Research article

JHEP|Reports

Impact of waiting time on post-transplant survival for recipients with hepatocellular carcinoma: A natural experiment randomized by blood group

Berend R Beumer,[1] Wojciech G Polak,[1] Robert A de Man,[2] Herold J Metselaar,[2] David van Klaveren,[3] Jeremy Labrecque,[4] Jan NM IJzermans[1,*]

[1]*Erasmus MC Transplant Institute, Department of Surgery, Division of HPB & Transplant Surgery, University Medical Centre Rotterdam, Rotterdam, the Netherlands;* [2]*Erasmus MC Transplant Institute, Department of Gastroenterology and Hepatology, University Medical Centre Rotterdam, Rotterdam, the Netherlands;* [3]*Department of Public Health, Erasmus MC - University Medical Centre, Rotterdam, the Netherlands;* [4]*Department of Epidemiology, Erasmus MC - University Medical Centre, Rotterdam, the Netherlands*

JHEP Reports **2023.** https://doi.org/10.1016/j.jhepr.2022.100629

Background & Aims: When listing for liver transplantation, one can transplant as soon as possible or introduce a test-of-time to better select patients, as the tumor's biological behavior is observed. Knowing the degree of harm caused by time itself is essential to advise patients and decide on the maximum duration of the test-of-time. Therefore, we investigated the causal effect of waiting time on post-transplant survival for patients with hepatocellular carcinoma.
Methods: We analyzed the UNOS-OPTN dataset and exploited a natural experiment created by blood groups. Relations between variables and assumptions were described in a causal graph. Selection bias was addressed by inverse probability weighting. Confounding was avoided using instrumental variable analysis, with an additive hazards model in the second stage. The causal effect was evaluated by estimating the difference in 5-year overall survival if all patients waited 2 months instead of 12 months. Upper bounds of the test-of-time were evaluated for probable scenarios by means of simulation.
Results: The F-statistic of the first stage was 86.3. The effect of waiting 12 months *vs.* 2 months corresponded with a drop in overall survival rate of 5.07% (95% CI 0.277–9.69) and 8.33% (95% CI 0.47–15.60) at 5- and 10-years post-transplant, respectively. The median survival dropped by 3.41 years from 16.21 years (95% CI 15.98–16.60) for those waiting 2 months to 12.80 years (95% CI 10.72–15.90) for those waiting 12 months.
Conclusions: From a patient's perspective, the choice between ablate-and-wait *vs.* immediate transplantation is in favor of immediate transplantation. From a policy perspective, the extra waiting time can be used to increase the utility of scarce donor livers. However, the duration of the test-of-time is bounded, and it likely should not exceed 8 months.
Impact and implications: When listing patients with liver cancer for transplantation, it is unclear whether a test-of-time or immediate transplantation offer better outcomes at the population level. In this study, we found that increased liver transplant waiting times are detrimental in patients with liver cancer. Furthermore, our simulation showed that a pre-operative observational period can be useful to ensure good donor liver allocation, but that its duration should not exceed 8 months.

Introduction

Patients with hepatocellular carcinoma (HCC) that are eligible for liver transplantation (LT) are placed on a waiting list as a donor liver is not instantly available. How this waiting list should be organized is a continuous topic of fierce debate. In the past two decades high quality research was aimed at answering major questions like: which subset of patients should be placed on the waiting list? What is the best way to treat patients prior to transplantation? And how do we prioritize high-risk patients to reduce waiting list mortality? Continued research on these major topics will keep advancing the discussion on how to best organize clinical care. Although there is a growing consensus about these major topics, a question fundamental to organizing the waiting list remains controversial. Namely, how much does the waiting cost a patient with HCC in terms of post-transplant survival[1,2]?

Knowing the cost of waiting is essential to determine the maximum size of the waiting list, or to confidently expand it. Additionally, the use of the so-called test-of-time is increasingly promoted.[3,4] The rationale of this policy is that in the perioperative observation period patients with the most aggressive forms of cancer are filtered out, hereby improving the allocation of scarce donor livers, and increasing the average post-transplant survival. However, the merit of this allocation policy depends on whether the possible harm of the additional

Keywords: Hepatocellular carcinoma; Liver transplantation; Waiting time; Instrumental variable analysis; Mendelian randomization; Causal inference; Quasi randomized controlled trial.
Received 1 September 2022; received in revised form 27 October 2022; accepted 27 October 2022; available online 22 November 2022
* Corresponding author. Address: Erasmus MC Transplant Institute, Dr. Molewaterplein 40, 3015 GD Rotterdam, Department of Surgery RG-219; Tel.: 010-7032396.
E-mail address: j.ijzermans@erasmusmc.nl (J.N.M. IJzermans)

