Learning, evaluating and analyzing a recommendation rule for early blood transfer in the ICU

Antoine Chambaz\textsuperscript{1}, N. Gatulle\textsuperscript{3}, J. Josse\textsuperscript{2}, P. Zhao\textsuperscript{2} and the Traumabase group

\textsuperscript{1} MAP5 (CNRS UMR 8145), Université Paris Cité
\textsuperscript{2} PreMeDICAL, Inria–Inserm
\textsuperscript{3} La Pitié–Salpêtrière, APHP

April 6th, 2023

Statlearn'23
Early blood transfer in the ICU

- Severely injured patients experiencing haemorrhagic shock often require massive transfusion.
- Early transfusion of blood products (plasma, red blood cells and platelets) is common and associated with improved outcomes in the hospital.
- However, determining a right amount of blood products is still a matter of scientific debate.
- Our objective is to build and analyze a recommendation rule for early transfusion of blood products to assist the medical team in the ICU.
Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
The Traumabase group

- Created in 2012, the French Traumabase group is a collaboration focusing on major trauma.
- The group's objectives are to improve major trauma care, to inform public health decisions and to facilitate research.
- Today,
  - 23 French trauma centers contribute to the Traumabase group.
  - The Traumabase data set gathers information on more than 40,000 trauma cases from admission until discharge from the ICU all over France.
The Traumabase data set

For each patient in the Traumabase data set, the information is naturally regrouped into four categories:

- **epidemiological data:**
  - age, gender, BMI, medical history, type of trauma (penetrating?), ...  

- **pre-hospital data:**
  - Glasgow score, pupillary abnormality, blood pressures, maximum heart rate, minimum peripheral capillary oxygen saturation (SpO2), ...  

- **admission data:**
  - same as above & differences pre-hospital vs. admission & temperature, indicator of hemorrhagic shock & early transfusion of blood products, ...  

- **hospital data:**
  - durations of stay in the ICU, in the hospital; survival in the hospital
Early transfusion of blood products

- Blood products:
  - plasma (PFC), red blood cells (CGR), platelets
  - rare, and conserved frozen; need up to 45 minutes to unfreeze

- Despite official recommendations, e.g.
  - Société Française d’Anesthésie Réanimation (2015): ratio PFC/GCR between 1 and 2
  - European 5th guideline on management of major bleeding and coagulopathy following trauma (2019): ratio PFC/GCR 2

  one observes a significant variability in early transfusion of blood products practice

- What could be a best practice is still a matter of scientific debate
Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
A recommendation rule: what is it?

- **Statistical model:**
  each patient contributes a data-structure $O := (W, T, Y) \sim P$ with
  - $W \in \mathcal{W}$: covariates (epidemiological data, pre-hospital data, admission data obtained before the early transfusion of blood products)
  - $T := (A, B, C) \in \mathcal{T}$: quantification of early transfusion of plasma ($A$), red blood cells ($B$) and platelets ($C$)
  - $Y \in \{0, 1\}$: indicator of survival in the hospital

- **Causal model:**
  each patient contributes a counterfactual data-structure $O := (W, (Y[t]), T, T) \sim P$ with
  - $Y[t] \in \{0, 1\}$: the potential/counterfactual indicator of survival in the hospital in a world where $T = t$ would be imposed
  - $Y := Y[T]$: the actual indicator of survival in the hospital, in the real world ("consistency assumption")

...
A recommendation rule: what is it?

- **Statistical model:**
  each patient contributes a data-structure \( O := (W, T, Y) \sim P \) with
  - \( W \in \mathcal{W} \): covariates (epidemiological data, pre-hospital data, admission data obtained before the early transfusion of blood products)
  - \( T := (A, B, C) \in \mathcal{T} \): quantification of early transfusion of plasma \((A)\), red blood cells \((B)\) and platelets \((C)\)
  - \( Y \in \{0, 1\} \): indicator of survival in the hospital

