by UNOS region, there was insufficient data granularity to study the potential impacts of differing radiology protocols, pathology methods, and transplantation techniques. Similarly, we were unable to explore in depth the associations between specific LRT approaches and the risk of understaging.

In conclusion, we report ongoing high proportions of HCC understaging, and further identify important risk factors for this error. Patients with elevated pre-LT AFP, multiple tumors, and LRT downstaging are at very high risk of being understaged. As such, additional scrutiny is warranted during the evaluation process in these cases. We also found significant regional and center-level heterogeneity in understaging, and present data suggesting that behavioral bias may be an influence, such that borderline patients are categorized as being within Milan criteria in the absence of clear imaging evidence. Prospective studies are needed to further explore these findings, and to propose interventions that may reduce the likelihood of understaging as well as achieve greater uniformity across UNOS regions and among transplant centers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Disclosures: Nadim Mahmud is supported by a National Institutes of Health T32 Research Training Grant (2-T32-DK007740-21A1).

Abbreviations:

AASLD	American Association for the Study of Liver Diseases
AFP	alpha-fetoprotein
BMI	body mass index
CI	confidence interval
CT	computed tomography
HCC	hepatocellular carcinoma
HR	hazard ratio
IQR	interquartile range
LIRADS	Liver Imaging and Reporting Data System
LRT	locoregional therapy
LT	liver transplantation

MELD	model for end-stage liver disease
MRI	magnetic resonance imaging
OR	odds ratio
UNOS	United Network for Organ Sharing

References

1. Lafaro KJ, Demirjian AN, Pawlik TM. Epidemiology of hepatocellular carcinoma. *Surgical Oncology Clinics* 2015;24:1–17. [PubMed: 25444466]
2. Biggins SW, Bambha K. MELD-based liver allocation: who is underserved?, In *Seminars in liver disease*, Copyright© 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY, 2006.
3. OPTN/UNOS Policy Notice Modification to Hepatocellular Carcinoma (HCC) Extension Criteria: Organ Procurement and Transplantation Network, 2018.
4. Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *New England Journal of Medicine* 1996;334:693–700.
5. Jang J, You C, Kim C, et al. Benefit of downsizing hepatocellular carcinoma in a liver transplant population. *Alimentary pharmacology & therapeutics* 2010;31:415–423. [PubMed: 19821808]
6. Yao FY, Mehta N, Flemming J, et al. Downstaging of hepatocellular cancer before liver transplant: long-term outcome compared to tumors within Milan criteria. *Hepatology* 2015;61:1968–1977. [PubMed: 25689978]
7. Schlansky B, Chen Y, Scott DL, et al. Waiting time predicts survival after liver transplantation for hepatocellular carcinoma: A cohort study using the United Network for Organ Sharing registry. *Liver Transplantation* 2014;20:1045–1056. [PubMed: 24838471]
8. Biggins SW. Futility and rationing in liver retransplantation: when and how can we say no? *Journal of hepatology* 2012;56:1404–1411. [PubMed: 22314427]
9. Mahmud N, Shaked A, Olthoff KM, et al. Differences in Posttransplant Hepatocellular Carcinoma Recurrence by Etiology of Liver Disease. *Liver Transplantation* 2019;25:388–398. [PubMed: 30362249]
10. Kashkoush S, El Moghazy W, Kawahara T, et al. Three-dimensional tumor volume and serum alpha-fetoprotein are predictors of hepatocellular carcinoma recurrence after liver transplantation: refined selection criteria. *Clinical transplantation* 2014;28:728–736. [PubMed: 24708263]
11. Serper M, Taddei TH, Mehta R, et al. Association of provider specialty and multidisciplinary care with hepatocellular carcinoma treatment and mortality. *Gastroenterology* 2017;152:1954–1964. [PubMed: 28283421]
12. Aufhauser DD Jr., Sadot E, Murken DR, et al. Incidence of Occult Intrahepatic Metastasis in Hepatocellular Carcinoma Treated With Transplantation Corresponds to Early Recurrence Rates After Partial Hepatectomy. *Ann Surg* 2018;267:922–928. [PubMed: 28085695]
13. Berry K, Ioannou GN. Serum alpha-fetoprotein level independently predicts posttransplant survival in patients with hepatocellular carcinoma. *Liver Transplantation* 2013;19:634–645. [PubMed: 23536495]
14. Mahmud N, John B, Taddei T, et al. Pre-transplant Alpha Fetoprotein is Associated with Post-transplant Hepatocellular Carcinoma Recurrence Mortality. *Clinical transplantation* 2019:e13634. [PubMed: 31177570]
15. Samoylova ML, Dodge JL, Vittinghoff E, et al. Validating posttransplant hepatocellular carcinoma recurrence data in the United Network for Organ Sharing database. *Liver Transplantation* 2013;19:1318–1323. [PubMed: 24039140]
16. Freeman RB, Mithoefer A, Ruthazer R, et al. Optimizing staging for hepatocellular carcinoma before liver transplantation: a retrospective analysis of the UNOS/OPTN database. *Liver transplantation* 2006;12:1504–1511. [PubMed: 16952174]