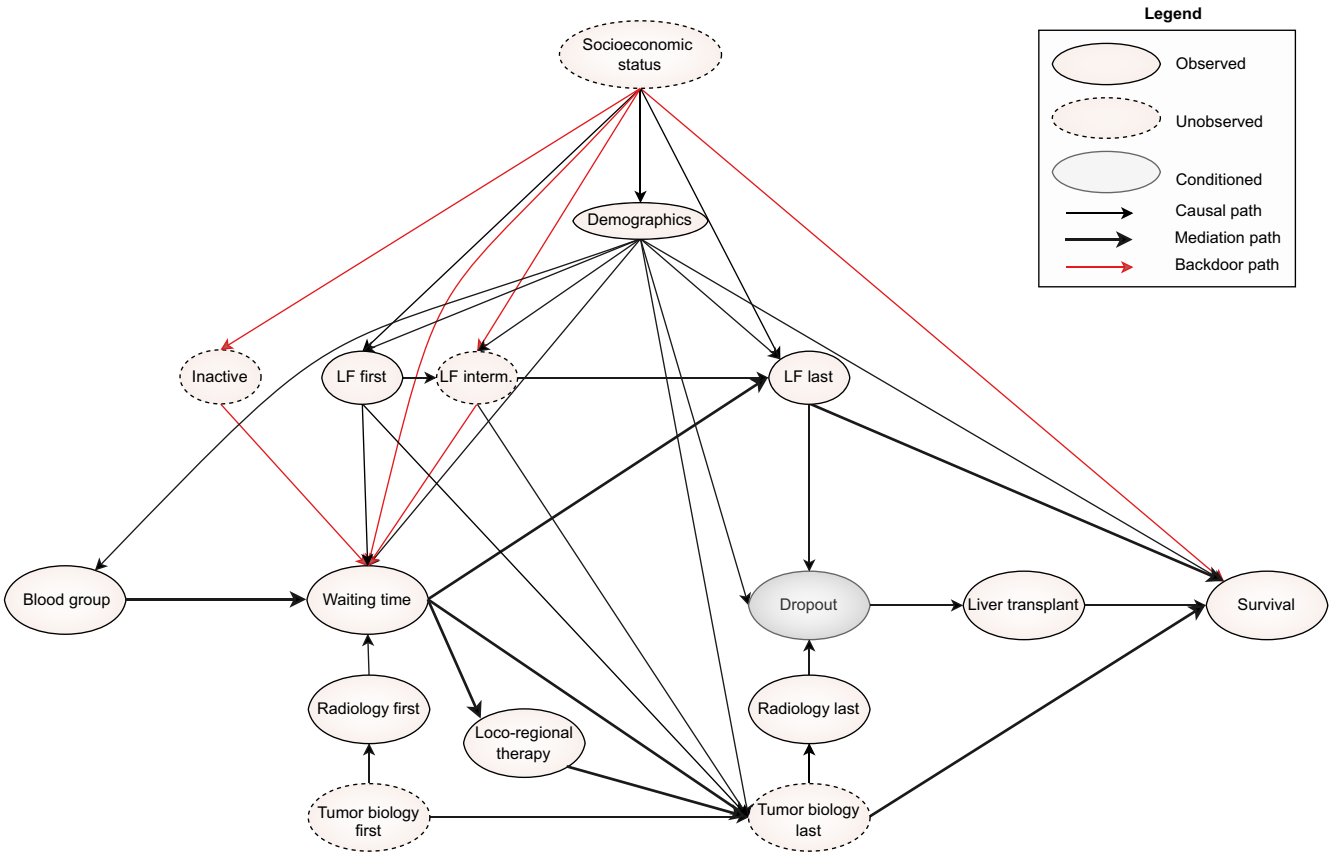

Fig. 1. The causal graph that guides the analysis to identify the harm of waiting time. LF, liver function.

waiting time endured by the full population can be offset with the better allocation of a fraction of the livers. Lastly, knowledge on the cost of waiting can aid transplant clinicians to recommend treatment for patients with multiple options. Patients with more treatment options are often diagnosed in an earlier stage with good liver function leading to a low ranking in the waiting list. Knowing the cost of waiting helps clinicians to identify a threshold at which resection or ablation is the better strategy compared to bridging therapy and eventually transplantation.

Establishing the causal effect of waiting time is, however, complicated by the prioritization schemes. To reduce waiting list mortality, the waiting list is currently organized such that the patients with the worst liver function get transplanted first.[5] Therefore, a naive survival comparison of patients transplanted within 6 months *vs.* those transplanted after more than 6 months, or a variation thereof, is guaranteed to result in downward biased estimates. In extreme cases, it would appear that waiting is beneficial. A further complicating feature is that socioeconomic mechanisms play an important role, crippling almost all empirical studies. Namely, it is known that patients with lower socioeconomic status are more often inactive on the waiting list due to, for example, problems with their insurance, incomplete screening, or medical non-compliance.[6] Additionally, lower frequency of check-up visits leads to a slower escalation of their priority in the waiting list if their liver function decreases. Even though a randomized controlled trial is the gold standard to avoid these biases, in this case it is less preferred due to ethical considerations and the time and expenditures involved. Studying natural experiments in large national datasets may offer a valid alternative to a prospective randomized controlled trial, avoiding confounding altogether.

In this case, blood groups create a natural randomized controlled trial with treatment arms that are equal in composition of (unobserved) confounders but differ with regard to waiting time. Key is that blood groups are determined at conception and follow the fundamental laws of inheritance.[7] Therefore, they can be expected to be independent of confounders. In other words, patients or doctors cannot choose the blood group. Additionally, the blood group itself does not directly influence the severity of the HCC, liver function, complications, or survival. The blood groups, however, do directly influence a patient's time to transplant. This is caused by the fact that patients with blood group AB can in case of emergency accept a donor organ from blood group A, B, AB, or O. While patients with blood group O can only receive a donor organ from blood group O. Therefore, the blood group could be used as an instrument for waiting time in an instrumental variable analysis. Hence, our research aim is to analyze the natural randomized controlled trial created by blood groups and estimate the causal effect of waiting time on postoperative overall survival for patients with HCC undergoing transplantation.

Patients and methods

Data

The reporting of this retrospective observational cohort study adheres to the STROBE guidelines (Table S1).[8] The protocol for this study was approved by the Medical Ethics Committee of Erasmus MC, Erasmus University Medical Centre, Rotterdam, the Netherlands (MEC-2020-0779), and adheres to the declaration of Helsinki. The data was supplied by the United Network for Organ Sharing (UNOS) of the Organ Procurement and Transplantation Network (OPTN) as of December 2, 2021. This study analyzed patients with HCC listed for LT in the period between 2000-2019. The records of patients were excluded if: the blood group or follow-up data was missing; patients were younger than 18 at the time of listing; the record did not describe the first transplantation; if the patient received a multi-organ transplantation; or if the patient received a living donor liver transplantation. Furthermore, the information from multiple listings (*e.g.*, due to hospital transfers) was aggregated for the relevant variables. Most importantly, waiting time was recalculated as the time difference between the date of first listing, out of all listings, and the date of transplantation.