- **Causal model:**
  each patient contributes a counterfactual data-structure \( \mathcal{O} := (W, (Y[t])_{t \in \mathcal{T}}, T) \sim P \) with
  - \( Y[t] \in \{0, 1\} \): the *potential/counterfactual* indicator of survival in the hospital in a world where \( T = t \) would be *imposed*
  - \( Y := Y[T] \): the *actual* indicator of survival in the hospital, in the real world ("consistency assumption")
  - the set of optimal actions for *that* patient: \( \arg \max_{t \in \mathcal{T}} Y[t] \)
  - the set of optimal *recommendation rules* at the population level:
    \[
    \arg \max_{r : \mathcal{W} \rightarrow \mathcal{T}} \mathbb{E}_P \{ Y[r(W)] \}
    \]
A more realistic notion of recommendation rule

Problem:

- viewing the data set as a collection of \(O_1, \ldots, O_i := (W_i, T_i, Y_i), \ldots, O_N \sim P\), the empirical law of \(T\) is highly concentrated around \(\{(a, b, c) = (q, q, q) : q \in \mathbb{N}\} \cap T\) (i.e., same number of bags for all blood products)
- hence would be unrealistic to look for a rule \(r: \mathcal{W} \rightarrow T\)
A recommendation rule

A more realistic notion of recommendation rule

- **Problem:**
  - viewing the data set as a collection of $O_1, \ldots, O_i := (W_i, T_i, Y_i), \ldots, O_N \overset{\text{ind}}{\sim} P$, the empirical law of $T$ is highly concentrated around $\{(a, b, c) = (q, q, q) : q \in \mathbb{N}\} \cap T$ (i.e., same number of bags for all blood products)
  - hence would be **unrealistic** to look for a rule $r : \mathcal{W} \to \mathcal{T}$

- **Proposed solution:**
  - define $\tilde{A} := 1\{A = 1\} + 21\{A \geq 2\}$
  - instead of $\arg \max_{r: \mathcal{W} \to \mathcal{T}} E_P \{Y[r(W)]\}$, target

$$r^*(P) := \arg \max_{r: \mathcal{W} \to \{0,1,2\}} E_P \{Y[\tilde{T}_r]\},$$

where $\tilde{T}_r|W \sim \mathcal{L}_P(T|\tilde{A} = r(W); W)$
Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
Identifiability and learning strategy

- Is $r^*(\mathbb{P})$ (a feature of $\mathbb{P}$) identifiable under $P$ (i.e., a feature of $P$)?
Identifiability and learning strategy

- Is $r^*(\mathbb{P})$ (a feature of $\mathbb{P}$) identifiable under $P$ (i.e., a feature of $P$)?
  - if there exists $\delta > 0$ such that $\delta \leq P(\tilde{A} = \tilde{a}|W)$ for all $\tilde{a} \in \{0, 1, 2\}$ $P$-as. (testable)
  - if $T \perp Y[t]|W$ for all $t \in T$ (“randomization assumption”, untestable)
- then, for every $r : \mathcal{W} \rightarrow \{0, 1, 2\}$,
  $$E_P\{Y[\tilde{T}_r]\} = E_P\{\tilde{Q}_P(r(W), W)\},$$
  where $\tilde{Q}_P(\tilde{A}, Y) := E_P(Y|\tilde{A}, W)$

  hence
  $$r^*(\mathbb{P})(W) = \arg\max_{\tilde{a} \in \{0, 1, 2\}} \tilde{Q}_P(\tilde{a}, W) =: \rho(P)(W)$$
Learning a recommendation rule

Identifiability and learning strategy

- Is $r^*(P)$ (a feature of $P$) identifiable under $P$ (i.e., a feature of $P$)?

