A Algorithms

A.1 Training the ITE model

In Algorithm 1 we give a detailed overview of how our ITE model (Section 4) is trained. There are two main blocks: first the organ clusters are trained; then, we train our ITE model using backpropagation. The GRL allows us to compartmentalise the gradient updates in three distinct updates: one for the outcome prediction parameters, θ_Y; one for the treatment classification parameters, θ_c, based on the organ clusters trained in the first block; and one for the representation parameters, θ_Φ, where the previous losses are combined as in (5). Of course, these three distinct updates are equal to one update over $\frac{\partial \mathcal{L}_{\text{ITE}}(\theta_\Phi, \theta_Y, \theta_p)}{\partial \theta_\Phi \partial \theta_Y \partial \theta_p}$, instead, but found this to abstract away to much detail at the expense of understanding. [47]

Input : model parameters, $\theta_Y, \theta_p, \theta_\Phi$
training data, $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{o}_i, y_i) \ : i = 1, ..., N\}$
amount of organ types, k
learning rate, δ

$c(\mathbf{O}) \leftarrow \text{KMeans}(k, \{\mathbf{o}_i : i = 1, .., N\} \subset \mathcal{D})$; `// Training the organ-clusters`

for $e = 1, ...,$ *max epochs* **do**
  for *Batch* $\mathcal{B} = \{(\mathbf{x}_j, \mathbf{o}_j, y_j, c(\mathbf{o}_j)) \ : j = 1, ..., |\mathcal{B}|\}$ *in epoch* **do**
    Compute $\mathcal{L}_{\text{MSE}}(\theta_\Phi, \theta_Y) = \frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} \mathcal{L}^j_{\text{MSE}}(\theta_\Phi, \theta_Y)$; `// Output prediction loss`
    Compute $\mathcal{L}_{\text{CE}}(\theta_\Phi, \theta_p) = \frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} \mathcal{L}^j_{\text{CE}}(\theta_\Phi, \theta_p)$; `// Treatment classification loss`
    $\theta_Y \leftarrow \theta_Y - \delta \frac{\partial \mathcal{L}_{\text{MSE}}(\theta_\Phi, \theta_Y)}{\partial \theta_Y}$; `// Update outcome prediction parameters`
    $\theta_p \leftarrow \theta_p - \delta \frac{\partial \mathcal{L}_{\text{CE}}(\theta_\Phi, \theta_p)}{\partial \theta_p}$; `// Update classification parameters`
    $\theta_\Phi \leftarrow \theta_\Phi - \delta \left(\frac{\partial \mathcal{L}_{\text{MSE}}(\theta_\Phi, \theta_Y)}{\partial \theta_\Phi} - \gamma \frac{\partial \mathcal{L}_{\text{CE}}(\theta_\Phi, \theta_c)}{\partial \theta_\Phi} \right)$; `// Update representation parameters`
  end
end

Algorithm 1: Training our ITE model.

A.2 Organ-to-patient assignment results

We describe how we evaluate the organ-to-patient assignment policies on real data in Algorithm 2. Should a patient-organ pair be present as-is in the test set, we base the score on the factual outcome, otherwise we use a counterfactual model to provide an outcome. Furthermore, after every organ assignment, we check whether the other patients have died while in $\mathcal{X}_Q$.

Input : data, $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{o}_i, y_i) \ : i = 1, ..., N\}$
test data, $\mathcal{D}_{\text{test}} \subset \mathcal{D}$
counterfactual model, $\hat{Y}$
to-evaluate policy, π