Outcome measures

The primary outcome measure was overall survival, defined as the time in days between the date of LT and the date of death or last follow-up. The primary aim was to estimate the average treatment effect (ATE) of waiting, defined as the change in hazard rate due to increases in waiting time by one day. To aid interpretation, the secondary aim was to estimate the difference in 5-year overall survival between if all patients waiting 2 months instead of 12 months, which approximately corresponds to the first and last quartile of the waiting time distribution. In addition, we investigated if tumor number, tumor size, alpha-fetoprotein (AFP), or the albumin-bilirubin (ALBI) score moderated the treatment effect. Lastly, for the test-of-time simulation we will evaluate the effect of waiting by taking the difference of the average lifetimes, defined as the area under the survival curve.[9]

Causal graph

Relations between variables were described in a causal graph and were based on expert opinion (Fig. 1). Variables that cover similar information and that share similar arrangements are grouped for clarity. An introduction to causal graphs can be found in Supplement 1. In addition, the rationale and a more detailed version of our causal graph are provided in Fig. S1. The causal graph shows that the relationship between waiting time and survival is biased by multiple mechanisms. More specifically, the dropout from the waiting list induces selection bias, as only those without the most aggressive sub-types and good liver function make it to the liver transplantation. In the analysis of post-transplant survival, by definition, only patients that received the transplantation are analyzed. Hereby we necessarily condition on the variable dropout depicted in the graph by shading in gray. Furthermore, we can observe that socioeconomic status is a common cause of waiting time and survival. The variables inactivity, intermediate liver function and socioeconomic status are not observed. Therefore, these paths cannot be blocked by means of conditioning and confounding remains. In the causal graph these backdoor paths are highlighted in red. Further, it should be noted that pre-operative loco-regional therapy (LRT) acts as a mediator, potentially modifying the effect of waiting time. Rather than focusing on the isolated effect of waiting time, this research will evaluate the total effect of waiting time. Hereby the effect of the LRT is absorbed into our measurement. Assuming pre-operative LRT is not harming patients, this will result in a conservative estimate of the isolated effect of waiting.

Statistical analysis

For the descriptive statistics, discrete data was represented in absolute numbers and percentages. Continuous data was represented using the first, second and third quartiles. Covariate balance over the blood groups was assessed using the standardized mean difference and the Kolmogorov-Smirnov statistic with thresholds 0.25 and 0.1, respectively.[10] Missing values were addressed using multiple imputation. As the fraction of missing data was 10%, the missing values were imputed 20 times with each imputation receiving 20 iterations.[11] Estimates from the repeated analysis were pooled using the Rubin Rules.[12,13]

Selection bias resulting from conditioning on patients that received a transplantation was addressed using inverse probability weighting (IPW).[14] The probability that a patient drops-out is estimated using a logistic regression model that included the variables: sex, age, BMI, functional status, life support, level of education, ethnicity, insurance type, transplant region, ALBI score (last), model for end-stage liver disease score (last), encephalopathy, ascites, cirrhosis, tumor number (last), tumor size (last), total tumor size (last), $\log_{10}$(AFP) (last).

Confounding bias was avoided using instrumental variable analysis, which is explained in Supplement 2.[15,16] The four essential instrumental variable assumptions are credible for the following reasons. First, the independence assumption was unlikely to be violated by the fact that the blood group of the child is exclusively determined by random selection of the parental alleles. Important to note, however, is that although the allele selection is random, there are slight differences in the distribution of A, B, and O alleles across transplant regions and ethnic groups. To ensure that no open backdoor paths exist the variables transplant region and ethnicity are absorbed into the conditioning set. Secondly, the relevance assumption was supported by the asymmetry in the compatible blood groups between donors and recipients. For example, a recipient with blood group AB can potentially receive a donor liver from someone with blood group AB, A, B, or O, while a recipient with blood group O can only receive a donor liver from a donor with blood group O. Hence, even if a recipient with blood group O is the highest on the waiting list, he or she might need to wait longer until a suitable organ is available in comparison to a recipient with blood group AB. Empirically, the relevance assumption will be assessed using the F-statistic of the first stage which corresponds to the strength of the association between the blood-group and waiting time. By rule of thumb the F-statistic should be above 10 for the first stage to have sufficient strength.[17] Thirdly, the exclusion assumption was justified as, to the best of our knowledge, no studies exist which describe a direct or indirect oncogenic interaction between the blood group antigen and the hepatocytes. In extension, there is no evidence that the blood group is related to HCC recurrence, metastasis, or mortality in any other way. In addition, after appropriate matching of donor and recipients, the blood group is not correlated to primary graft dysfunction or rejection. An overview of the addressed and avoided backdoor paths is shown in Fig. S1. Lastly,

Table 1. Descriptive characteristics stratified by blood group.