  - then, for every $r : \mathcal{W} \to \{0, 1, 2\}$,
    \[
    E_P\{Y[\tilde{T}_r]\} = E_P\{\tilde{Q}_P(r(W), W)\},
    \]
    where $\tilde{Q}_P(\tilde{A}, Y) := E_P(Y|\tilde{A}, W)$
    
  hence
    \[
    r^*(P)(W) = \arg\max_{\tilde{a} \in \{0, 1, 2\}} \tilde{Q}_P(\tilde{a}, W) =: \rho(P)(W)
    \]

- Consequence: an estimator $\tilde{Q}_n$ of $\tilde{Q}_P$ yields an estimator $r_n$ of $\rho(P)$,

  \[
  r_n : w \mapsto \arg\max_{\tilde{a} \in \{0, 1, 2\}} \tilde{Q}_n(\tilde{a}, w)
  \]
Learning a recommendation rule (1/3)

The **missSuperLearner R-package**

- Building an estimator $\bar{Q}_n$ of $\bar{Q}_P$ is a regression task
- There exist many (many (many)) algorithms to learn $\bar{Q}_P$
- Instead of choosing one algorithm, we learn which one performs best for the task at hand
- We rely on super learning (van der Laan et al, 2007, {...}):
  - an *aggregation procedure* based on cross-validation and the choice of a *loss function* tailored to the task at hand
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}}$
Learning a recommendation rule (1/3)

The missSuperLearner R-package

- Building an estimator $\bar{Q}_n$ of $\bar{Q}_P$ is a regression task
- There exist many (many (many)) algorithms to learn $\bar{Q}_P$
- Instead of choosing one algorithm, we learn which one performs best for the task at hand
- We rely on super learning (van der Laan et al, 2007, {...}):
  - an aggregation procedure based on cross-validation and the choice of a loss function tailored to the task at hand
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}}$
- As often, there are a lot of missing data in our data set
Learning a recommendation rule (1/3)

The missSuperLearner R-package

- Building an estimator $\tilde{Q}_n$ of $\bar{Q}_P$ is a regression task
- There exist many (many (many)) algorithms to learn $\bar{Q}_P$
- Instead of choosing one algorithm, we learn which one performs best for the task at hand
- We rely on super learning (van der Laan et al, 2007, {...}):
  - an aggregation procedure based on cross-validation and the choice of a loss function tailored to the task at hand
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}}$
- As often, there are a lot of missing data in our data set
- We adapt the SuperLearner R-package (Polley et al., 2021) and propose the missSuperLearner R-package
  - which deals with missing data
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}} \circ \text{Algo}_{\text{filling in}}$
Learning a recommendation rule (1/3)

The missSuperLearner R-package

- Building an estimator $\tilde{Q}_n$ of $\bar{Q}_P$ is a regression task
- There exist many (many (many)) algorithms to learn $\bar{Q}_P$
- Instead of choosing one algorithm, we learn which one performs best for the task at hand
- We rely on super learning (van der Laan et al, 2007, {...}):
  - an aggregation procedure based on cross-validation and the choice of a loss function tailored to the task at hand
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}}$
- As often, there are a lot of missing data in our data set
- We adapt the SuperLearner R-package (Polley et al., 2021) and propose the missSuperLearner R-package
  - which deals with missing data
  - where the candidate algorithms write as $Q = \text{Algo}_{\text{regression}} \circ \text{Algo}_{\text{screening}} \circ \text{Algo}_{\text{filling in}}$
- A question remains: which loss function?
Value of a recommendation rule

Given a recommendation rule \( r : \mathcal{W} \to \{0, 1, 2\} \), the value of \( r \) is

\[
\mathcal{V}_r(P) := E_P\{\bar{Q}_P(r(W), W)\}
\]

- note: \( \mathcal{V}_r(P) \leq \mathcal{V}_{\rho(P)}(P) \) by definition of \( \rho(P) \)
Learning a recommendation rule (2/3)

Value of a recommendation rule

- Given a recommendation rule \( r : \mathcal{W} \rightarrow \{0,1,2\} \), the value of \( r \) is

\[
\mathcal{V}_r(P) := \mathbb{E}_P\{\bar{Q}_P(r(W), W)\}
\]

- note: \( \mathcal{V}_r(P) \leq \mathcal{V}_{\rho(P)}(P) \) by definition of \( \rho(P) \)

- Back to the question: which loss function?
Learning a recommendation rule (2/3)

Value of a recommendation rule

- Given a recommendation rule \( r : \mathcal{W} \to \{0, 1, 2\} \), the value of \( r \) is

\[
\mathcal{V}_r(P) := \mathbb{E}_P\{\bar{Q}_P(r(W), W)\}
\]

- note: \( \mathcal{V}_r(P) \leq \mathcal{V}_\rho(P) \) by definition of \( \rho(P) \)

- Back to the question: which loss function?