17. Harper AM, Edwards E, Washburn WK, et al. An early look at the Organ Procurement and Transplantation Network explant pathology form data. *Liver Transplantation* 2016;22:757–764. [PubMed: 26970341]
18. Ecker BL, Hoteit MA, Forde KA, et al. Patterns of discordance between pretransplant imaging stage of hepatocellular carcinoma and posttransplant pathologic stage: a contemporary appraisal of the Milan criteria. *Transplantation* 2018;102:648–655. [PubMed: 29319629]
19. Shah SA, Tan JC, McGilvray ID, et al. Accuracy of staging as a predictor for recurrence after liver transplantation for hepatocellular carcinoma. *Transplantation* 2006;81:1633–1639. [PubMed: 16794527]
20. Riaz A, Miller FH, Kulik LM, et al. Imaging response in the primary index lesion and clinical outcomes following transarterial locoregional therapy for hepatocellular carcinoma. *Jama* 2010;303:1062–1069. [PubMed: 20233824]
21. Forner A, Ayuso C, Varela M, et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? *Cancer* 2009;115:616–623. [PubMed: 19117042]
22. Kim SH, Choi BI, Lee JY, et al. Diagnostic accuracy of multi-/single-detector row CT and contrast-enhanced MRI in the detection of hepatocellular carcinomas meeting the milan criteria before liver transplantation. *Intervirology* 2008;51:52–60. [PubMed: 18544949]
23. Krinsky G Imaging of dysplastic nodules and small hepatocellular carcinomas: experience with explanted livers. *Intervirology* 2004;47:191–198. [PubMed: 15383729]
24. Hussain HK, Barr DC, Wald C. Imaging techniques for the diagnosis of hepatocellular carcinoma and the evaluation of response to treatment, In *Seminars in liver disease*, Thieme Medical Publishers, 2014.
25. Davenport MS, Khalatbari S, Liu PS, et al. Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. *Radiology* 2014;272:132–142. [PubMed: 24555636]
26. Baron RL. The radiologist as interpreter and translator. *Radiology* 2014;272:4–8. [PubMed: 24956044]
27. Samoylova ML, Nigrini MJ, Dodge JL, et al. Biases in the reporting of hepatocellular carcinoma tumor sizes on the liver transplant waiting list. *Hepatology* 2017;66:1144–1150. [PubMed: 28520210]
28. Moshiri F, Salvi A, Gramantieri L, et al. Circulating miR-106b-3p, miR-101-3p and miR-1246 as diagnostic biomarkers of hepatocellular carcinoma. *Oncotarget* 2018;9:15350. [PubMed: 29632649]
29. Qin M, Liu G, Huo X, et al. Hsa_circ_0001649: A circular RNA and potential novel biomarker for hepatocellular carcinoma. *Cancer Biomarkers* 2016;16:161–169. [PubMed: 26600397]
30. Sohn W, Kim J, Kang SH, et al. Serum exosomal microRNAs as novel biomarkers for hepatocellular carcinoma. *Experimental & molecular medicine* 2015;47:e184. [PubMed: 26380927]
31. Wald C, Russo MW, Heimbach JK, et al. New OPTN/UNOS policy for liver transplant allocation: standardization of liver imaging, diagnosis, classification, and reporting of hepatocellular carcinoma: Radiological Society of North America, Inc., 2013.