	AB	B	A	O	SMD (max)	KS (max)
n	**1,797**	**5,877**	**16,818**	**21,202**		
Period, n (%)						
[2000,2005]	268 (15)	966 (16)	2,642 (16)	3,391 (16)	0.042	0.015
(2005,2010]	458 (25)	1,435 (24)	4,119 (24)	5,258 (25)	0.025	0.011
(2010,2015]	596 (33)	1,926 (33)	5,485 (33)	6,872 (32)	0.016	0.008
(2015,2020]	475 (26)	1,550 (26)	4,572 (27)	5,681 (27)	0.018	0.008
Male						
Missing (%)	0 (0)	0 (0)	0 (0)	0 (0)	–	–
n (%)	1,404 (78)	4,513 (77)	13,061 (78)	16,165 (76)	0.045	0.019
Age						
Missing (%)	0 (0)	0 (0)	0 (0)	0 (0)	–	–
Q1 | Q2 | Q3	55 | 60 | 64	54 | 59 | 64	54 | 59 | 64	54 | 59 | 64	0.066	0.037
Height (m)						
Missing (%)	3 (0)	29 (0)	59 (0)	58 (0)	0.058	0.003
Q1 | Q2 | Q3	165 | 173 | 180	165 | 173 | 178	167 | 173 | 180	165 | 173 | 178	0.109	0.046
Weight (kg)						
Missing (%)	3 (0)	30 (1)	64 (0)	63 (0)	0.059	0.003
Q1 | Q2 | Q3	72 | 84 | 97	70 | 82 | 96	73 | 85 | 98	73 | 84 | 97	0.148	0.025
Ethnicity, n (%)						
Missing (%)	0 (0)	0 (0)	0 (0)	0 (0)	–	–
White	1,163 (65)	3,157 (54)	12,105 (72)	12,908 (61)	0.381	0.183
Black	175 (10)	840 (14)	1,043 (6)	2,149 (10)	0.270	0.081
Hispanic	146 (8)	772 (13)	2,398 (14)	4,359 (21)	0.361	0.124
Asian	292 (16)	1,045 (18)	1,056 (6)	1,433 (7)	0.362	0.125
Native American	1 (0)	15 (0)	90 (1)	182 (1)	0.123	0.008
Pacific Islander	10 (1)	15 (0)	46 (0)	45 (0)	0.061	0.003
Multiracial	10 (1)	33 (1)	80 (0)	126 (1)	0.016	0.001
MELD score						
Missing (%)	18 (1)	47 (1)	188 (1)	198 (1)	0.033	0.003
Q1 | Q2 | Q3	8 | 11 | 15	8 | 11 | 15	8 | 11 | 15	8 | 11 | 15	0.035	0.025
ALBI score						
Missing (%)	13 (0.7)	29 (0.5)	115 (0.7)	119 (0.6)	0.029	0.002
Q1 | Q2 | Q3	-2.4 | -1.9 | -1.4	-2.4 | -1.9 | -1.3	-2.3 | -1.8 | -1.3	-2.3 | -1.8 | -1.3	0.106	0.032
Tumor #, n (%)						
Missing	438 (24)	1,105 (19)	3,126 (19)	3,868 (18)	0.154	0.061
1	1,032 (57)	3,622 (62)	10,337 (61)	13,196 (62)	0.015	0.018
2	234 (13)	849 (14)	2,449 (15)	3,053 (14)		
3	90 (5)	282 (5)	856 (5)	1,024 (5)		
4	0 (0)	13 (0)	39 (0)	45 (0)		
≥5	3 (0)	6 (0)	11 (0)	16 (0)		
Tumor size (cm)						
Missing (%)	438 (24)	1,105 (19)	3,126 (19)	3,868 (18)	0.154	0.061
Q1 | Q2 | Q3	2 | 2 | 3	2 | 2 | 3	2 | 2 | 3	2 | 2 | 3	0.022	0.028
Log_{10}(AFP) (ng/ml)						
Missing (%)	546 (30.4)	1,473 (25.1)	4,220 (25.1)	5,318 (25.1)	0.121	0.053
Q1 | Q2 | Q3	0.7 | 1 | 1.6	0.7 | 1 | 1.6	0.7 | 1 | 1.6	0.7 | 1 | 1.6	0.035	0.024
LRT						
Missing	465 (26)	1,388 (24)	3,905 (23)	5,109 (24)	0.062	0.027
n (%)	916 (69)	3,278 (73)	9,496 (74)	11,849 (74)	0.109	0.049
Waiting time (months)						
Missing (%)	0 (0)	0 (0)	0 (0)	0 (0)	–	–
Q1 | Q2 | Q3	1 | 3 | 8	2 | 7 | 13	3 | 8 | 15	3 | 8 | 16	0.329	0.252
Dropout						
Missing	22 (1)	126 (2)	438 (3)	594 (3)	0.108	0.016
n (%)	304 (17)	1,428 (25)	4,470 (27)	6,159 (30)	0.297	0.128
n	**1,465**	**4,306**	**11,857**	**14,372**		
Median FU [95%CI] (years)	6 [5.8 - 6.3]	5.9 [5.8 - 6]	5.9 [5.9 - 6]	5.9 [5.8 - 6]		
Death						
Missing (%)	0 (0)	0 (0)	0 (0)	0 (0)		
n (%)	402 (27)	1,205 (28)	3,512 (30)	4,317 (30)		
Median OS [95%CI] (years)	14.5 [13.5–n.a.]	14.3 [12.9–15.9]	12.7 [12.3–13.3]	12.8 [12.3–13.3]		
5-year OS [95%CI]	0.75 [0.73–0.78]	0.75 [0.74–0.77]	0.74 [0.73–0.75]	0.74 [0.73–0.74]		

The main characteristics stratified by blood group. Tumor number, tumor size, MELD score, ALBI score, and AFP represent the measurement at listing.
ALBI score, albumin-bilirubin; AFP, alpha-fetoprotein; FU, follow-up; KS, Kolmogorov-Smirnov; MELD, model for end-stage liver disease; OS, overall survival; SMD, standardized mean difference.

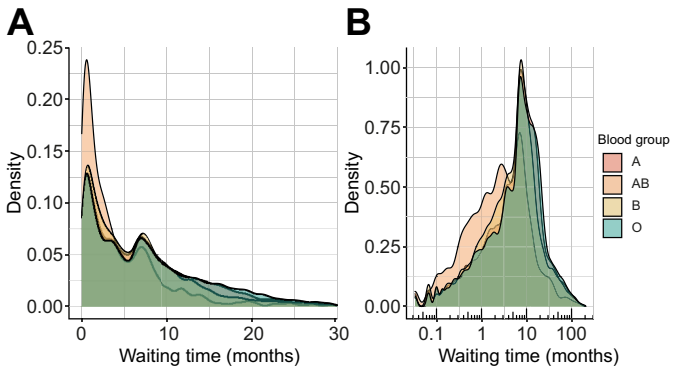

Fig. 2. Waiting time distribution. The distribution of the waiting time on the linear scale (A) and on the logarithmic scale (B).

the monotonicity assumption is warranted. Although the waiting times vary between individuals, the waiting time would not increase if a patient were, hypothetically, assigned a more favorable blood group.[18] The validity of this assumption will be empirically assessed by plotting the cumulative density function of the waiting time stratified by blood group.

To estimate the treatment effect of waiting we used two-stage predictor substitution.[19] The first stage consisted of a linear regression model in which waiting time was regressed on the blood group, transplant region, and ethnicity. In the second stage an additive hazards model was used to regress the time-to-event information on the predicted value of the first stage, transplant region, and ethnicity.[20] Standard errors were obtained using bootstrap resampling. Heterogeneity in the treatment effect was investigated on a variable-by-variable basis for which the population was subsetted based on the quartiles. For the lowest and highest subset, the instrumental variable analysis was repeated. More formally the instrumental variable analysis was extended with the addition of interaction terms to investigate if the ATE was moderated by tumor number, tumor size, AFP, or the ALBI score, which were measured prior to listing.