  - archetypal example: in a regression framework, the least-squares loss function

\[
\bar{Q} \mapsto \ell_2(\bar{Q}) \quad \text{s.t.} \quad \ell_2(\bar{Q}) : (w, t, y) \mapsto (y - \bar{Q}(\tilde{a}, w))^2
\]

  induces the least-squares risk \( \mathcal{R}_P^{\ell_2} : \bar{Q} \mapsto \mathbb{E}_P\{\ell_2(\bar{Q})(O)\} \)
Learning a recommendation rule (2/3)

Value of a recommendation rule

- Given a recommendation rule \( r : \mathcal{W} \to \{0, 1, 2\} \), the value of \( r \) is

\[
\mathcal{V}_r(P) := E_P\{\bar{Q}_P(r(W), W)\}
\]

- note: \( \mathcal{V}_r(P) \leq \mathcal{V}_{\rho(P)}(P) \) by definition of \( \rho(P) \)

- Back to the question: which loss function?

  - archetypal example: in a regression framework, the least-squares loss function

\[
\bar{Q} \mapsto \ell_2(\bar{Q}) \quad \text{s.t.} \quad \ell_2(\bar{Q}) : (w, t, y) \mapsto (y - \bar{Q}(\tilde{a}, w))^2
\]

  induces the least-squares risk

\[
\mathcal{R}_P^{\ell_2} : \bar{Q} \mapsto E_P\{\ell_2(\bar{Q})(O)\}
\]

  - by analogy, the loss function

\[
r \mapsto \ell_{\bar{Q}}(r) \quad \text{s.t.} \quad \ell_{\bar{Q}}(r) : (w, t, y) \mapsto -\bar{Q}(r(w), w)
\]

  induces the negative-value risk

\[
\mathcal{R}_P^{\ell_{\bar{Q}}} : r \mapsto E_P\{\ell_{\bar{Q}}(r)(O)\}
\]

  - note: if \( \bar{Q} = \bar{Q}_P \), then \( \mathcal{R}_P^{\ell_{\bar{Q}}} = -\mathcal{V}_r(P) \)

  - challenge: the loss function is indexed by the nuisance parameter \( \bar{Q} \)
Learning a recommendation rule (3/3)

Let $B_n \in \{0, 1\}^n$ be a random vector drawn independently of $O_1, \ldots, O_n$, $n := 0.7 \times N$, and

- $P_n$, the data set $\{O_1, \ldots, O_n\}$ (viewed as a prob. measure)
- $P^0_{n,B_n}$, the $B_n$-specific training data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 0\}$ (viewed as a prob. measure)
- $P^1_{n,B_n}$, the $B_n$-specific testing data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 1\}$ (viewed as a prob. measure)
Learning a recommendation rule (3/3)

A tailored cross-validated risk

- Let $B_n \in \{0, 1\}^n$ be a random vector drawn independently of $O_1, \ldots, O_n$, $n := 0.7 \times N$, and
  - $P_n$, the data set $\{O_1, \ldots, O_n\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^0$, the $B_n$-specific training data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 0\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^1$, the $B_n$-specific testing data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 1\}$ (viewed as a prob. measure)

- Inspired by the negative-value risk $\mathcal{R}_P^\ell \hat{Q} : r \mapsto E_P\{\ell \hat{Q}(r)(O)\}$, we choose the following cross-validated risk: for any algorithm $\text{Algo}$ designed to learn $\hat{Q}_P$,
  
  $$\mathcal{R}_{P,n}(\text{Algo}) := E_{B_n}\left[ E_{P_{n,B_n}^1} \left\{ \ell \hat{Q}_n^*(r_0^0)(O) \right\} \right]$$