for $t = 0, 1, ..., |\mathcal{D}_{test}|$ **do**
  $\mathcal{X}_Q^t \leftarrow \mathcal{X}_Q^{t-1} \cup \{\mathbf{X}^t\} \in \mathcal{D}_{\text{test}}$;
  $\mathbf{o} \leftarrow \mathbf{O}^t \in \mathcal{D}_{\text{test}}$;
  if $\mathbf{o} \neq \emptyset$ **then**
    $\mathbf{x} \leftarrow \pi(\mathbf{o})$;
    if $(\mathbf{x}, \mathbf{o}) \in \mathcal{D}_{test}$ **then**
      score $\leftarrow$ score $+ \mathcal{D}_{\text{test}}(Y|\mathbf{X}, \mathbf{O})$; `// Patient-organ is in test-set`
    else
      score $\leftarrow$ score $+ \hat{Y}(\mathbf{X}, \mathbf{O})$; `// Patient-organ pair not in test-set`
    end
  end
  for *all* $\mathbf{x} \in \mathcal{X}_Q^t$ **do**
    if $\mathbf{x}$ *died in* $\mathcal{X}_Q^t$ **then**
      if $\mathbf{x}$ *did not receive an organ in* $\mathcal{D}_{test}$ **then**
        score $\leftarrow$ score $+ \mathcal{D}_{\text{test}}(Y|\mathbf{X})$; `// Patient-organ is in test-set`
      else
        score $\leftarrow$ score $+ \hat{Y}(\mathbf{X})$; `// Patient-organ pair not in test-set`
      end
    end
  end
end

Algorithm 2: Evaluation of a policy, π. $\hat{Y}$ is trained using $\mathcal{D}_{\text{train}}$, where $\mathcal{D}_{\text{test}} \cup \mathcal{D}_{\text{train}} = \mathcal{D}$ and $\mathcal{D}_{\text{test}} \cap \mathcal{D}_{\text{train}} = \emptyset$.

B Benchmarks and ablations: details

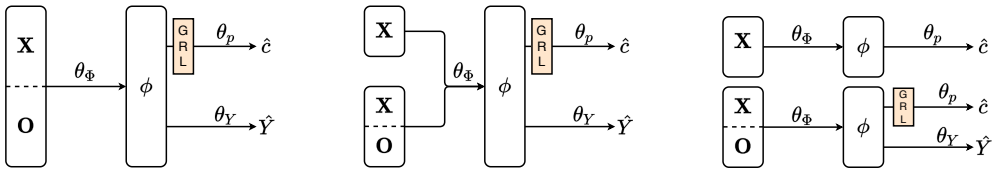
B.1 ITE model

We compare our ITE model with ConfidentMatch, and a multitask network which predicts the outcome per organ-type (based on the organ clusters, $c(\mathbf{O})$).

ConfidentMatch [10, 28]. ConfidentMatch is an ensemble prediction method where the trainingdata is divided in partitions and a predictor method is fitted on every partition. Given an hypothesis space (i.e., prediction methods of a certain VC-dimension), and a maximum partition count, ConfidentMatch optimises a prediction loss over different compositions of the partitions and combinations of predictors from the hypothesis space.

We set the maximum partition count to the number of organ-clusters we used when comparing to our ITE model. For example, when we fit our KMeans cluster with $k = 15$, we restrict ConfidentMatch's partition count to 15. Furthermore, we set the hypothesis space to a `RandomForestRegressor`, support vector machine for regression (`SVR`), and a multi-layered perceptrion (`MLP`) [63].

Multi-task network. Given the organ-clusters, $c(\mathbf{O})$, we train a multi-task network to predict outcomes for every organ-cluster. Specifically, we employ a hard-parameter sharing methodology for our mutli-task network. As we have only one factual outcome for every patient, we set the counterfactuals (i.e. the other organ-clusters) to the prediction when computing the loss. As such,



(a) Padded network. (b) Siamese network. (c) Two networks.

Figure 5: Networks used in the ablation study.

the loss for the counterfactuals remains 0, such that the gradient update is only based on the factual outcome.

B.2 Ablation study networks

As we have in Figure 3, we illustrate the architectures for the ablation studies of our ITE model in Figure 5.

B.3 Organ-to-patient assignment policy benchmarks

Best match (BM) – $\pi_{\text{BM}}(\mathcal{X}_Q, \mathbf{O}) \coloneqq \text{argmax}_{\mathbf{X}\in\mathcal{X}_Q} \{Y^{\mathbf{O}}|\mathbf{X}\}$
A common organ-to-patient assignment policy, selecting whomever is associated with highest life expectancy, given the available organ. We use the same policy for ConfidentMatch, but let $\hat{Y}$ be estimated as in Yoon et al. [10].