Lastly, we performed a *post hoc* simulation to investigate the conditions under which the test-of-time results in a survival benefit given the causal effect of waiting time on post-transplant survival. For this, following the rationale of the test-of-time,[3,21,22] the population of patients was viewed as a mixture of patients

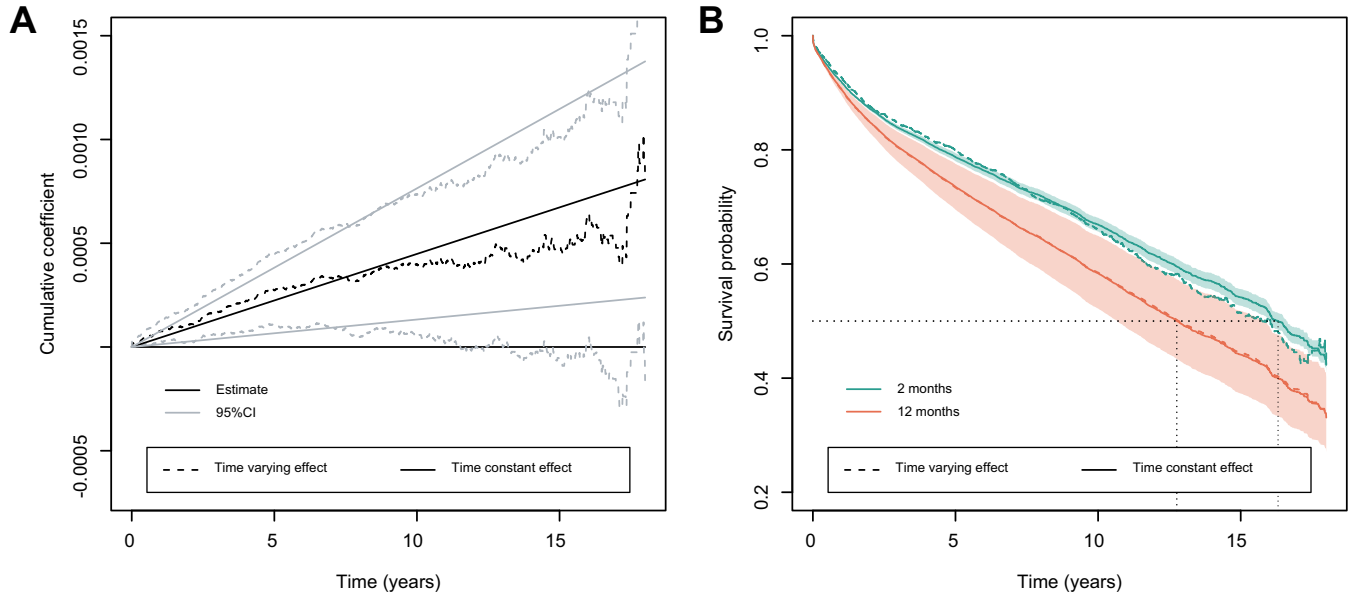

Fig. 3. Average treatment effect of waiting time. The average treatment effect of waiting time on post-transplant survival. (A) The time constant and time varying cumulative hazard function for waiting time. (B) The cumulative coefficient for waiting time is translated to the survival scale and shows a contrast between if all patients wait 2 months *vs.* 12 months.

Table 2. Heterogenous treatment comparison of quantiles.

	Q1				Q4			
	Level	**F**	**$\gamma^{*}10^{-4}$ [95%CI]**	**Δ5 year OS %**	**Level**	**F**	**$\gamma^{*}10^{-4}$ [95%CI]**	**Δ5 year OS %**
Tumor number	1	63	0.17 [-0.22; 0.55]	1.87	2	24	1.03 [0.10; 1.97]	11.8
Tumor size (cm)	1.8	32	0.35 [-0.34; 1.04]	4.02	2.9	24	0.54 [-0.19; 1.27]	6.0
Log_{10}(AFP)	0.7	24	0.28 [-0.32; 0.88]	3.38	1.59	25	0.62 [-0.19; 1.44]	6.4
ALBI score	-2.37	31	0.39 [-0.16; 0.94]	4.7	-1.33	18	0.43 [-0.37; 1.24]	4.6

The harm of waiting for each of the variables within the lowest and the highest subgroup based on the quartiles.
ALBI score, albumin-bilirubin; AFP, alpha-fetoprotein; OS, overall survival.

Table 3. Heterogenous treatment effect regression.

	$\gamma * 10^{-4}$	[95%CI]	*p* value
W	0.38	[0.06; 0.82]	0.004
W*tumor number	0.48	[-0.27; 1.20]	0.115
Tumor number	70	[26; 114]	<0.001
W*tumor size	0.08	[-0.48; 0.64]	0.938
Tumor size	0.12	[-0.55; 0.79]	<0.001
W*$\log_{10}$(AFP)	0.12	[-0.55; 0.79]	0.908
Log_{10}AFP	120	[82; 160]	<0.001
W*ALBI	0.11	[-0.65; 0.87]	0.936
ALBI	81	[20; 140]	<0.001

The regression with interaction terms to study the heterogenous effect of waiting. *p* values were obtained using the Wald test. Significance testing was performed using the Wald test.
ALBI score, albumin-bilirubin; AFP, alpha-fetoprotein; W, waiting time.

with aggressive, and non-aggressive HCC. The derivation of the upper bound is presented in Supplement 3. Only the cost of waiting was studied and quantified in this research. For the other components we used credible ranges to calculate the upper bound. More specifically, for the difference in average lifetime between patients with aggressive and non-aggressive HCC we used a range of 7 to 16 years; for the proportion of aggressive cases without the test-of-time we used a range of 2 to 30%; and lastly, we used a range of 10 to 70% for the reduction of aggressive cases with the test-of-time.