---

A. Chambaz  
Early blood transfer in the ICU  
4/6/23
Learning a recommendation rule (3/3)

A tailored cross-validated risk

- Let $B_n \in \{0, 1\}^n$ be a random vector drawn independently of $O_1, \ldots, O_n$, $n := 0.7 \times N$, and
  - $P_n$, the data set $\{O_1, \ldots, O_n\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^0$, the $B_n$-specific training data set $\{O_i: 1 \leq i \leq n \text{ s.t. } B_n(i) = 0\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^1$, the $B_n$-specific testing data set $\{O_i: 1 \leq i \leq n \text{ s.t. } B_n(i) = 1\}$ (viewed as a prob. measure)

- Inspired by the negative-value risk $R_P^{\ell Q}: r \mapsto E_P \{\ell Q(r)(O)\}$, we choose the following cross-validated risk: for any algorithm $\text{Algo}$ designed to learn $\bar{Q}_P$,

$$R_{P,n}(\text{Algo}) := E_{B_n} \left[ E_{P_{n,B_n}^1} \left\{ \ell \bar{Q}_n^*(r_n^0)(O) \right\} \right]$$

cross-validation

- training data

- rule derived from $\bar{Q}_n^0 := \text{Algo}(P_{n,B_n}^0)$

- optimal fluctuation of $\bar{Q}_n^0$

- targeting $\mathcal{V}_{r_n^0}(P)$

- testing data
Learning a recommendation rule (3/3)

A tailored cross-validated risk

- Let $B_n \in \{0, 1\}^n$ be a random vector drawn independently of $O_1, \ldots, O_n$, $n := 0.7 \times N$, and
  - $P_n$, the data set $\{O_1, \ldots, O_n\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^0$, the $B_n$-specific training data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 0\}$ (viewed as a prob. measure)
  - $P_{n,B_n}^1$, the $B_n$-specific testing data set $\{O_i : 1 \leq i \leq n \text{ s.t. } B_n(i) = 1\}$ (viewed as a prob. measure)

- Inspired by the negative-value risk $\mathcal{R}_P^{\ell,Q} : r \mapsto E_P\{\ell_{\bar{Q}}(r)(O)\}$, we choose the following cross-validated risk: for any algorithm $\text{Algo}$ designed to learn $\bar{Q}_P$,

\[
\mathcal{R}_{P,n}(\text{Algo}) := E_{B_n}\left[ E_{P_{n,B_n}^1}\{\ell_{\bar{Q}_n^*}(r_{n,0}^0)(O)\} \right]
\]

- The “discrete” super learner: $\text{Algo}_{SL} := \arg\min_{\text{Algo}} \mathcal{R}_{P,n}(\text{Algo})$

- Our recommendation rule: the rule $r_n$ derived from $\bar{Q}_n = \text{Algo}_{SL}(P_n)$
Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
Estimating the value of $r_n$

- From now on, we consider $r_n$ as a fixed recommendation rule
- What is its value $\nu_{r_n}(P)$? We estimate it by CV-TMLE (Zheng & van der Laan, 2011), using $O_{n+1}, \ldots, O_N$
Estimating the value of $r_n$

- From now on, we consider $r_n$ as a fixed recommendation rule.
- What is its value $\mathcal{V}_{r_n}(P)$? We estimate it by CV-TMLE (Zheng & van der Laan, 2011), using $O_{n+1}, \ldots, O_N$.
- Let $B_N \in \{0, 1\}^{N-n}$ be a random vector drawn independently of $O_{n+1}, \ldots, O_N$, $n := 0.7 \times N$, and
  - $P_N$, data set $\{O_{n+1}, \ldots, O_N\}$ excluding observations with missing data
  - $P^0_{N,B_N}$, the corresponding $B_N$-specific training data set
  - $P^1_{N,B_N}$, the corresponding $B_N$-specific testing data set
Estimating the value of $r_n$