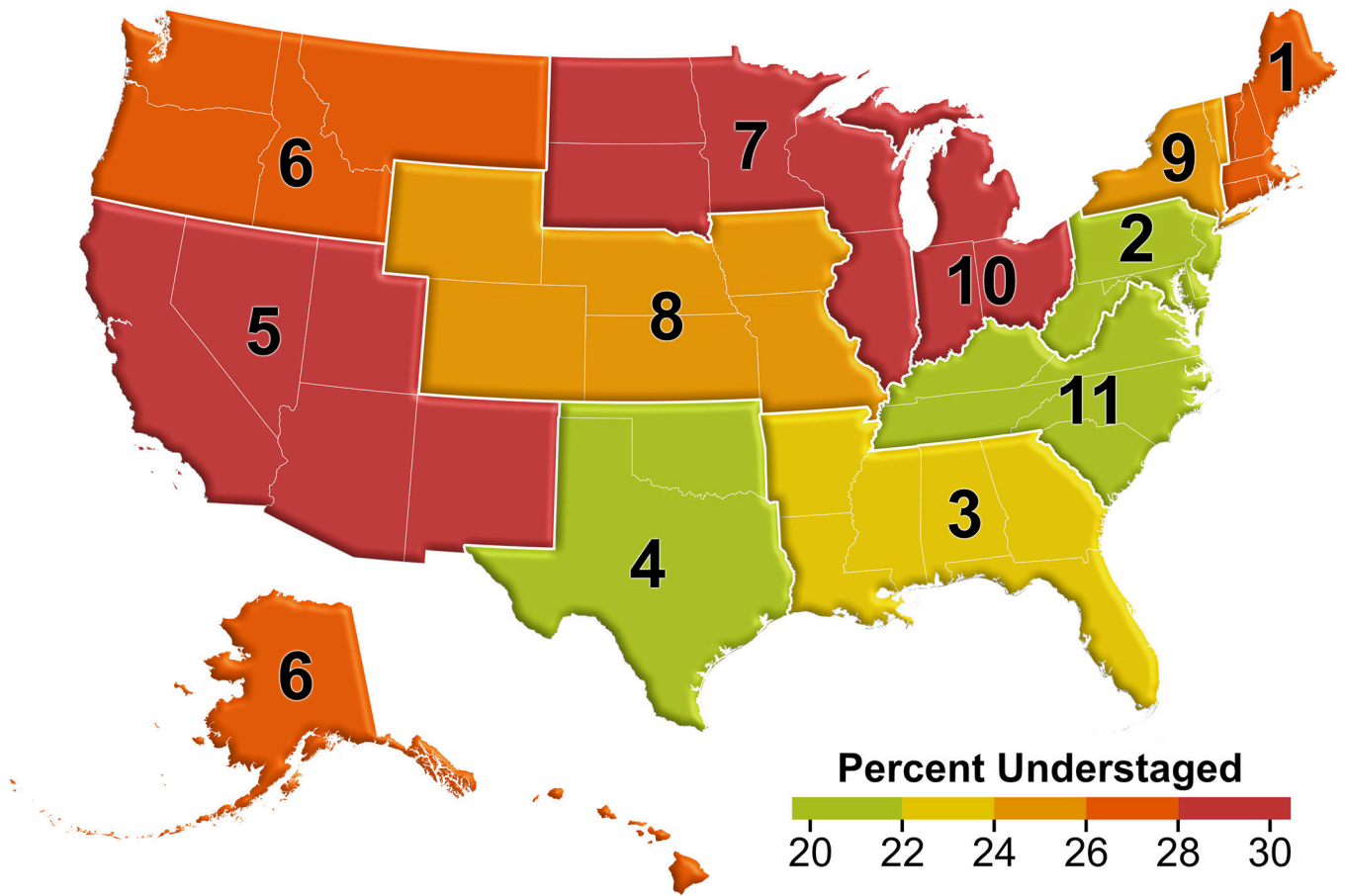


Figure 1 –.
Variation in Proportion of Understaged Patients by UNOS Region

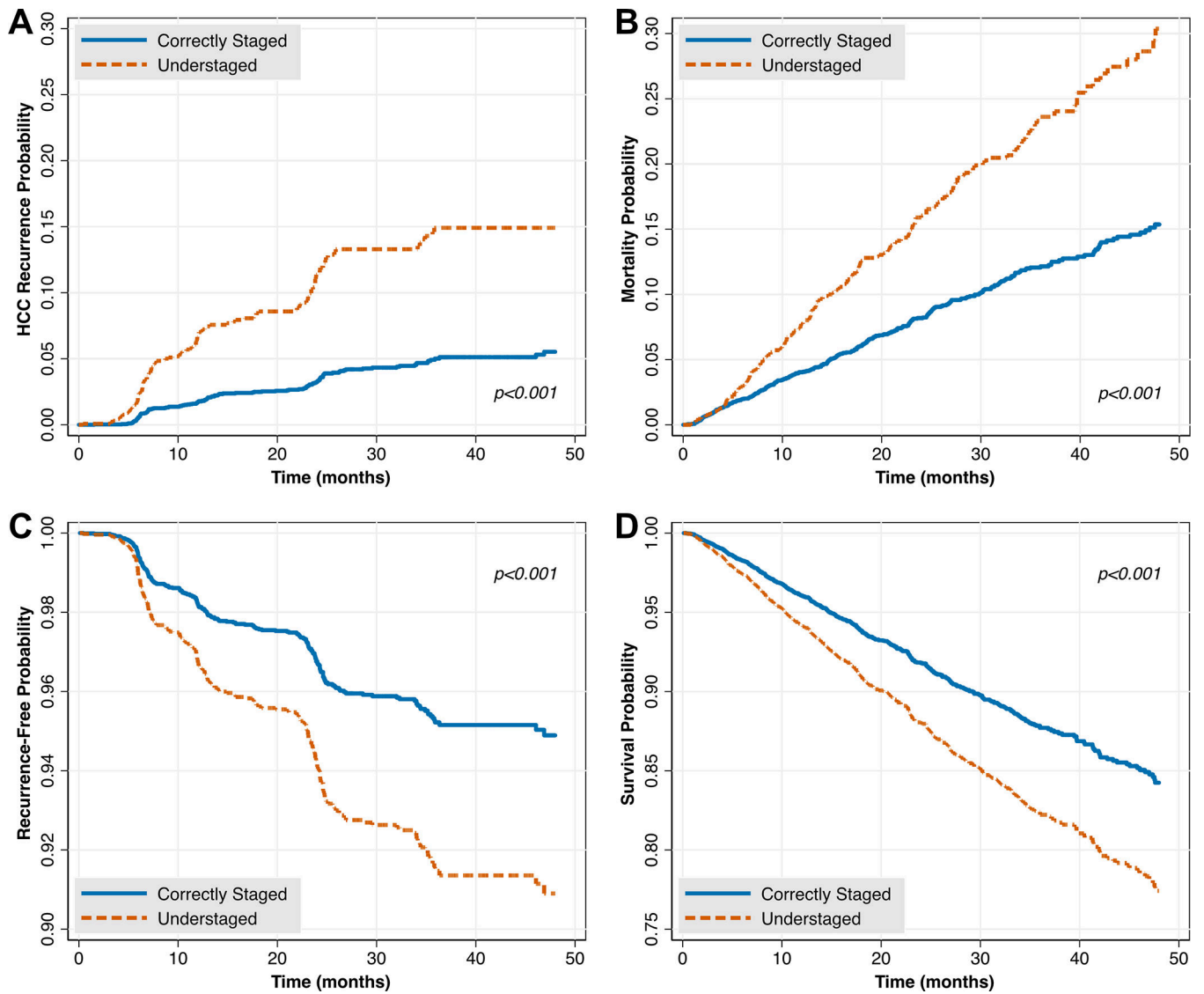


Figure 2 –.