Results

Of the 273,134 patients in the UNOS dataset, 45,694 patients were included in the study. All included patients were used in the IPW model, and 32,000 patients that had undergone transplantation entered the instrumental variable analysis (Fig. S2). Of the included patients, the median waiting time was 7.4 months with an IQR of 12 months. The descriptive statistics, stratified by blood group, showed no major imbalance in important confounders across the blood groups (Table 1). Blood groups were not uniformly distributed across different ethnicities. Furthermore, differences across the blood groups exist in terms of waiting time, dropout rate, and post-transplant survival. A more detailed overview of the covariate balance statistics for each blood group comparison is provided in Fig. S3.

In the first stage, the mean waiting time at each of the levels of the instrument was significantly different with respect to the reference category (AB). Furthermore, the F-statistic of 86.3 indicated that the blood group as an instrument is sufficiently associated with waiting time to be used in instrumental variable analysis. The underlying source of variation was confirmed to be the asymmetric blood group compatibility (Table S2). This results in an overrepresentation of patients with AB and B in the first quartile, while more patients with blood groups O and A are left in the last quartile (Fig. 2). The empirical cumulative distribution functions did not show any evidence that the monotonicity assumption was violated, as the lines overlap at the extremes but do not cross (Fig. S4).

The instrumental variable analysis indicates that waiting longer reduces post-transplant survival (Fig. 3). The time varying estimates were increasing, and the confidence bounds did not include the zero line. The time constant effect, $\gamma^{*}10^{-4}$, was estimated to be 0.44 (95% CI 0.12–0.76; p <0.001). On the survival scale, the effect of waiting 12 months *vs.* 2 months corresponded with a drop in overall survival of 5.07% (95% CI 0.277–9.69) at 5 years post transplantation and 8.33% (95% CI 0.47–15.60) at 10 years post transplantation. The median survival dropped by 3.41 years from 16.21 years (95% CI 15.98–16.60) for those waiting 2 months to 12.80 years (95% CI 10.72–15.90) for those waiting 12 months and the difference in mean lifetime was 1.36 years (95% CI 0.26–2.34).

Table 2 shows the descriptive statistics regarding the heterogeneity of the treatment effect. All estimates indicate a higher impact of waiting on patients that were in the highest quartile (*i.e.*, sicker patients) in comparison to patients that were in the lowest quartile (*i.e.*, less sick patients). However, due to the reduction in sample size, the confidence intervals of the ATE in the subgroups increased. The estimate for waiting time in the Q4 subgroup of tumor number is the only subset in which the subgroup ATE is significantly different from zero, with $\gamma^{*}10^{-4}$ of 1.03 (95% CI 0.10–1.79). The heterogeneous treatment effect was more formally investigated using interaction terms shown in Table 3. These are all positive, in concordance with the observation that sicker patients are harmed more by waiting. However, none of the interaction terms attained statistical significance. The tests therefore remain inconclusive with respect to whether the harm of waiting is equal for all, or if some patients are more affected.

The results of the simulation investigating the conditions under which test-of-time results in a survival benefit are presented in Table S3 and S4. The simulation shows that given a proportion of aggressive cases of 12% and a reduction of 50% with the test-of-time, the upper bound of the waiting time is 3 to 8 months, with the point estimate depending on the difference in average survival between the aggressive and non-aggressive cases. Beyond this upper bound, the harm of waiting is no longer compensated for by the improved allocation of donor livers.

Discussion

The aim of our research was to estimate the causal effect of waiting time on post-transplant survival. We conclude that waiting is harmful for patients undergoing transplantation for HCC, with the overall survival rate at 5 years dropping by an estimated 5% and median survival dropping by 3.41 years if all patients wait 12 instead of 2 months. We reason that the impaired survival could be biologically explained by the increased time at risk of micro metastatic spread. This is in line with our subgroup analysis indicating that a large tumor burden is likely to aggravate the harm of waiting. Although our research by itself does not recommend a specific liver transplant policy, it does provide essential information needed to formulate conditions which policies need to meet in order to result in a net benefit. These conditions necessarily need to consider both the impact of dropout and the post-transplant survival. In this research, we focused on the latter and are hereby closer to answering the question of whether the harm of all patients having to wait longer can be offset by the improved allocation of a fraction of the livers. Besides quantifying the harm of waiting on post-transplant survival, to the best of our knowledge, we provide the first causal graph in the field. This graph represents a model of reality and simultaneously explicates our assumptions. The graph is an important starting point for all future empiric liver transplantation research in order to avoid spurious results.

In the literature, the effect of waiting time on survival is studied from two different perspectives. Despite the perspectives

being linked, they are, and should be treated as, distinct. The first perspective is concerned with the causal effect on a patient. That perspective is studied here. The second perspective is concerned with the measurement of the association. Most of the research focuses on the second perspective.[23] All these studies find that waiting longer is associated with improved post-transplant survival[3,24–28] or no harm.[29,30] Many acknowledge that the cause of this association has no oncological basis. They explain the association by the fact that the patients with the most aggressive HCC do not make it to the transplantation and are thus filtered out of the post-transplant survival analysis. This leads to an important difference between the two perspectives. Given a different waiting time policy, for a patient the causal effect of 1 month of extra waiting stays the same, while the strength of the association measured changes depending on the degree of dropout that is characteristic to a particular waiting policy. Given this, we support the assertion of Metha *et al.*[22] that this association should be seen as a perceived risk of transplanting patients too early and not as an actual risk for the patient themself. It is, namely, possible that all patients are harmed by waiting longer but that the measured average post-transplant survival increases due to the shifting proportion of aggressive/non-aggressive cases. Despite this, previous studies addressed the selection bias due to dropout only by means of performing an additional intention-to-treat analysis. Although, the intention-to-treat analysis allows for the inclusion of all patients, it describes a less defined treatment or exposure. In this case the treatment becomes a mixture of waiting time only, and waiting time followed by transplantation. Yet the proportion of these is again dependent on the waiting policy, hereby distorting the measurement of the causal effect.