First in first out (FIFO) – $\pi_{\text{FIFO}}(\mathcal{X}_Q, \mathbf{O}) \coloneqq \text{argmin}_{\mathbf{X}\in\mathcal{X}_Q} \{s_{\mathbf{X}}\}$
FIFO is a naive scheduling algorithm that simply selects the oldest addition to the queue, based on $s_{\mathbf{X}}$ representing the time of entry, whenever an organ becomes available.

Sickest person first (SPF) – $\pi_{\text{SPF}}(\mathcal{X}_Q, \mathbf{O}) \coloneqq \text{argmin}_{\mathbf{X}\in\mathcal{X}_Q} \{Y^{\emptyset}|\mathbf{X}\}$
Like FIFO, we relate SPF to common queueing strategies, where SPF relates to a prioritised queue. While a measure of sickness is not directly observable, we can approximate it with a patient's estimated life expectancy, $\hat{Y}^{\emptyset}$. That is, a patient with lower life expectancy is considered sicker than a patient with higher life expectancy.

Incremental survival (IS) – $\pi_{\text{IS}}(\mathcal{X}_Q, \mathbf{O}) \coloneqq \text{argmax}_{\mathbf{X}\in\mathcal{X}_Q} \{Y^{\mathbf{O}} - Y^{\emptyset}|\mathbf{X}\}$
Currently employed as policy in the UK, is an estimate of incremental survival rates for an individual patient [9]. In effect, this is a first step towards a counterfactual based approach, though it should be noted that in Neuberger et al. [9], assignment bias was not balanced from the dataset.

C Additional results

C.1 Organ-to-patient assignment

We present a detailed breakdown of the results for our organ-to-patient assignment policy. Specifically, we report: (i) the premature deaths in $\mathcal{X}_Q$ and $\mathcal{X}_M$— where any death in $\mathcal{X}_Q$ and any death before 5 years in $\mathcal{X}_M$ is considered premature; and (ii) the average time alive in $\mathcal{X}_Q$ and $\mathcal{X}_M$— where patients in $\mathcal{X}_M$ are not considered for the average time alive in $\mathcal{X}_Q$.

While these results are informative on the performance of all policies, they require careful attention. For example, a high life expectancy in $\mathcal{X}_Q$ is not necessarily a good property of a policy, as this means that potentially healthier patients are dying before they receive a transplant-organ. Similarly for deathrates in $\mathcal{X}_Q$, where a low deathrate in $\mathcal{X}_Q$ could indicate a waste of tranplant-organs as the policy might prefer sicker patients, with lower life expectancy in $\mathcal{X}_M$.

As mentioned in our related works (Section 2), deciding the objective brings forth an ethical discussion and should be done with great care. By reporting these details, we offer argumentation for clinicians interested in implementing a policy for scarce medical resources. While OrganITE is clearly the best policy when optimising the total population life years (as was our objective, cfr. Section 3), some situations might favor SPF as it allows sicker patients to be treated first. An example of such a situation could be pain relief, where patients suffering the most are aided first.

Table 4: Organ-to-patient evaluation on synthetic data over 10 different folds. Lower is better above the dotted line, and higher is better below the dotted line.

Using our ITE model	FIFO	SPF	BM	IS	CM	OrganITE
Population life years	83509	92153	104889	111228	110129	**112359**
Deaths in $\mathcal{X}_Q$	0.2646	0.2309	0.2357	0.2067	0.2038	**0.1926**
Deaths before 5 years in $\mathcal{X}_M$	0.1683	0.1869	0.1702	0.1593	0.1891	**0.1472**
Avg. days alive in $\mathcal{X}_Q$	32.49	32.38	32.81	32.65	33.12	**37.19**
Avg. years alive in $\mathcal{X}_M$	4.347	4.138	5.088	5.057	5.165	**5.905**
Using TransplantBenefit						
Population life years	n.a.	71953	86664	81813	n.a.	**106392**
Deaths in $\mathcal{X}_Q$	n.a.	**0.1587**	0.2179	0.3201	n.a.	0.3346
Deaths before 5 years in $\mathcal{X}_M$	n.a.	0.3201	0.3152	**0.3048**	n.a.	0.3055
Avg. days alive in $\mathcal{X}_Q$	n.a.	24.46	11.55	20.52	n.a.	**31.50**
Avg. years alive in $\mathcal{X}_M$	n.a.	4.222	5.181	4.572	n.a.	**5.785**

Results on synthetic data. In Table 4 we report a detailed breakdown of the results presented in the upper part of Table 3 in our main text. From this we learn that OrganITE's performance is better across all reported metrics when using our ITE model. However, when using TransplantBenefit we notice weaker performance in death rates, especially in $\mathcal{X}_Q$. This small performance drop is made up for by a significant increase in expected life years after transplantation.

