Liver and Intestinal Organ Transplantation Committee

Spring 2018
Recent Policy Implementation: Automatic Approval of HCC Exceptions

- Policy implemented December 12, 2017
- Upper limit of AFP 1,000 that is allowed for standard MELD exception (may be treated, and if responds to below 500, also eligible)
- Standardized down-staging policy uniform across regions: patients who present outside of T2 criteria but within down-staging criteria now eligible for standard MELD exception if they are successfully treated and demonstrate a reduction of tumor burden to within T2 criteria.
Recent Public Comment Proposals

- **Modification to Hepatocellular Carcinoma (HCC) Extension Criteria**
  - Proposal provides automatic extension of a HCC exception score for candidates with HCC lesions who met criteria for T2 at initial application, who subsequently fall below T2 lesion criteria at time of their extension because of liver-directed therapies.
  - Intended to revise an effect of HCC down-staging policy implemented Dec 12, 2017 whereby candidates with existing HCC appeals, who were treated to below T2, were no longer auto-approved at next extension.
  - Supportive public comment
- Implemented Feb. 5, 2018 if approved by Executive Committee
Recent policy approval: Enhancing Liver Distribution

- Approved by Board - Dec 2017
- **3 MELD or PELD** points to candidates within the circle or the DSA
- Proximity circles with 150 nautical mile radius around the donor hospital
  - Circles may extend out of the region
Recent policy approval: Enhancing Liver Distribution

- Expanded regional sharing - **Share 32**
  - Adult candidates within the region and/or circle with calculated score 32 or higher (including proximity points)
  - Pediatric candidates within the region and/or circle based on calculated or exception score
  - Adult HAT exception candidates

- Separate allocation for DCD donors and donors at least 70 years old

- Implementation third quarter 2018, after NLRB
Upcoming Policy Implementation: NLRB

- Initial phase will be implemented in 3rd quarter 2018, prior to liver distribution changes

- NLRB
  - Establishes a NLRB with 3 specialty review boards
  - Scores for standardized exceptions will be tied to the median MELD at transplant in the DSA
  - Implementation will occur in phases

- Instructional offerings will be provided to help members prepare for impact of these policy changes
NLRB: Structure

- NLRB is comprised of 3 specialty review boards
  - Adult HCC
  - Adult Other Diagnosis
  - Pediatrics

- Representation
  - Every liver transplant program may appoint a representative

- Reps Responsibilities
  - Reps must vote within 7 days on all exception requests
  - Non-responsiveness may result in suspension of program’s participation in NLRB
NLRB: Structure

- Voting
  - Exception request is randomly assigned to five reps of the appropriate board

- Appeal Process
  - The same five reps of the original request review the appeal

- ART
  - If appeal is denied, a conference call may be requested with the Appeals Review Team (ART)
  - All NLRB members are assigned to serve one month each year on the ART (9 member teams, require 5 for quorum). Conference calls will be held at a fixed day each week and cancelled only if there are no cases
  - Following ART denial, program may initiate final appeal to the Liver Committee
NLRB: Details

- Changes how cases are reviewed and scores awarded
  - Median MELD at transplant (MMaT) score by DSA, minus 3 points, will be the default score for many candidates with standard exceptions
  - Eliminates the “MELD elevator”
- NLRB will review non-standard exception requests
  - Typically related to the MMaT
  - Guidance documents have been created to help the NLRB and to help centers and these can be found on the OPTN website
    https://optn.transplant.hrsa.gov/resources/guidance/liver-review-board-guidance/
NLRB: Details

MMaT Calculation

- OPTN will re-calculate MMaT every 180 days using the previous 365-day cohort.
- At 180 day update, candidates with existing standardized score exceptions will be adjusted.
## Adult Standard Exception Points