 The Impact of Understaging on Post-Transplant HCC Recurrence and Death in Unadjusted (A and B, respectively) and Adjusted Models (C and D, respectively)

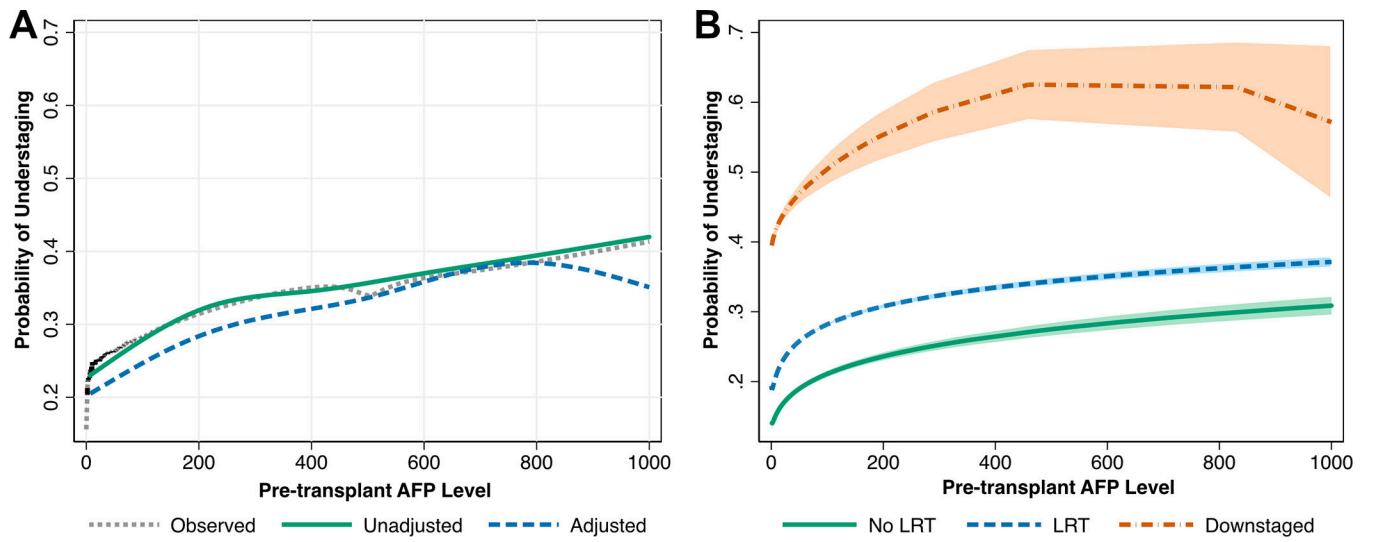


Figure 3 –.
Relationship between Pre-LT AFP and Probability of Understaging, Presented Overall (A)
and Stratified by Receipt of Pre-LT Therapy (B)