Results on real data. We argue that OrganITE's performance is a result of balancing the various aspects taken into account in organ-to-patient assignment (cfr. Section 4.1). Leveraging this balance results in less deaths and high life expectancy, making OrganITE such a successful assignment-policy.

This balance is visible in Table 5, where we find OrganITE to excel in life expectancy post-transplantation, while maintaining decent performance in the other performance indicators. For example, notice how OrganITE is best in life expectancy post-transplantion across all counterfactual models, while performing: best or second best in premature deaths in $\mathcal{X}_M$; never worst in death rates for $\mathcal{X}_Q$ (even second best for TB as the counterfactual model); and never worst in life expectancy in $\mathcal{X}_Q$ (third using TB as the counterfactual model).

Furthermore, notice how SPF has higher life expectancy and lower death rates in $\mathcal{X}_Q$, while performing very poorly in total population life years. SPF's performance is due to the aforementioned greedy approach to selecting the sickest patients in the waiting queue, $\mathcal{X}_Q$.

Table 5: Organ-to-patient evaluation on real data over 10 different folds. Lower is better above the dotted line, and higher is better below the dotted line.

Nearest neighbor counterfactual	FIFO	SPF	BM	IS	OrganITE
Population life years	94062	83198	100249	102303	**108217**
Deaths in $\mathcal{X}_Q$	0.4414	**0.4055**	0.4209	0.4236	0.4233
Deaths before 5 years in $\mathcal{X}_M$	0.2196	0.2422	0.2093	0.1972	**0.1787**
Avg. days alive in $\mathcal{X}_Q$	28.78	**28.96**	27.83	28.12	28.15
Avg. years alive in $\mathcal{X}_M$	3.805	3.363	4.057	4.138	**4.378**
ITE model counterfactual					
Population life years	95646	84757	92948	105866	**107623**
Deaths in $\mathcal{X}_Q$	0.4294	**0.4081**	0.4189	0.4245	0.4257
Deaths before 5 years in $\mathcal{X}_M$	0.1995	0.2455	0.1996	**0.1794**	0.1806
Avg. days alive in $\mathcal{X}_Q$	28.70	**28.96**	28.33	28.13	28.05
Avg. years alive in $\mathcal{X}_M$	3.875	3.436	3.765	4.282	**4.349**
TB counterfactual					
Population life years	89979	83347	99259	101585	**102773**
Deaths in $\mathcal{X}_Q$	0.4294	**0.4041**	0.4277	0.4167	0.4113
Deaths before 5 years in $\mathcal{X}_M$	0.2195	0.2653	0.2296	0.1994	**0.1906**
Avg. days alive in $\mathcal{X}_Q$	27.21	**28.83**	22.34	25.48	26.77
Avg. years alive in $\mathcal{X}_M$	3.676	3.348	4.008	4.109	**4.153**

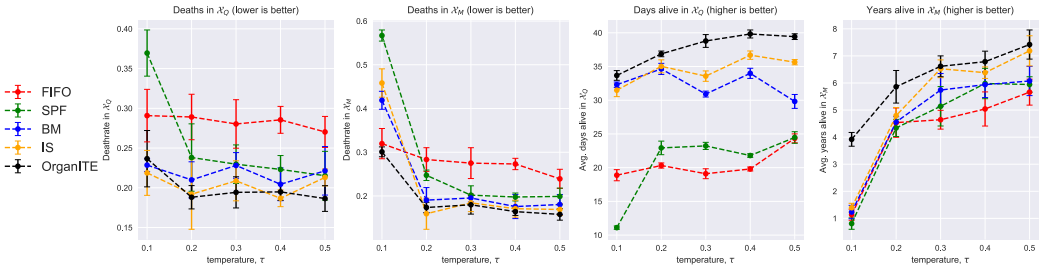


Figure 6: Policy performance indicators in function of temperature, τ, as in (7). From left to right: Deaths in $\mathcal{X}_Q$, Deaths in $\mathcal{X}_M$, Days alive in $\mathcal{X}_Q$, and Years alive in $\mathcal{X}_M$. For deaths, lower is better; for time alive, higher is better.