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Current Exception Points Assignment</th>
<th>Recommended Proposed Exception Points Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholangiocarcinoma</td>
<td>MELD 22 (w/ 10% point escalator)</td>
<td>MMaT – 3 for DSA</td>
</tr>
<tr>
<td>Cystic Fibrosis</td>
<td>MELD 22 (w/ 10% point escalator)</td>
<td>MMaT – 3 for DSA</td>
</tr>
<tr>
<td>Familial amyloid polyneuropathy</td>
<td>MELD 22 (w/ 10% point escalator)</td>
<td>MMaT – 3 for DSA</td>
</tr>
<tr>
<td>Hepatic artery thrombosis</td>
<td>MELD 40</td>
<td>MELD 40 for DSA</td>
</tr>
<tr>
<td>Hepatopulmonary syndrome</td>
<td>MELD 22 (w/ 10% point escalator if PaO₂ remains under 60 mmHg)</td>
<td>MMaT – 3 for DSA</td>
</tr>
<tr>
<td>Portopulmonary hypertension</td>
<td>MELD 22 (w/ 10% point escalator if repeat heart cath shows MPAP &lt;35)</td>
<td>MMaT – 3 for DSA</td>
</tr>
<tr>
<td>Primary Hyperoxaluria</td>
<td>MELD 28 (w/ 10% point escalator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>HCC</td>
<td>Delay 6 months, then 28, 30, 32, 34</td>
<td>MMaT - 3 for DSA (after delay)</td>
</tr>
</tbody>
</table>

MMaT = Median MELD at Transplant
## Pediatric Standard Exception Points for Candidates 12-17 years old

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Current Exception Points Assignment</th>
<th>Recommended Proposed Initial Exception Points Assignment for 12-17 year olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholangiocarcinoma</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>Cystic Fibrosis</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>Familial amyloid polyneuropathy</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>Hepatic artery thrombosis (not meeting 1A criteria)</td>
<td>MELD 40</td>
<td>MELD or a PELD 40</td>
</tr>
<tr>
<td>Hepatopulmonary syndrome</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>Metabolic Disease</td>
<td>MELD/PELD 30, then status 1B after 30 days</td>
<td>MMaT for DSA, then 1B after 30 days</td>
</tr>
<tr>
<td>Portopulmonary hypertension</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for DSA</td>
</tr>
<tr>
<td>Primary Hyperoxaluria</td>
<td>MELD 28/PELD 41 (w/ 10% elevator)</td>
<td>MMaT for DSA + 3</td>
</tr>
<tr>
<td>HCC</td>
<td>MELD 28/PELD 41 (w/ elevator)</td>
<td>MELD or a PELD 40</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Current Exception Points Assignment</td>
<td>Recommended Proposed Initial Exception Points Assignment for less than 12 year olds</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cholangiocarcinoma</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for region</td>
</tr>
<tr>
<td>Cystic Fibrosis</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for region</td>
</tr>
<tr>
<td>Familial amyloid polyneuropathy</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for region</td>
</tr>
<tr>
<td>Hepatic artery thrombosis (not meeting 1A criteria)</td>
<td>MELD 40</td>
<td>MELD or a PELD 40</td>
</tr>
<tr>
<td>Hepatopulmonary syndrome</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for region</td>
</tr>
<tr>
<td>Metabolic Disease</td>
<td>MELD/PELD 30, then status 1B after 30 days</td>
<td>MMaT for region, then 1B after 30 days</td>
</tr>
<tr>
<td>Portopulmonary hypertension</td>
<td>MELD 22/PELD 28 (w/ 10% elevator)</td>
<td>MMaT for region</td>
</tr>
<tr>
<td>Primary Hyperoxaluria</td>
<td>MELD 28/PELD 41 (w/ 10% elevator)</td>
<td>MMaT for region + 3</td>
</tr>
<tr>
<td>HCC</td>
<td>MELD 28/PELD 41 (w/ elevator)</td>
<td>MELD or a PELD 40</td>
</tr>
</tbody>
</table>
Questions?