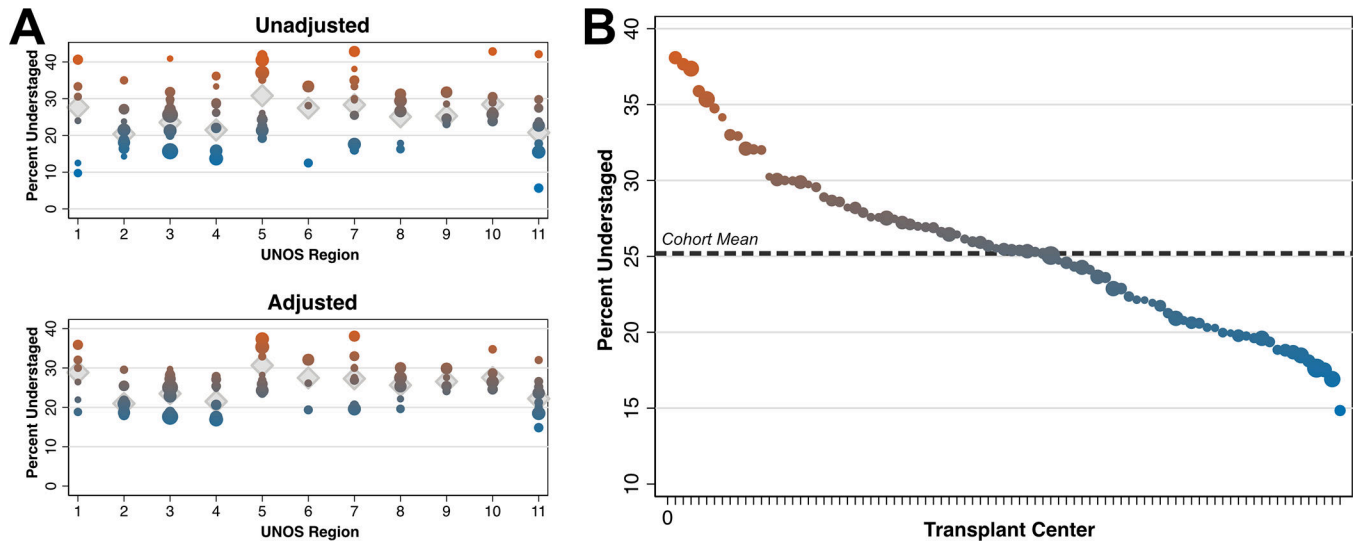


Figure 4 –.
Variation in Percent Understaging by Transplant Center, Stratified by UNOS Region (A) and Sorted by Transplant Center (B)
* Note that each dot represents an individual transplant center, and dot size represents transplant volume. In panel A, diamond symbols indicate unadjusted and adjusted understaging percent by UNOS region.