C.2 Density importance

Adjusting the density $p(\mathbf{O})$ in the synthetic experiment (Table 4) allows us to report on how important this density is to OrganITE and other organ-assignment policies, in terms of the performance metrics presented above. From Figure 4 we learn that OrganITE's expected total population life years seems much less affected by more extreme organ-densities, when compared to the other organ-to-patient assignment policies. Here, we report the same breakdown of our result as we have above, providing a more detailed argument for the organ-to-patient assignment policies.

What is a more extreme density? First, we clarify that by introducing a temperature parameter, we make some organs more rare than other organs. Specifically, because the weights, $\mathbf{w}$, in (7) sum to one ($\sum \mathbf{w} = 1$ making $0 \leq \mathbf{w}_i \leq 1 \forall \mathbf{w}_i$), and every component of $\mathbf{w}$ is divided by $\tau \in (0, 1]$, we make some organ clusters less likely or more likely. As such, leaving patients in need of otherwise more frequent organs (as they are present in $\mathcal{D}$), now less likely to receive a suitable organ.

Why is this important? Having more patients in need of rare organs, requires a policy to explicitly handle these deficits in a way that gives higher priority to these patients, at the cost of potentially multiple better matches with the available organ. From this experiment it is clear that using an estimate of $p(\mathbf{O})$ allows OrganITE to make more informed decisions on how to distribute organs in such an extreme condition.

Results. Consider Figure 6 where we breakdown the result presented in Figure 4 as we did for the results above. From this breakdown we notice that OrganITE's performance does drop when the conditions get more severe, however, they remain much more stable when compared to the other policies.

Another observation we make is how FIFO seems largely unaffected in $\mathcal{X}_Q$. We argue this is due to FIFO being a queue based policy, effectively assigning organs to patients without any estimate of $\hat{Y}$. Naturally, FIFO is triumphed in performance by all estimation based policies, i.e., IS, BM, SPF, and OrganITE; though this might explain why FIFO is performing slightly better in Figure 4 given low τ. Furthermore we notice that IS' performance is fairly close to OrganITE's performance when τ increases, suggesting the density might matter less when the organ-density converges to a uniform density. We reason as such, as IS resembles OrganITE most closely as it too uses an ITE estimation rather than mere prediction.

D Hyperparameters

Table 6: Hyperparameters ITE model and ablations

	Padded network	Two models	Siamese network
Layers	• Φ:(32, 16, 16) • c:(16, 16, k) • Y:(16, 16, 1)	• $\Phi^{\emptyset}$:(16, 16, 16) • $\Phi^{\mathbf{O}}$:(32, 16, 16) • c:(16, 16, k) • $Y^{\emptyset}$:(16, 16, 1) • $Y^{\mathbf{O}}$:(16, 16, 1)	• Φ:(16 and 32, 16, 16) • c:(16, 16, k) • $Y^{\emptyset}$:(16, 16, 1)
Activation	ReLU	ReLU	ReLU
Learning rate	0.0004	$\emptyset$: 0.0001; $\mathbf{O}$:0.0004	0.0004
γ	0.25	0.25	0.25
max-epochs	60	60	60
batch-size	100	100	100

Table 7: Hyperparameters multitask network

Layers	activation	learning rate	max-epochs	batch-size
(32, 16, 16, 16, 16, k)	ReLU	0.0004	60	100

Table 8: Hyperparameters OrganITE

a	b	α_1	α_2	KDE bandwith	KDE kernel
1	1	1	1	1	Gaussian

E Data

Table 9: Features used in real data experiments.

