

Systematic Bias in Surgeons' Predictions of the Donor-Specific Risk of Liver Transplant Graft Failure

Michael L. Volk,¹ Meghan Roney,³ and Robert M. Merion²

¹Divisions of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor MI;

²Divisions of Transplantation Surgery, University of Michigan Health System, Ann Arbor MI; and

³Center for Bioethics and Social Sciences in Medicine, University of Michigan, Ann Arbor MI

The decision to accept or decline a liver allograft for a patient on the transplant waiting list is complex. We hypothesized that surgeons are not accurate at predicting donor-specific risks. Surgeon members of the American Society of Transplant Surgeons were invited to complete a survey in which they predicted the 3-year risk of graft failure for a 53-year-old man with alcoholic cirrhosis and a Model for End-Stage Liver Disease score of 21 with a liver from (1) a 30-year-old local donor with traumatic brain death or (2) a 64-year-old regional donor with brain death from a stroke. Complete responses were obtained from 201 surgeons, whose self-reported case volume represents the majority of liver transplants in the United States. The surgeon-predicted 3-year risk of graft failure varied widely (more than 10-fold). In scenario 1, 90% of the respondents provided lower estimates of the graft failure risk than the literature-derived estimate of 21% ($P < 0.001$). In scenario 2, 96% of the responses were lower than the literature-derived estimate of 40% ($P < 0.001$). In conclusion, transplant surgeons vary widely in their predictions of the donor-specific risk of graft failure, and they demonstrate a systematic bias toward inaccurately low estimates of graft failure, particularly for higher risk organs. *Liver Transpl* 19:987-990, 2013. © 2013 AASLD.

Received March 4, 2013; accepted May 21, 2013.

Deceased donor livers available for transplantation vary widely in quality. Donor characteristics such as age, cause of death, and ischemia time can make the difference between a 20% rate of graft failure and a 40% rate of graft failure within 3 years after transplantation.¹

Each time an organ is offered, the surgeon and the potential recipient must decide whether to accept that offer or wait in the hope that a better one will come along. These decisions are high-risk ones; a recent study revealed that 84% of patients who die on the

waiting list have previously declined at least 1 organ offer.² These decisions are also complex ones. Surgeons must incorporate multiple donor factors, recipient factors, and donor-recipient interactions as well as the local magnitude of the organ shortage and various technical and logistical concerns. Thus, it is perhaps not surprising that decisions about organ quality vary widely by transplant center and are susceptible to cognitive biases and external forces such as policy changes and competition between centers.³⁻⁵

Additional Supporting Information may be found in the online version of this article.

Abbreviations: ASTS, American Society of Transplant Surgeons; MELD, Model for End-Stage Liver Disease.

This study was funded by the National Institute of Diabetes and Digestive and Kidney Diseases (K23-K085204 to Michael L. Volk).

The authors have no relevant conflicts of interest.

Address reprint requests to Michael L. Volk, M.D., M.Sc., Division of Gastroenterology and Hepatology, University of Michigan Health System, 1500 East Medical Center Drive, 3912 TC, Ann Arbor MI 48109. Telephone: 734-232-1071; FAX: 734-936-8944; E-mail: mvolk@med.umich.edu

DOI 10.1002/lt.23683

View this article online at wileyonlinelibrary.com.

LIVER TRANSPLANTATION.DOI 10.1002/lt. Published on behalf of the American Association for the Study of Liver Diseases

For these reasons, we hypothesized that surgeons are not accurate at predicting donor-specific risks. We performed a nationwide survey to test this hypothesis.

MATERIALS AND METHODS

Survey Design

Surgeon members of the American Society of Transplant Surgeons (ASTS) were invited by e-mail to complete an online survey in which they were provided clinical scenarios and asked to predict the probability of death or graft failure (hereafter simply called *graft failure*). E-mails were sent in an anonymous fashion via the ASTS administration. The survey, which is shown in the supporting information, was designed to test the following primary hypotheses:

- Hypothesis 1. The variance between surgeons in estimates of the probability of graft failure would be high.
- Hypothesis 2. As a group, the surgeon-predicted graft failure rate for higher risk organs would be systematically low in comparison with quantitative metrics such as the donor risk index.