Julie Heimbach, MD
Committee Chair
Heimbach.Julie@mayo.edu

Matt Prentice, MPH
Committee Liaison
Matthew.Prentice@unos.org
Toward an accelerated adoption of data-driven findings in medicine

Research, skepticism, and the need to speed up public visibility of data-driven findings

Uri Kartoun

© Springer Nature B.V. 2018

Abstract

To accelerate the adoption of a new method with a high potential to replace or extend an existing, presumably less accurate, medical scoring system, evaluation should begin days after the new concept is presented publicly, not years or even decades later. Metaphorically speaking, as chameleons capable of quickly changing colors to help their bodies adjust to changes in temperature or light, health-care decision makers should be capable of more quickly evaluating new data-driven insights and tools and should integrate the highest performing ones into national and international care systems. Doing so is essential, because it will truly save the lives of many individuals.

Keywords Clinical informatics · Prediction modeling · Electronic medical records · Machine-learning · Data-mining · Cirrhosis · Liver transplantation

Throughout history, skepticism has played an important role in evaluating a variety of phenomena. In medicine, some scientists have occasionally been dismissed as irrational only to be proven right many years later. For instance, Galen, a second-century philosopher and physician, believed that the liver was the source of all veins and the principle organ for blood production (ElMaghawry et al. 2014). Though most of Galen’s writings were incorrect, people still held strong to his beliefs even 1500 years later. Dr. William Harvey was the first to describe blood circulation to the heart, brain, and body in detail. In 1628, in his book, De Motu Cordis (On the Motion of the Heart and Blood), describing the structure of the heart and arteries, he posited for the first time that blood passed through the heart, not the liver as previously believed. Harvey’s findings were ridiculed, and many doctors in the seventeenth century noted that they would “rather err with Galen than proclaim the truth with Harvey.” (Bushak 2015).

Another example of skepticism in medicine concerns non-alcoholic fatty liver disease (NAFLD). Until a few decades ago, the scientific community was undecided about whether NAFLD is actually a clinical condition. An NAFLD diagnosis has important health and clinical implications because it is a risk factor for the development of diseases such as type 2 diabetes mellitus and an independent risk factor for cardiovascular-related mortality and all-cause mortality (Musso et al. 2011; Byrne and Targher 2015). Nonalcoholic steatohepatitis, the progressive form of NAFLD, can result in cirrhosis and hepatocellular carcinoma and is estimated to become the leading indication for liver transplant in the United States by 2020 (Charlton 2008).

Recent remarkable advancements in computer hardware and software and the growing accessibility of electronic medical records (EMRs) have accelerated research on predicting patient outcomes. Such advances have allowed the rapid development of massive-scale predictive models—powerful resources to study disease complications at the population level. Such models have proved highly useful to discovering or confirming disease correlations, sub-categories of diseases, and adverse drug events. The model of the end-stage liver disease (MELD) risk score, for instance, is one of the most important and widely used risk prediction scores in medicine. Unlike in the case of other scores, a patient’s MELD score may indicate the likelihood of a major clinical event for the patient. MELD determines the patient’s rank on the organ allocation waiting list; notably, since 2002, MELD has played a crucial role in determining...
which patient on a waiting list will be the next to receive a liver transplant (Kamath and Kim 2007).

Combining the ability to store and rapidly process the records of millions of individuals by accessing the repositories of Massachusetts General Hospital (MGH), Brigham and Women’s Hospital (BWH), and the IBM Explorys Platform using machine-learning algorithms has helped us create a new and highly accurate score to predict short-term mortality in cirrhosis patients (Kartoun et al. 2017). We took an unbiased approach to the discovery of biomarkers. In this approach, we filtered a large collection of medical records through a feature-selection algorithm and identified a small set of variables that could serve as the most efficient predictors for a given medical outcome. We used the traditional supervised-learning paradigm to assess accuracy and applied standard statistical methods to assess the validity of our approach. We realized that combining the components of MELD with several easily accessible variables would enable us to construct a new score that would be approximately 10% more accurate. We named our new score MELD-Plus. MELD-Plus is an attempt to create a new mortality prediction risk score in cirrhosis. Our unbiased data-driven approach, which involves the use of an algorithm to select predicting variables as well as the large and independent databases used for validation, makes our score a useful tool that could truly save lives. Furthermore, the fact that MELD-Plus’s variables are available for any patient (including total bilirubin, creatinine, albumin, INR, WBC, sodium, total cholesterol, length of stay, and age) makes it easy to calculate the patient’s mortality risk using Excel or to deploy on any digital health repository.