- From now on, we consider $r_n$ as a **fixed** recommendation rule.
- What is its value $\mathcal{V}_{r_n}(P)$? We estimate it by CV-TMLE (Zheng & van der Laan, 2011), using $O_{n+1}, \ldots, O_N$.
- Let $B_N \in \{0, 1\}^{N-n}$ be a random vector drawn independently of $O_{n+1}, \ldots, O_N$, $n := 0.7 \times N$, and
  - $P_N$, data set $\{O_{n+1}, \ldots, O_N\}$ excluding observations with missing data
  - $P_{N, B_N}^0$, the corresponding $B_N$-specific training data set
  - $P_{N, B_N}^1$, the corresponding $B_N$-specific testing data set
- The estimator of $\mathcal{V}_{r_n}(P) := E_P\{\bar{Q}_P(r_n(W), W)\}$ is
  $$\nu_N := E_{B_N} \left[ E_{P_{N, B_N}^1} \{\bar{Q}_N^*(r_n(W), W)\} \right]$$
Evaluating the recommendation rule

Estimating the value of $r_n$

- From now on, we consider $r_n$ as a fixed recommendation rule.
- What is its value $\mathcal{V}_{r_n}(P)$? We estimate it by CV-TMLE (Zheng & van der Laan, 2011), using $O_{n+1}, \ldots, O_N$.
- Let $B_N \in \{0, 1\}^{N-n}$ be a random vector drawn independently of $O_{n+1}, \ldots, O_N$, $n := 0.7 \times N$, and
  - $P_N$, data set $\{O_{n+1}, \ldots, O_N\}$ excluding observations with missing data
  - $P_{N,B_N}^0$, the corresponding $B_N$-specific training data set
  - $P_{N,B_N}^1$, the corresponding $B_N$-specific testing data set
- The estimator of $\mathcal{V}_{r_n}(P) := E_P\{\bar{Q}_P(r_n(W), W)\}$ is
  \[
  \nu_N := E_{B_N} \left[ E_{P_{N,B_N}^1} \{\bar{Q}_{N,B_N}^*(r_n(W), W)\} \right]
  \]
  output of super learner trained on $P_{N,B_N}^0$
  cross-validation testing data
  optimal fluctuation of $\bar{Q}_N^0$
  targeting $\mathcal{V}_{r_n}(P)$

Under mild assumptions, $\nu_N$ is asymptotically Gaussian with an asymptotic variance that can be conservatively estimated.
Evaluating the recommendation rule

Estimating the value of $r_n$

1. From now on, we consider $r_n$ as a fixed recommendation rule.
2. What is its value $\mathcal{V}_{r_n}(P)$? We estimate it by CV-TMLE (Zheng & van der Laan, 2011), using $O_{n+1}, \ldots, O_N$.
3. Let $B_N \in \{0, 1\}^{N-n}$ be a random vector drawn independently of $O_{n+1}, \ldots, O_N$, $n := 0.7 \times N$, and
   - $P_N$, data set $\{O_{n+1}, \ldots, O_N\}$ excluding observations with missing data
   - $P_{N,B_N}^0$, the corresponding $B_N$-specific training data set
   - $P_{N,B_N}^1$, the corresponding $B_N$-specific testing data set
4. The estimator of $\mathcal{V}_{r_n}(P) := E_P\{\tilde{Q}_P(r_n(W), W)\}$ is
   $$\nu_N := E_{B_N} \left[ E_{P_{N,B_N}^1} \{\tilde{Q}_{N}^*(r_n(W), W)\} \right]$$

5. Under mild assumptions, $\nu_N$ is asymptotically Gaussian with an asymptotic variance that can be conservatively estimated.
Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
Variable importance measures (1/2)
Definition and estimation

- Our recommendation rule $r_n$ is a complex object. How does it function? What is the influence of each covariate on $r_n(W)$? What does that even mean?
Variable importance measures (1/2)

Definition and estimation

- Our recommendation rule $r_n$ is a complex object. How does it function? What is the influence of each covariate on $r_n(W)$? What does that even mean?