Three scenarios were presented. The first 2 scenarios were constructed on the basis of the following literature evidence. In scenario 1:

- Average-risk recipient: a 53-year-old man with diabetes and alcoholic cirrhosis complicated by ascites and encephalopathy who has a Model for End-Stage Liver Disease (MELD) score of 21.⁵
- Low-risk donor: a 30-year-old white male with brain death from a gunshot wound, local share (donor risk index 1.0, 3-year graft failure risk 21%).¹

In scenario 2:

- Average-risk recipient: a 53-year-old man with diabetes and alcoholic cirrhosis complicated by ascites and encephalopathy who has a MELD of score 21.⁵
- High-risk donor: a 64-year-old black male with brain death from a stroke, regional share (donor risk index 2.3, 3-year graft failure risk 40%).¹

The order of these scenarios was randomly alternated in order to test for the phenomenon of anchoring. We also hypothesized that surgeons would weigh posttransplant outcomes more heavily than pretransplant outcomes, and we tested this hypothesis by presenting a third scenario:

- Donor: 64 years old.
- Recipient A: hepatitis C virus cirrhosis and a MELD score of 32.
or
- Recipient B: Alcoholic cirrhosis and a MELD score of 17.

Finally, respondents were asked what percentage of time visual inspection plays a dominant role in the

acceptance decision. Given the anonymous nature of the survey, we did not know which of the 1029 individuals in the ASTS database were actively performing liver transplantation. Therefore, the e-mail requested participation only from surgeons who were currently performing liver transplantation. In order to estimate the response rate among surgeons who were actively performing liver transplantation, we asked respondents to report their personal liver transplant volume and compared the sum of the responses to national data. This study was exempted from oversight by our institutional review board. Only the clinical characteristics were provided, not the risk information.

Statistical Analysis

In order to test hypothesis 1, responses of graft failure estimates were displayed graphically, and the variance in the responses was compared visually to that of random chance. Twenty of 201 responses were outliers and were presumed to reflect inadvertent surgeon responses to the probability of graft survival rather than graft failure; those responses were inverted to graft failure for the primary analyses, and sensitivity analyses were also performed through the exclusion of those respondents. In order to test hypothesis 2, a comparison of the responses and the literature-derived estimates was performed with the Kolmogorov-Smirnov test for equality of distribution. The Student *t* test was used to determine whether responses were influenced by the order of the scenarios, and linear regression was used to determine any association between responses and surgeon characteristics such as transplant volume and time since the completion of a fellowship.

RESULTS

E-mails were sent to 1029 ASTS members, and complete responses were obtained from 201 individuals who reported that they were currently performing liver transplantation. On the basis of the self-reported case volume, these 201 surgeons were responsible for 6156 of the 6342 liver transplants (97%) performed in the United States in 2011. The median time since the completion of a fellowship was 11 years, whereas the median time was 15 years in the entire ASTS database ($P < 0.001$). Almost 90% of the respondents (180/201) indicated that the surgeon fielding the offer was the same one performing transplantation at their center.

Surgeons' predictions of the 3-year risk of death or graft failure varied widely and were systematically low in comparison with the literature-derived estimates, as shown in Table 1 and Fig. 1. Figure 1 displays in histogram format the responses with overlaid normal distribution curves, and it demonstrates that the variation in responses approximates what would be expected by random chance. In scenario 1, 90% of the respondents provided lower estimates of the graft failure risk than the literature-derived estimate of

TABLE 1. Surgeons' Predictions of the 3-Year Risk of Graft Failure Versus Estimates from the Literature

Scenario	Median Response (%)	Estimate From Literature (%)	P Value
1. 30-year-old local white donor with brain death from trauma and average-risk recipient	15	21	<0.001
2. 64-year-old regional black donor with brain death from a stroke and average-risk recipient	20	40	<0.001

NOTE: The surgeon predictions were systematically low, particularly for the higher risk scenario.

21%. In scenario 2, 96% of the responses were lower than the literature-derived estimate of 40%. These differences between surgeons' predictions and literature-derived estimates were statistically significant ($P < 0.001$ for both comparisons).

Respondents who received scenario 1 first provided a mean graft failure estimate of 13.7%, whereas the mean estimate was 13.8% from those who received scenario 1 second ($P = 0.9$). Respondents who received scenario 2 second provided a mean estimate of 19.8%, whereas the mean estimate was 23.2% from those who received scenario 2 first ($P = 0.02$). This is suggestive evidence that the responses to scenario 2 were influenced by the question order and that surgeons' decisions about high-risk organs may be anchored more by their most recent experience than by their overall experience and published literature. A sensitivity analysis, in which we excluded subjects who appeared to have responded with predictions of graft survival rather than graft failure, did not change the results (data not shown). None of the individual variables (years of practice, individual case volumes, person who fields offers, or opinions regarding visual inspection) were significantly associated with responses to the clinical scenarios (data not shown).