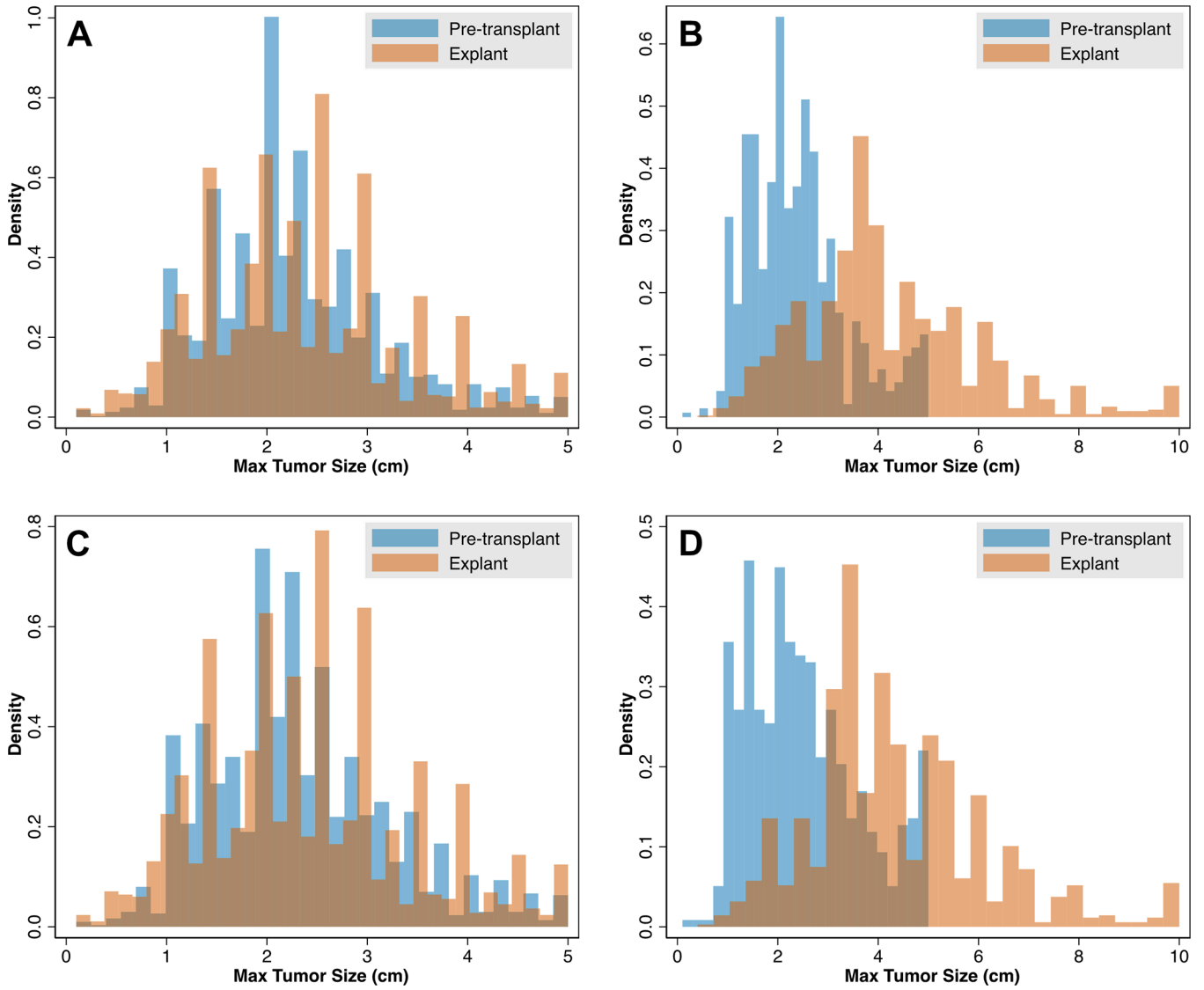


Figure 5 –. Maximum Tumor Sizes on Imaging versus Explant in Correctly Staged (A) and Understaged (B) Patients, and Maximum Tumor Sizes when Limited to Patients with a Single Tumor in Correctly Staged (C) and Understaged Patients (D)

Table 1 -

Characteristics of Correctly versus Understaged Patients

	Correctly Staged (N = 4071)	Understaged (N =1353)	p-value
Age at Transplant, median (IQR)	61 (56, 64)	61 (57, 65)	0.06
Male Sex	3085 (75.8%)	1122 (82.9%)	<0.001
Race			0.90
White	2797 (68.7%)	911 (67.3%)	
Black	408 (10.0%)	137 (10.1%)	
Hispanic	530 (13.0%)	187 (13.8%)	
Asian	280 (6.9%)	98 (7.2%)	
Other	56 (1.4%)	20 (1.5%)	
BMI at Listing, median (IQR)	28.34 (25.15, 31.93)	28.33 (25.22, 31.97)	0.88
Etiology of Liver Disease			0.47
Hepatitis C	2487 (61.1%)	826 (61.0%)	
Hepatitis B	212 (5.2%)	77 (5.7%)	
Alcohol	361 (8.9%)	103 (7.6%)	
Non-alcoholic steatohepatitis	441 (10.8%)	152 (11.2%)	
Autoimmune	104 (2.6%)	27 (2.0%)	
Other	466 (11.4%)	168 (12.4%)	
MELD at Listing, median (IQR)	11.00 (8.00, 15.00)	11.00 (8.00, 15.00)	0.70
MELD at Transplant, median (IQR)	11.00 (8.00, 14.00)	11.00 (8.00, 15.00)	0.72
Pre-transplant AFP, median (IQR)	8.00 (4.00, 20.00)	10.00 (4.00, 31.00)	<0.001
Pre-transplant Therapy			<0.001
None	575 (14.1%)	143 (10.6%)	
LRT Bridging	3432 (84.3%)	1150 (85.0%)	
LRT Downstaging	64 (1.6%)	60 (4.4%)	
UNOS Region			<0.001
1	170 (4.2%)	65 (4.8%)	
2	520 (12.8%)	133 (9.8%)	
3	705 (17.3%)	217 (16.0%)	
4	479 (11.8%)	131 (9.7%)	
5	532 (13.1%)	237 (17.5%)	
6	148 (3.6%)	56 (4.1%)	
7	339 (8.3%)	134 (9.9%)	
8	281 (6.9%)	94 (6.9%)	
9	183 (4.5%)	62 (4.6%)	
10	272 (6.7%)	108 (8.0%)	
11	442 (10.9%)	116 (8.6%)	
Max Pre-LT Tumor Size (cm), median (IQR)	1.60 (0.00, 2.40)	1.50 (0.00, 2.50)	0.65
Number of Viable Tumors at Transplant			<0.001