Our preceding manuscript drafts, in which we outlined a better scoring system than MELD, raised significant skepticism from reviewers and editors. Although we were invited to present our earlier findings at a medical informatics conference (Kartoun et al. 2016), leading medical journals repeatedly criticized our work. The criticism always had a reasonable rationale, but our findings and the proposition for an alternative score did not change throughout our resubmissions and were, therefore, kept out of the public eye. Eventually we successfully published our study in October 2017 in PLoS ONE, a peer-reviewed journal.

The main criticism of our initial manuscript was valid: until 2016 we had access to only one source of data (MGH/BWH), and our claim of generalizability was indeed weak. Another criticism was that the interest in such scores might be limited to individual clinicians who are making decisions. This claim, however, may rule out the usefulness of any other type of risk score as well. Another concern was that our predictive model contained too many variables, reducing its practicability in day-to-day use. Such criticism was valid if powerful computers capable of instantaneously processing tremendous collections of EMRs were not in such broad use, as they are today. Future risk scores will likely be composed of tens of thousands of patient characteristics and be calculated automatically as an integrated component of an EMR system to provide real-time decision support to monitor a disease or to prioritize organ transplant candidacy.

Finally, we faced criticism that several of the variables that we used (all selected by a feature-selection algorithm) were associated with cardiovascular risk rather than liver-related mortality. Strikingly, researchers from the Cleveland Clinic validated another of our liver-related studies in which we strengthened the existing knowledge and discovered new biomarkers regarding the interplay between cardiovascular risk and liver disease. Both studies were published in The American Journal of Gastroenterology (Corey et al. 2016; Mehta et al. 2016). Medical publications that describe unbiased approaches to feature selection for developing new scores or to classifying diseases more accurately are rare. A few, however, have been published, including, for instance, a prediction model for 30-day readmission for heart failure patients (Kartoun et al. 2015) and models to classify rheumatoid arthritis (Liao et al. 2010), Crohn’s disease, and ulcerative colitis (Ananthakrishnan et al. 2013). Although we were not criticized explicitly for favoring a data-driven unbiased approach rather than relying on domain expertise, it could have been our use of an approach not yet broadly accepted that have raised further criticism.

Furthermore, our approach also relied on a new text-processing method that we developed to accurately extract concepts from clinical narrative notes. The method, text nailing (TN), raised skepticism in reviewers of medical informatics journals who claimed that TN “relies on simple tricks to simplify the text,” and “leans heavily on human annotation.” TN indeed may seem just like a trick of the light at first glance, but it is actually a fairly sophisticated method that finally caught the attention of more adventurous reviewers and editors who ultimately accepted it for publication (Kartoun 2017a, b). We found TN to be highly accurate, outperforming traditional machine-learning algorithms in multiple scenarios, such as extracting family history of coronary artery disease (Corey et al. 2016), classifying patients with sleep disorders (Beam et al. 2017; Kartoun et al. 2018), and improving the accuracy of the Framingham risk score for patients with NAFLD (Simon et al. 2017).
Toward an accelerated adoption of data-driven findings in medicine