Recipient		Organ		Cause of death	
Name	Mean (std.)	Name	Mean (std.)	Cause	Proportion
Height	168.6 (17.6)	BMI	25.8 (4.95)	Intracranial haemorrhage	57.4%
Gender	37.3% male	Gender	53.3% male	Hypoxic brain damage - all causes	13.8%
Haemoglobin	11.5 (3.8)	Cause of death	see right	Other trauma - accident	3.4%
White blood cells	5.6 (3.8)	Age	46.8 (15.99)	Intracranial - type unclassified (CVA)	3.3%
Platelets	123.5 (90.5)	Donor type	84.7% brain dead	Unspecified	3.1%
Serum urea	6.3 (5.6)		13.7% circulatory death	Trauma RTA - car	2.9%
Serum creatinine	84.9 (43.7)		1.13% living	Intracranial thrombosis	2.1%
Serum albumin	31.9 (6.7)		0.39% domino	Trauma RTA - pedestrian	1.8%
INR	1.4 (0.5)			Living donor	1.4%
Serum bilirubin	87.0 (119.0)	Meningitis	1.4 %		
Serum sodium	136.2 (4.8)	Brain tumour	1.4%		
Serum potassium	4.2 (0.53)	Trauma RTA - motorbike	1.1%		
PO2	12.5 (3.44)			Under 1% not reported.	
AFP level	26.0 (286.37)				

References

[1] James Neuberger. Liver transplantation in the united kingdom. *Liver Transplantation*, 22(8): 1129–1135, 2016.

[2] Muhammad F Dawwas, Alexander E Gimson, James D Lewsey, Lynn P Copley, and Jan HP van der Meulen. Survival after liver transplantation in the united kingdom and ireland compared with the united states. *Gut*, 56(11):1606–1613, 2007.

[3] Rachel J Johnson, Lisa L Bradbury, Kate Martin, James Neuberger, et al. Organ donation and transplantation in the uk—the last decade: a report from the uk national transplant registry. *Transplantation*, 97:S1–S27, 2014.

[4] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3076–3085. JMLR. org, 2017.

[5] Ahmed M. Alaa and Mihaela van der Schaar. Bayesian inference of individualized treatment effects using multi-task gaussian processes. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3424–3432. Curran Associates, Inc., 2017.

[6] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=ByKWUeWA-.

[7] Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Representation learning for treatment effect estimation from observational data. In *Advances in Neural Information Processing Systems*, pages 2633–2643, 2018.

[8] Ioana Bica, James Jordon, and Mihaela van der Schaar. Estimating the effects of continuous-valued interventions using generative adversarial networks. In *Advances in Neural Information Processing Systems*, 2020.

[9] James Neuberger, Alex Gimson, Mervyn Davies, Murat Akyol, John O'Grady, Andrew Burroughs, Mark Hudson, UK Blood, et al. Selection of patients for liver transplantation and allocation of donated livers in the uk. *Gut*, 57(2):252–257, 2008. ISSN 0017-5749. doi: 10.1136/gut.2007.131730. URL https://gut.bmj.com/content/57/2/252.

[10] Jinsung Yoon, Ahmed M Alaa, Martin Cadeiras, and Mihaela Van Der Schaar. Personalized donor-recipient matching for organ transplantation. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[11] John P Dickerson, Ariel D Procaccia, and Tuomas Sandholm. Price of fairness in kidney exchange. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1013–1020, 2014.

[12] Floris Devriendt, Darie Moldovan, and Wouter Verbeke. A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics. *Big data*, 6(1):13–41, 2018.

[13] Ahmed M Alaa, Michael Weisz, and Mihaela Van Der Schaar. Deep counterfactual networks with propensity-dropout. *arXiv preprint arXiv:1706.05966*, 2017.

[14] Dimitris Bertsimas, Nathan Kallus, Alexander M Weinstein, and Ying Daisy Zhuo. Personalized diabetes management using electronic medical records. *Diabetes care*, 40(2):210–217, 2017.

[15] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

[16] Fredrik Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *International conference on machine learning*, pages 3020–3029, 2016.

[17] Fredrik D Johansson, Uri Shalit, Nathan Kallus, and David Sontag. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *arXiv preprint arXiv:2001.07426*, 2020.

[18] Yao Zhang, Alexis Bellot, and Mihaela van der Schaar. Learning overlapping representations for the estimation of individualized treatment effects. In *Proceedingsof the twenty-third international conference on artificial intelligence and statistics*, 2020.