	Correctly Staged (N = 4071)	Understaged (N =1353)	p-value
1	3485 (85.6%)	1085 (80.2%)	
2	464 (11.4%)	189 (14.0%)	
3	122 (3.0%)	79 (5.8%)	
Most Recent Imaging Prior to LT			0.52
CT	1451 (35.8%)	497 (36.8%)	
MRI	2603 (64.2%)	855 (63.2%)	
Time on Waiting List (months), median (IQR)	6.60 (3.22, 13.31)	6.93 (3.25, 13.24)	0.71
Time from Imaging to LT (days), median (IQR)	73.00 (46.00, 99.00)	75.00 (48.00, 101.00)	0.09
Tumor Location			0.18
Left Lobe	965 (24.9%)	312 (23.1%)	
Right Lobe	2905 (75.1%)	1039 (76.9%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2 –

Post-Transplant Kaplan-Meier Outcomes for Correctly Staged versus Understaged Patients

	Correctly Staged (N = 4071)	Understaged (N =1353)	p-value
HCC Recurrence (%)			<0.001
12 months	1.72 (1.35 – 2.19)	6.82 (5.53 – 8.39)	
24 months	3.23 (2.66 – 3.93)	11.37 (9.50 – 13.57)	
36 months	5.01 (4.18 – 6.00)	14.91 (12.51 – 17.74)	
Death (%)			<0.001
12 months	4.03 (3.46 – 4.70)	7.74 (6.39 – 9.36)	
24 months	8.18 (7.27 – 9.20)	16.04 (13.90 – 18.48)	
36 months	12.06 (10.82 – 13.43)	23.31 (20.41 – 26.56)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3 –Mixed-Effects Logistic Regression Model for Understaging[†]

Variable	Odds Ratio	95% CI	p-value
Age (per 10 years)	1.13	(1.03 – 1.25)	0.01 *
Male Sex	1.61	(1.36 – 1.89)	<0.001 *
Pre-transplant Therapy			
None	(ref)	(ref)	
LRT Bridging	1.36	(1.10 – 1.67)	<0.01 *
LRT Downstaging	4.03	(2.65 – 6.11)	<0.001 *
RCS(AFP)[‡]			<0.001 *
Number of Viable Tumors at Transplant			
1	(ref)	(ref)	
2	1.40	(1.15 – 1.70)	<0.01 *
3	2.13	(1.57 – 2.89)	<0.001 *

* Significant at the alpha=0.05 level

[†] Transplant center and UNOS region designated as random intercepts

[‡] AFP modeled as a 3-knot restricted cubic spline function (knots at 3, 8, and 83)

[§] The following variables were not significant in univariate analysis: race/ethnicity, body mass index, etiology of liver disease, MELD score, maximum tumor size on imaging, pre-transplant imaging modality, time on waiting list, time from imaging to transplant