In addition, because these studies did not use a causal graph, it remains unclear if they sufficiently accounted for confounding bias, as illustrated by the widely varying conditioning sets. The study of Halazun *et al.*[24] addressed confounding in more detail by investigating survival differences between long and short waiting time regions. They suggested a natural experiment was exploited. Hereby implying that patients are randomly assigned to either a long or short waiting time region, resulting in an equal distribution of confounders. However, it remains unclear if the composition of patients was actually comparable, as socioeconomic factors also differ from region to region and do affect survival. As the socioeconomic factors were not part of their conditioning set, confounding biases are still present. Therefore, the causal claim of their research was not well supported. Even though the considerations were more complex, involving equal access and the utility of scarce donor organs, the consensus of the associational studies played an important role in the adoption of the mandatory 6-month waiting policy by the UNOS.

Several authors did, however, study the first perspective using causal inference. Recently, Nagai *et al.* used a before-after study design to investigate the mandatory 6-month waiting policy.[31] Their primary aim was to investigate waiting list mortality and dropout, however, as a secondary outcome they also investigated post-transplant outcomes. The authors reported that the policy change had no effect on the post-transplant mortality. However, a few limitations, inherent to the before-after study design, made interpretation of their results more difficult. The before-after study design was likely biased by trends in post-transplant survival. As the inclusion period spanned more than 5 years, and improvements in terms of imaging and clinical care were not controlled for, the estimates are likely biased toward the null (*i.e.*, the group with a longer waiting time got improved medical treatment in comparison with the group that waited less time). Furthermore, apart from changes in the treatment technique, the composition of the before and after group changed. The authors anticipated a higher dropout rate and better post-transplant outcomes, due to exclusion of the most aggressive HCC during the extended waiting period. However, their results showed both lower dropout and better post-transplant survival in the group subjected to the mandatory 6-month waiting policy. Care could have been improved, but Nagai *et al.* speculated that after the policy change doctors became more reluctant to place patients with advanced disease on the waiting list, knowing that exception points are only assigned after 6 months of waiting. In their analysis they did not address the trends in treatment outcomes or the multitude of selection processes. In addition, they recommended that a longer follow-up was needed to study post-transplant survival in more detail.

An alternative causal inference technique is to exploit the exogenous variation in a treatment created by an instrumental variable. The research by Everhart *et al.* was the first to analyze the survival of patients stratified by blood group.[1] The analysis was, however, limited to stratification. Several other aspects of the study are problematic for identifying the harm of waiting in terms of survival for patients with HCC. Most importantly, the study described patients treated two decades ago (1990-1993) and indications for transplantation and the waiting list policy have changed since. For this reason, their study also did not include patients with HCC and is therefore not representative. Further, their sample size of 673 patients was limited. In addition, the logistic model they used to study time-to-event data was likely biased due to more frequent censoring of patients with a longer survival. Later, Howard expanded on the work of Everhart *et al.* and used blood groups as an instrument in a two-stage regression approach.[32] However, this analysis did not focus on post-transplant survival, but on graft failure within 3 months post-transplant. Our research extends the use of blood groups as an instrumental variable for post-transplant survival.

A strength of our analysis is that our results are subject to minimal (unobserved) confounding. In addition, selection bias was addressed using IPW and missing values were imputed using multiple imputation. However, it is important to note that our research has several limitations. First, the causal graph presented here is our view on reality and we realize that undoubtedly more detail can be added. In the graph the most important simplification we made was that the complexities involved with waiting time being a time varying treatment were simplified to where it is being assigned at once and at listing. We recognize that in reality the waiting time of a patient is not known at the time of listing and that the waiting time depends on intermediate examinations of which the frequency is tailored to the individual patient.

Another limitation is that our analysis of the heterogeneous treatment effect involves parametric assumptions. Although we

expect changes in the treatment effect to be smooth, they are not necessarily linear. Alternatives such as performing the analysis using a subset, kernel smoothing, or K-nearest neighbours are more flexible. However, for this research we valued the ability of interaction terms to test if the treatment effect changed relative to the average treatment effect. Yet we acknowledge that the power of interaction terms to detect heterogeneity is limited.

Our work can be extended in several ways. Our research was quantitative in nature, determining how much waiting harms post-transplant survival. Qualitative research focusing on why waiting harms survival should be performed in a dataset with high quality data regarding follow-up, recurrence and metastatic spread. Another future study, with a more advanced instrumental variable analysis, could unentangle the isolated effects of LRT and waiting time on survival. In the current study, the total or combined effect is analyzed, but quantification of the isolated effect is certainly of value. Furthermore, we advise to repeat the analysis in a non-American dataset and investigate the generalizability of our results.

Overall, we conclude that for transplant candidates with HCC, a prolonged waiting period reduces post-transplant survival. From a patient's perspective, all else being equal, the choice between first LRT and eventually transplantation *vs.* immediate transplantation is in favor of immediate transplantation, due to the harm of waiting time. Yet we stress that the reduction in survival is limited compared to the survival gain from liver transplantation. From a policy perspective, we realize that extra waiting time can be desirable such that more biologically aggressive cancers are filtered out and the utility of the scarce donor livers is increased. Among all suggestions for selection criteria the ones incorporating (a variation of) the test-of-time might be the fairest. Nevertheless, the duration of the test-of-time is bounded, and we highlight that more research is needed to identify the optimal writing time.

Abbreviations

AFP, alpha-fetoprotein; ALBI score, albumin-bilirubin score; ATE, average treatment effect; HCC, hepatocellular carcinoma; IPW, inverse probability weighting; LT, liver transplantation; LRT, loco-regional therapy; OPTN, Organ Procurement and Transplantation Network; UNOS, United Network for Organ Sharing.

Financial support

The authors received no financial support to produce this manuscript.

Conflict of interest

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/disclosure-of-interest/and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

Beumer had the original idea. Beumer, IJzermans, Polak, De Man, Metselaar, Labrecque, and Van Klaveren designed the study. Beumer, IJzermans and Polak helped in the organization of the study and the collection of the data. Beumer performed the statistical analysis which was checked by Van Klaveren and Labrecque. Beumer wrote the first version of the manuscript. The manuscript was critically reviewed and improved upon by all co-authors. IJzermans as guarantor accepts the full responsibility for the work and the conduct of the study, had access to the data, and controlled the decision to publish.