- Inspired by (Williamson et al., 2021, 2022) we introduce, for every $W_s \subset W$,

$$
\psi_s(P) := \frac{E_P \left\{ [r_n(W) - E_P(r_n(W)|W_s)]^2 \right\}}{\text{Var}_P(r_n(W))} \in [0, 1]
$$
Our recommendation rule \( r_n \) is a complex object. How does it function? What is the influence of each covariate on \( r_n(W) \)? What does that even mean?

Inspired by (Williamson et al., 2021, 2022) we introduce, for every \( W_s \subset W \),

\[
\psi_s(P) := \frac{E_P \left\{ [r_n(W) - E_P(r_n(W)|W-s)]^2 \right\}}{\text{Var}_P(r_n(W))} \in [0, 1]
\]

- if \( \psi_s(P) = 0 \), then \( W_s \) plays no role in \( r_n(W) \)
- if \( \psi_s(P) \neq 0 \), then we can build a consistent, asymptotically Gaussian estimator \( \psi_{N,s} \) of \( \psi_s(P) \)
- if \( \psi_s(P) = 0 \), then \( \psi_{N,s} \) is no longer asymptotically Gaussian!
- testing “\( \psi_s(P) = 0 \)” against “\( \psi_s(P) \neq 0 \)” is not easy...
Variable importance measures (2/2)

Testing

Inspired by (Hudson et al., 2022) we note that, for every $W_s \subset W$,

if $\Psi_s(P) = 0$, then for each $h : \mathcal{W} \to \mathbb{R}$,

$$\Phi_{s,h}(P) := E_P \left\{ [r_n(W) - E_P(r_n(W)|W_{-s})] h(W) \right\} = 0$$
Variable importance measures (2/2)

Testing

- Inspired by (Hudson et al., 2022) we note that, for every \( W_s \subset W \),
  \[
  \text{if } \Psi_s(P) = 0, \text{ then for each } h : \mathcal{W} \to \mathbb{R},
  \Phi_{s,h}(P) := E_P \{ [r_n(W) - E_P(r_n(W)|W-s)] h(W) \} = 0
  \]

- So, given a RKHS and its approximation \( \mathcal{H} \) generated by the \( K \) first eigenfunctions (with a constraint on the norms of its elements), we decide to test
  \[
  H_{0,s} : \text{“ sup } \Phi_{s,h}(P) = 0" \text{ against } \text{“ sup } \Phi_{s,h}(P) \neq 0"
  \]
  and to reject \( \Psi_s(P) = 0 \) for \( \Psi_s(P) \neq 0 \) if we reject \( H_{0,s} \) for \( \neg H_{0,s} \)
Variable importance measures (2/2)

Testing

- Inspired by (Hudson et al., 2022) we note that, for every $W_s \subset W$,

  \[
  \text{if } \Psi_s(P) = 0, \text{ then for each } h : \mathcal{W} \rightarrow \mathbb{R}, \Phi_{s,h}(P) := E_P \{ [r_n(W) - E_P(r_n(W)|W-s)] h(W) \} = 0
  \]

- So, given a RKHS and its approximation $\mathcal{H}$ generated by the $K$ first eigenfunctions (with a constraint on the norms of its elements), we decide to test

  \[
  H_{0,s} : \text{“ sup } \Phi_{s,h}(P) = 0 \text{” against “ sup } \Phi_{s,h}(P) \neq 0 \text{”}
  \]

  and to reject “$\Psi_s(P) = 0$” for “$\Psi_s(P) \neq 0$” if we reject $H_{0,s}$ for $\neg H_{0,s}$

- Using influence-curve-based estimators of any $\Phi_{s,h}(P)$ ($h \in \mathcal{H}$), we can build an estimator $\omega_N$ of $\sup_{h \in \mathcal{H}} \Phi_{s,h}(P)$ by solving an optimization programme
Variable importance measures (2/2)

Testing

- Inspired by (Hudson et al., 2022) we note that, for every $W_s \subset W$,
  
  \[
  \text{if } \Psi_s(P) = 0, \text{ then for each } h : W \rightarrow \mathbb{R}, \\
  \Phi_{s,h}(P) := E_{P} \{[r_n(W) - E_{P}(r_n(W)|W-s)] h(W)\} = 0
  \]