explicit thresholds for listing a patient and be expressed through objective and measurable medical criteria” (Institute of Medicine 1999). Independently, scientists reported in 2000 on a well-validated model and on the creation of a new equation to calculate survival probabilities for patients following a transjugular intrahepatic portosystemic shunt placement (Malinchoc et al. 2000). In February 27, 2002, this equation was selected to serve as the basis for the new allocation policy (Freeman et al. 2002). The equation, forming MELD, has become the standard by which priorities are determined in donor liver allocation, and as expected, implementation of MELD led to an immediate reduction in liver transplant waiting list registrations for the first time in the history of liver transplantation (with a 12% decrease in 2002) (Kamath and Kim 2007). In subsequent years, multiple studies proposed that the incorporation of sodium into the original MELD equation could significantly improve prediction accuracy for liver disease. For instance, a study published in the New England Journal of Medicine in 2008 estimated that using an extended version of MELD, one that incorporated serum sodium levels, would save 90 lives in the period from 2005 to 2006 (Kim et al. 2008). Additional studies supported the usefulness of sodium to improve prediction performance for liver disease (Ruf et al. 2005; Londoño et al. 2007; Luca et al. 2007). The MELD-Na score, an equation that incorporates sodium into MELD, was finally adopted in 2014 (Mulligan and Hirose 2014).

Why did it take many years to adopt MELD-Na, a score that was created by using a data-driven approach, instead of starting to use it, say, in 2008, right after multiple studies demonstrated the advantage of using sodium to improve the prediction accuracy of MELD? The lives of hundreds would have been saved if MELD-Na was in use starting in 2008 rather than in 2014. The reason for the delay was most likely to let the scientific community assess and discuss further the combination’s potential usefulness as well as its drawbacks, a consideration undertaken by a large number of independent investigators and through the use of patient data captured at multiple health systems. Only after broad scientific evidence had been accumulated, was the United Network for Organ Sharing (UNOS) convinced to extend MELD to MELD-Na. Furthermore, UNOS estimated that MELD-Na was expected to save between 50 and 60 lives per year (Mulligan and Hirose 2014), and relevant to MELD-Plus, our experiments demonstrate that while MELD-Na performed slightly better than MELD, MELD-Plus performed significantly better than MELD (>10% better) (Kartoun et al. 2017). Thus, MELD-Plus, if incorporated into hospital systems, could save hundreds of patients every year in the United States alone. Furthermore, as an encouraging first step toward adoption, a very recent study reported that MELD-Plus plays a predictive role in the occurrence of post-liver transplantation acute kidney injury, proposing a broader usefulness beyond mortality prediction (Tudoroiu et al. 2018).

The adoption of MELD-Na would have been faster if the scientific community had been able to publish convincing studies earlier to assess the contribution of sodium to MELD. Organizations such as The American Medical Information Association (AMIA) have encouraged universities as well as commercial companies to form “Challenges,” such as the de-identification and the smoking status challenges (Uzuner et al. 2007, 2008). Such challenges have resulted in a variety of high-impact papers that have significantly enhanced the medical informatics subdomain, as well as the entire health-care domain. If UNOS had worked more collaboratively with AMIA, as well as with The Institute of Electrical and Electronics Engineers (IEEE)’s Engineering in Medicine and Biology Society, new challenges could have been formed, with titles such as “The MELD-Plus Challenge” or “The Liver Disease Challenge,” inviting investigators from all around the globe to assess current scores and propose new scores that might even outperform MELD-Plus. Additional associations, such as the Association for Computing Machinery (ACM), might be encouraged to be involved in such efforts focused on computational assessments of health. Such initiatives could help accelerate the adoption of health-related data-driven findings, as these challenges are expected to produce scientific papers faster and thus support or rule out the usefulness of the newest findings.

In a desirable future scenario, UNOS may decide to replace MELD (or its subsequent score, MELD-Na) with MELD-Plus or even with more advanced futuristic scores that may be developed by other researchers that incorporate, for instance, additional behavioral and genetic aspects. Hypothetically, we can imagine a patient a decade from now who is in need of a liver replacement. That patient might feel encouraged if MELD-Plus was in use, determining more accurately his or her rank on the waiting list. MELD-Plus will not cure that patient, of course, but its ability to assess the severity of a condition more precisely could mean that the patient might wait 2 months less for a new liver than if the original MELD was in use. MELD-Plus, therefore, could save the patient’s life.