Data availability statement

The data of this study were supplied by United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network in December 2020. The patient level data are available at this institution upon reasonable request.

Statistical script

The R source file is available as online supplement entitled: *Effect of waiting time on survival - blood group IV - UNOS.R.*

Transparency statement

Beumer and IJzermans, as the lead authors, affirm that the manuscript is an honest, accurate, and transparent account of the study; that no important aspects of the study have been omitted.

Disclaimer

The data reported here have been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government.

Acknowledgements

Dr. Naghi, Dr. O'Neill, and Mrs. Fresina are gratefully acknowledged for proofreading the manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2022.100629.

References

[1] Everhart JE, Lombardero M, Detre KM, Zetterman RK, Wiesner RH, Lake JR, et al. Increased waiting time for liver transplantation results in higher MORTALITY1. Transplantation 1997;64(9):1300–1306.
[2] Freeman Jr RB, Edwards EB. Liver transplant waiting time does not correlate with waiting list mortality: implications for liver allocation policy. Liver Transplant 2000;6(5):543–552.
[3] Roberts JP, Venook A, Kerlan R, Yao F. Hepatocellular carcinoma: ablate and wait *vs.* rapid transplantation. Liver Transplant 2010;16(8):925–929.
[4] Kulik L, Abecassis M. Living donor liver transplantation for hepatocellular carcinoma. Gastroenterology 2004;127(5):S277–S282.
[5] OPTN-UNOS. OPTN bylaws effective Dec 6 2021. *Allocation of Livers and liver-intestines*. 2021.
[6] Bryce CL, Chang C-CH, Angus DC, Arnold RM, Farrell M, Roberts MS. The effect of race, sex, and insurance status on time-to-listing decisions for liver transplantation. J Transplant 2010;2010.
[7] Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr 2016;103(4):965–978.
[8] Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573–577.
[9] Efron B. The two sample problem with censored data. In: Paper presented at: proceedings of the fifth Berkeley symposium on mathematical statistics and probability; 1967.
[10] Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv Outcomes Res Methodol 2001;2(3):169–188.
[11] Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007;8(3):206–213.
[12] Rubin DB. Multiple imputation for nonresponse in surveys, vol. 81. John Wiley & Sons; 2004.

[13] Van Buuren S. Flexible imputation of missing data. CRC press; 2018.
[14] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004:615–625.
[15] Angrist JD. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. Am Econ Rev 1990:313–336.
[16] Angrist JD, Keueger AB. Does compulsory school attendance affect schooling and earnings? Q J Econ 1991;106(4):979–1014.
[17] Staiger DO, Stock JH. Instrumental variables regression with weak instruments. USA: National Bureau of Economic Research Cambridge, Mass.; 1994.
[18] Angrist JD, Graddy K, Imbens GW. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Rev Econ Stud 2000;67(3):499–527.
[19] Tchetgen EJT, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. Epidemiology (Cambridge, Mass) 2015;26(3):402.
[20] Aalen OO. A linear regression model for the analysis of life times. Stat Med 1989;8(8):907–925.
[21] Schwartz M, Florman S. Optimal liver allocation for hepatocellular carcinoma: hurry up and wait, but which one when?, vol. 14. Wiley Online Library; 2014. p. 1479–1480.
[22] Mehta N, Heimbach J, Lee D, Dodge JL, Harnois D, Burns J, et al. Wait time of less than 6 and greater than 18 months predicts hepatocellular carcinoma recurrence after liver transplantation: proposing a wait time "sweet spot". Transplantation 2017;101(9):2071–2078.
[23] Lai Q, Lerut J, Study EHCLTE. Waiting time and transplantation for hepatocellular cancer: a balance between tempus fugit and carpe diem. Hepatology (Baltimore, Md) 2015;61(4):1438–1439.
[24] Halazun KJ, Patzer RE, Rana AA, Verna EC, Griesemer AD, Parsons RF, et al. Standing the test of time: outcomes of a decade of prioritizing patients with hepatocellular carcinoma, results of the UNOS natural geographic experiment. Hepatology 2014;60(6):1957–1962.
[25] Salvalaggio P, Felga G, Axelrod D, Della Guardia B, Almeida M, Rezende M. List and liver transplant survival according to waiting time in patients with hepatocellular carcinoma. Am J Transplant 2015;15(3):668–677.
[26] Schlansky B, Chen Y, Scott DL, Austin D, Naugler WE. Waiting time predicts survival after liver transplantation for hepatocellular carcinoma: a cohort study using the U nited N etwork for O rgan S haring registry. Liver Transplant 2014;20(9):1045–1056.
[27] Samoylova ML, Dodge JL, Yao FY, Roberts JP. Time to transplantation as a predictor of hepatocellular carcinoma recurrence after liver transplantation. Liver Transplant 2014;20(8):937–944.
[28] San Miguel C, Vilchez A, Villegas T, Granero K, Becerra A, Lopez M, et al. Influence of waiting list in recurrence disease of hepatocellular carcinoma. In: Paper presented at: transplantation proceedings; 2015.
[29] Palmer WC, Lee D, Burns J, Croome K, Rosser B, Patel T, et al. Liver transplantation for hepatocellular carcinoma: impact of wait time at a single center. Ann Hepatol 2017;16(3):402–411.
[30] Hogen R, Lo M, DiNorcia J, Ji L, Genyk Y, Sher L, et al. More than just wait time? Regional differences in liver transplant outcomes for hepatocellular carcinoma. Transplantation 2019;103(4):747–754.
[31] Nagai S, Kitajima T, Yeddula S, Salgia R, Schilke R, Abouljoud MS, et al. Effect of mandatory 6-month waiting period on waitlist and transplant outcomes in patients with hepatocellular carcinoma. Hepatology 2020;72(6):2051–2062.
[32] Howard D. The impact of waiting time on liver transplant outcomes. Health Serv Res 2000;35(5 Pt 2):1117.