[19] Ioana Bica, Ahmed M Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BJg866NFvB.

[20] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[21] Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani Wu, and Steve Altschuler. Multi-domain adversarial learning. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Sklv5iRqYX.

[22] Ya Li, Xinmei Tian, Mingming Gong, Yajing Liu, Tongliang Liu, Kun Zhang, and Dacheng Tao. Deep domain generalization via conditional invariant adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 624–639, 2018.

[23] Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

[24] Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

[25] Patrick Schwab, Lorenz Linhardt, Stefan Bauer, Joachim M Buhmann, and Walter Karlen. Learning counterfactual representations for estimating individual dose-response curves. *arXiv preprint arXiv:1902.00981*, 2019.

[26] Michael Malinchoc, Patrick S Kamath, Fredric D Gordon, Craig J Peine, Jeffrey Rank, and Pieter CJ Ter Borg. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology*, 31(4):864–871, 2000.

[27] W Ray Kim, Scott W Biggins, Walter K Kremers, Russell H Wiesner, Patrick S Kamath, Joanne T Benson, Erick Edwards, and Terry M Therneau. Hyponatremia and mortality among patients on the liver-transplant waiting list. *New England Journal of Medicine*, 359(10):1018–1026, 2008.

[28] Jinsung Yoon, William R. Zame, Amitava Banerjee, Martin Cadeiras, Ahmed M. Alaa, and Mihaela van der Schaar. Personalized survival predictions via trees of predictors: An application to cardiac transplantation. *PLOS ONE*, 13(3):1–19, 03 2018. doi: 10.1371/journal.pone.0194985. URL https://doi.org/10.1371/journal.pone.0194985.

[29] M Pérez-Ortiz, Manuel Cruz-Ramírez, María Dolores Ayllón-Terán, N Heaton, Rubén Ciria, and César Hervás-Martínez. An organ allocation system for liver transplantation based on ordinal regression. *Applied Soft Computing*, 14:88–98, 2014.

[30] John Rawls. *A theory of justice*. Harvard university press, 2009.

[31] Allan Gibbard. The prospective pareto principle and equity of access to health care. *The Milbank Memorial Fund Quarterly. Health and Society*, 60(3):399–428, 1982. ISSN 01601997. URL http://www.jstor.org/stable/3349800.

[32] Arnold C Harberger. Monopoly and resource allocation. In *Essential Readings in Economics*, pages 77–90. Springer, 1995.

[33] Marc Gravel, Jean Marc Martel, Raymond Nadeau, Wilson Price, and Richard Tremblay. A multicriterion view of optimal resource allocation in job-shop production. *European journal of operational research*, 61(1-2):230–244, 1992.

[34] Leonidas Georgiadis, Michael J Neely, Leandros Tassiulas, et al. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends® in Networking*, 1(1):1–144, 2006.

[35] Joseph Y Hui. Resource allocation for broadband networks. *IEEE Journal on selected areas in communications*, 6(9):1598–1608, 1988.

[36] Milton C Weinstein. Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care*, 6(1):93–103, 1990.

[37] HT Engelhardt. Allocating scarce medical resources and the availability of organ transplantation. *New England Journal of Medicine*, 311(1):66–71, 1984.

[38] Tom Koch. Normative and prescriptive criteria: the efficacy of organ transplantation allocation protocols. *Theoretical Medicine*, 17(1):75–93, 1996.

[39] Robert D Truog, Dan W Brock, Deborah J Cook, Marion Danis, John M Luce, Gordon D Rubenfeld, Mitchell M Levy, et al. Rationing in the intensive care unit. *Critical care medicine*, 34(4):958–963, 2006.

[40] Ezekiel J Emanuel and Alan Wertheimer. Who should get influenza vaccine when not all can? *Science*, 312(5775):854–855, 2006.

[41] Arthur L Caplan. Organ transplant rationing: a window to the future? *Health progress (Saint Louis, Mo.)*, 68(5):40–45, 1987.

[42] Govind Persad, Alan Wertheimer, and Ezekiel J Emanuel. Principles for allocation of scarce medical interventions. *The Lancet*, 373(9661):423–431, 2009.