In scenario 3, respondents were asked to choose whether a liver from a 64-year-old donor should go to (A) a 53-year-old recipient with hepatitis C and a laboratory MELD score of 32 or (B) a 53-year-old recipient with alcoholic cirrhosis and a laboratory MELD score of 17. As shown in Table 2, 74% chose recipient A, and this suggests that most surgeons adhere to the spirit of allocation rules by considering the risk of death on the waiting list more than the predicted posttransplant outcome.

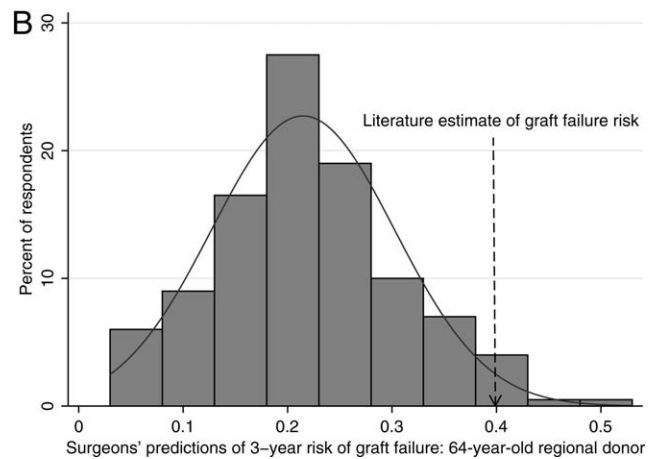
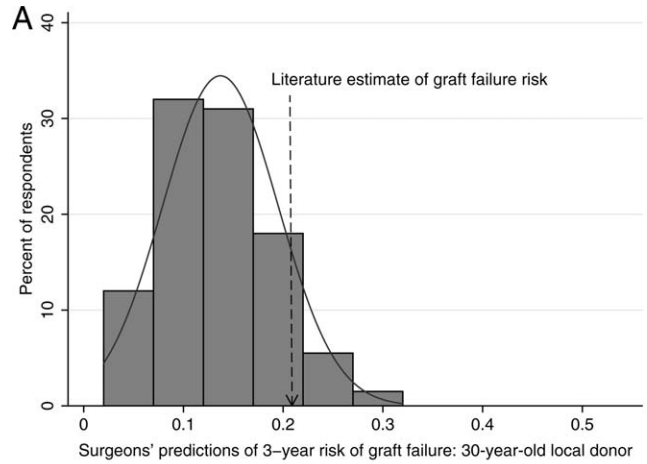


Figure 1. Surgeons' predictions of the 3-year risk of death or graft failure with overlaid normal distribution curves for (A) a 30-year-old local donor and (B) a 64-year-old regional donor.

TABLE 2. Recipient Choices for a Liver From an Older Donor

Preferred Recipient for a Liver From a 64-Year-Old Donor	Respondents [n (%)]
53-year-old woman with hepatitis C virus cirrhosis and a MELD score of 32	149 (74)
53-year-old woman with alcoholic cirrhosis and a MELD score of 17	52 (26)

There was a bimodal distribution of responses regarding the role of a visual inspection of the donor liver in the decision to transplant a particular organ, as shown in Fig. 2. Two-thirds of the respondents replied that a visual inspection played a critical role in <40% of cases, whereas one-fifth replied that it played a critical role in 80% to 100% of cases.

DISCUSSION

This study has demonstrated that liver transplant surgeons vary widely in their estimates of the probability of graft failure in specific clinical scenarios.

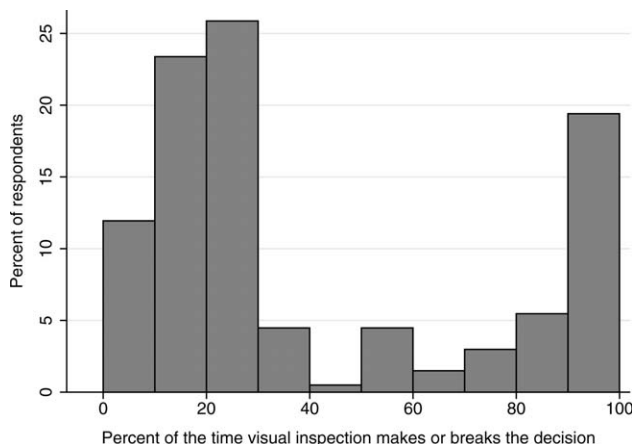


Figure 2. Surgeons' opinions of the importance of visual inspection in deciding whether to accept liver allografts.