On the one hand, skeptics are often proven wrong as science advances. For instance, it took years for the mainstream scientific community to accept Harvey’s contributions over Galen’s. On the other hand, skepticism in medicine is essential, especially regarding questionable treatments and methods and the potential effects of using new medicines. Advances in medicine that have raised significant skepticism include, for example, a human head transplant operation proposed by neurosurgeon Dr. Sergio Canavero (former director of the Turin Advanced Neuromodulation Group, Italy) or a new approach to slow the progression of Alzheimer’s disease proposed by Dr.
Dale Bredesen (University of California, Los Angeles). The development of new risk scores, by contrast, and especially those that are based on components of similar widely used scores, such as MELD, should not be interpreted as questionable and thus should be expected to raise only minor levels of skepticism. Regardless of any specific disease, when a new score is introduced publicly in a peer-reviewed scientific journal, the scientific community would benefit from the availability of mechanisms that could evaluate the scores more rapidly, considering data derived from multiple health-care systems. Especially, official organizations that rely on the scores (e.g., the American Diabetes Association, the American Heart Association, UNOS) would benefit from such mechanisms. Combining the most advanced data-driven algorithms with human expertise is expected to achieve a desirable increase in knowledge, and this could potentially affect decisions to either replace or extend current scoring methods. Incorporating data-driven scores adjusted by human expertise is essential, especially within the context of medical ethics, and will help in deciding which research findings may be worth accelerating and what safeguards need to be provided. For instance, a better scoring system might result in some patients being pushed up the queue for transplantation, but it would result in other patients being pushed down. Additional components coming into play regarding medical scoring systems must consider the ability to assess the true risk of learning algorithms (Kartoun 2018) and the potential for sociological biases at the individual level, as well as at the social level. An algorithm favoring a candidate for transplant based on his or her political views, family status, or sexual preference are just a few examples for such potential biases.

To accelerate the adoption of a new method with a high potential to replace or extend an existing, presumably less accurate, medical scoring system, evaluation should begin days after the new concept is presented publicly, not years. The true risk of learning algorithms (Kartoun 2018) and the potential for sociological biases at the individual level, as well as at the social level. An algorithm favoring a candidate for transplant based on his or her political views, family status, or sexual preference are just a few examples for such potential biases.

To accelerate the adoption of a new method with a high potential to replace or extend an existing, presumably less accurate, medical scoring system, evaluation should begin days after the new concept is presented publicly, not years. The true risk of learning algorithms (Kartoun 2018) and the potential for sociological biases at the individual level, as well as at the social level. An algorithm favoring a candidate for transplant based on his or her political views, family status, or sexual preference are just a few examples for such potential biases.

To accelerate the adoption of a new method with a high potential to replace or extend an existing, presumably less accurate, medical scoring system, evaluation should begin days after the new concept is presented publicly, not years. The true risk of learning algorithms (Kartoun 2018) and the potential for sociological biases at the individual level, as well as at the social level. An algorithm favoring a candidate for transplant based on his or her political views, family status, or sexual preference are just a few examples for such potential biases.

To accelerate the adoption of a new method with a high potential to replace or extend an existing, presumably less accurate, medical scoring system, evaluation should begin days after the new concept is presented publicly, not years. The true risk of learning algorithms (Kartoun 2018) and the potential for sociological biases at the individual level, as well as at the social level. An algorithm favoring a candidate for transplant based on his or her political views, family status, or sexual preference are just a few examples for such potential biases.

Funding The author received honoraria and travel funding from The American Association for the Study of Liver Diseases (October 2017).

Compliance with ethical standards

Conflict of interest The author has declared that no competing interests exist. The author confirms that the commercial affiliation with IBM does not alter his adherence to all Medicine, Health Care and Philosophy policies.

References


Bushak, L. 2015. Mad Scientist: 6 scientists who were dismissed as crazy, only to be proven right years later. Medical Daily, Nov 18, 2015.