[43] Valerie A Luyckx, Dominique E Martin, Mohammed Rafique Moosa, Aminu K Bello, Ezequiel Bellorin-Font, Tak Mao Chan, Rolando Claure-Del Granado, Walter Douthat, Somchai Eiam-Ong, Felicia U Eke, et al. Developing the ethical framework of end-stage kidney disease care: from practice to policy. *Kidney International Supplements*, 10(1):e72–e77, 2020.

[44] Jessica B Kramer, Douglas E Brown, and Piroska K Kopar. Ethics in the time of coronavirus: Recommendations in the covid-19 pandemic. *Journal of the American College of Surgeons*, 2020.

[45] Amy Solnica, Leonid Barski, and Alan Jotkowitz. Allocation of scarce resources during the covid-19 pandemic: a jewish ethical perspective. *Journal of Medical Ethics*, 2020.

[46] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[47] Judea Pearl. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

[48] Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076, 09 1962. doi: 10.1214/aoms/1177704472. URL https://doi.org/10.1214/aoms/1177704472.

[49] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837, 09 1956. doi: 10.1214/aoms/1177728190. URL https://doi.org/10.1214/aoms/1177728190.

[50] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

[51] Peter Schulam and Suchi Saria. Reliable decision support using counterfactual models. In *Advances in Neural Information Processing Systems*, pages 1697–1708, 2017.

[52] Michael L Schilsky and Maryam Moini. Advances in liver transplantation allocation systems. *World journal of gastroenterology*, 22(10):2922, 2016.

[53] Jan Lerut, Vincent Karam, Valérie Cailliez, Henri Bismuth, Wojciech G Polak, Bridget Gunson, Rene Adam, and Intestine Transplantation Association (ELITA) European Liver. What did the european liver transplant registry bring to liver transplantation? *Transplant International*, 2020.

[54] DE Schaubel, MK Guidinger, SW Biggins, JD Kalbfleisch, EA Pomfret, P Sharma, and RM Merion. Survival benefit-based deceased-donor liver allocation. *American Journal of Transplantation*, 9(4p2):970–981, 2009.

[55] John M Coombes and James F Trotter. Development of the allocation system for deceased donor liver transplantation. *Clinical Medicine & Research*, 3(2):87–92, 2005.

[56] Alexander Gimson. Development of a uk liver transplantation selection and allocation scheme. *Current Opinion in Organ Transplantation*, 25(2):126–131, 2020.

[57] Ina Jochmans, Marieke van Rosmalen, Jacques Pirenne, and Undine Samuel. Adult liver allocation in eurotransplant. *Transplantation*, 101(7):1542–1550, 2017.

[58] Alessandro Vitale, Michael L Volk, Tullia Maria De Feo, Patrizia Burra, Anna Chiara Frigo, Rafael Ramirez Morales, Luciano De Carlis, Luca Belli, Michele Colledan, Stefano Fagiuoli, et al. A method for establishing allocation equity among patients with and without hepatocellular carcinoma on a common liver transplant waiting list. *Journal of Hepatology*, 60(2):290–297, 2014.

[59] K Ross, RE Patzer, DS Goldberg, and RJ Lynch. Sociodemographic determinants of waitlist and posttransplant survival among end-stage liver disease patients. *American Journal of Transplantation*, 17(11):2879–2889, 2017.

[60] Allyson Hart, David P Schladt, Jessica Zeglin, Joshua Pyke, W Ray Kim, John R Lake, John P Roberts, Ryutaro Hirose, David C Mulligan, Bertram L Kasiske, et al. Predicting outcomes on the liver transplant waiting list in the united states: accounting for large regional variation in organ availability and priority allocation points. *Transplantation*, 100(10):2153, 2016.

[61] Gwilym J Webb, James Hodson, Abhishek Chauhan, John O'Grady, James M Neuberger, Gideon M Hirschfield, and James W Ferguson. Proximity to transplant center and outcome among liver transplant patients. *American Journal of Transplantation*, 19(1):208–220, 2019.

[62] Eric J Keller, Paul Y Kwo, and Paul R Helft. Ethical considerations surrounding survival benefit–based liver allocation. *Liver Transplantation*, 20(2):140–146, 2014.

[63] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.