Furthermore, as a group, surgeons made systematically low estimates of graft failure probability in comparison with evidence-based estimates from the literature, particularly for higher risk organs. These findings suggest that surgeons are not accurate at predicting donor-specific risks, and they may provide a partial explanation for the wide variability in organ acceptance practices.^{2,5}

These data should not be interpreted as critical of surgeons but should instead highlight the complexity of organ offer decisions. Currently, the myriad data available with an organ offer are evaluated with mental math and gestalt opinion. Such situations, particularly when the risks are high, lead to numerous human inconsistencies and biases, which are the topic of an entire field of study termed *behavioral economics*.⁶ We hypothesize that the availability of a point-of-care decision aid could improve the consistency and accuracy of organ acceptance decisions and thus potentially improve patient outcomes. Such a tool would be intended not to replace clinical judgment but rather to augment it. In fact, the literature on physician decision support suggests that in many situations, the judgment of the so-called expert physicians is aided the most.⁷ We are currently developing such a tool that estimates the probability of survival for a given patient by accepting a given organ offer versus waiting for another one to come along.

The main limitation of this study is the lack of a true gold standard for expected rates of graft failure. The scenarios were created to correspond to categories of the donor risk index, which was derived from data that are now more than 10 years old. Additionally, some of the variation in the responses may reflect true differences in outcomes between centers. Furthermore, because of space constraints and in order to limit the response burden, the scenarios lacked many clinical details that would normally accompany an organ offer. Therefore, these findings may reflect in part the limitations of currently available prognostic tools. However, the lack of precision

in the gold standard is unlikely to fully explain the more than 10-fold variation in risk estimation by respondents. Finally, none of these limitations can explain the systematic underestimation of risk by the respondents.

The choice of a survey study is a second limitation, in that respondents may have been systematically different from nonrespondents. Although we received responses from only 201 of 1029 ASTS members, the most appropriate denominator would have been the number of ASTS members actively performing liver transplantation. This number is unknown, although it is certainly less than the total ASTS membership. The self-reported personal case volumes of respondents may be overestimated because there are approximately 105 liver transplant centers in the United States and many centers have more than 2 surgeons. Nonetheless, the case volume calculation does suggest that the respondents included the majority of surgeons in the United States actively performing liver transplantation. Finally, we chose the endpoint of 3-year graft survival because this is most relevant to donor quality: short-term outcomes are driven largely by recipient and operative characteristics, whereas intermediate-term outcomes are significantly influenced by disease recurrence and other factors mediated by donor characteristics.¹ Risk factors for 1- and 3-year graft failure are highly correlated, so we feel that it is unlikely that the findings would have been different had 1-year graft failure been the primary outcome.

In summary, transplant surgeons are not accurate at predicting donor-specific risks. These findings suggest that organ acceptance decisions may be improved by a point-of-care decision support tool.

REFERENCES

1. Feng S, Goodrich NP, Bragg-Gresham JL, Dykstra DM, Punch JD, DeRoy MA, et al. Characteristics associated with liver graft failure: the concept of a donor risk index. *Am J Transplant* 2006;6:783-790.
2. Lai JC, Feng S, Roberts JP. An examination of liver offers to candidates on the liver transplant wait-list. *Gastroenterology* 2012;143:1261-1265.
3. Volk ML, Lok AS, Pelletier SJ, Ubel PA, Hayward RA. Impact of the Model for End-Stage Liver Disease allocation policy on the use of high-risk organs for liver transplantation. *Gastroenterology* 2008;135:1568-1574.
4. Duan KI, Englesbe MJ, Volk ML. Centers for Disease Control 'high-risk' donors and kidney utilization. *Am J Transplant* 2010;10:416-420.
5. Volk ML, Reichert HA, Lok AS, Hayward RA. Variation in organ quality between liver transplant centers. *Am J Transplant* 2011;11:958-964.
6. Schnier KE, Cox JC, McIntyre C, Ruhil R, Sadiraj V, Turgeon N. Transplantation at the nexus of behavioral economics and health care delivery. *Am J Transplant* 2013;13:31-35.
7. Brailer DJ, Kroch E, Pauly MV. The impact of computer-assisted test interpretation on physician decision making: the case of electrocardiograms. *Med Decis Making* 1997;17:80-86.