- So, given a RKHS and its approximation $\mathcal{H}$ generated by the $K$ first eigenfunctions (with a constraint on the norms of its elements), we decide to test

  \[
  H_{0,s} : " \sup_{h \in \mathcal{H}} \Phi_{s,h}(P) = 0" \quad \text{against} \quad " \sup_{h \in \mathcal{H}} \Phi_{s,h}(P) \neq 0"
  \]

  and to reject "$\Psi_s(P) = 0$" for "$\Psi_s(P) \neq 0$" if we reject $H_{0,s}$ for $\neg H_{0,s}$

- Using influence-curve-based estimators of any $\Phi_{s,h}(P)$ ($h \in \mathcal{H}$), we can build an estimator $\omega_N$ of $\sup_{h \in \mathcal{H}} \Phi_{s,h}(P)$ by solving an optimization programme

- A multiplier bootstrap procedure allows to approximate the law of $\omega_N$ under the null
Variable importance measures (2/2)

Testing

- Inspired by (Hudson et al., 2022) we note that, for every $W_s \subset W$,

  \[
  \text{if } \Psi_s(P) = 0, \text{ then for each } h : \mathcal{W} \to \mathbb{R}, \\
  \Phi_{s,h}(P) := \mathbb{E}_P \{ [r_n(W) - \mathbb{E}_P(r_n(W)|W-s)] h(W) \} = 0
  \]

- So, given a RKHS and its approximation $\mathcal{H}$ generated by the $K$ first eigenfunctions (with a constraint on the norms of its elements), we decide to test

  \[
  H_{0,s} : \text{“} \sup_{h \in \mathcal{H}} \Phi_{s,h}(P) = 0\text{”} \quad \text{against} \quad \text{“} \sup_{h \in \mathcal{H}} \Phi_{s,h}(P) \neq 0\text{”}
  \]

  and to reject “$\Psi_s(P) = 0$” for “$\Psi_s(P) \neq 0$” if we reject $H_{0,s}$ for $\neg H_{0,s}$

- Using influence-curve-based estimators of any $\Phi_{s,h}(P)$ ($h \in \mathcal{H}$), we can build an estimator $\omega_N$ of $\sup_{h \in \mathcal{H}} \Phi_{s,h}(P)$ by solving an optimization programme

- A multiplier bootstrap procedure allows to approximate the law of $\omega_N$ under the null

- Thus we can test $H_{0,s}$ against $\neg H_{0,s}$, and provide $p$-values
Discussion

Early blood transfer in the ICU

Traumabase: a French observatory for major trauma

A recommendation rule

Learning a recommendation rule

Evaluating the recommendation rule

Analyzing the recommendation rule

Discussion
Discussion

- Work still in progress:
  - theory ✓
  - coding ✓
  - simulation study ✓
  - real data cleaning ✓
  - real data application ongoing

A few insights:
- The empirical probability of survival equals ∼90%: difficult to perform as well in ABC (more severe) patients,
- The empirical probability of survival equals ∼70%: perhaps less challenging to do as well or even better,
- We rely on ∼20 Algo regression, ∼5 Algo screening, 3 Algo filling in results of simulation study encouraging choice of $H$ is a delicate matter (chose $K = 20$). In preliminary results,- the estimated values of the rules were quite close to one another
- Across the validation data, the recommendation rule’s suggestions differed in law from the actual interventions.

Thank you very much for your attention. Any question?
Discussion

- Work still in progress:
  - theory ✓
  - coding ✓
  - simulation study ✓
  - real data cleaning ✓
  - real data application ongoing

- A few insights:
  - the empirical probability of survival equals \( \sim 90\% \): difficult to perform as well
  - the empirical probability of survival equals \( \sim 70\% \): perhaps less challenging to do as well or even better
  - we rely on \( \sim 20 \) Algo regression, \( \sim 5 \) Algo screening, 3 Algo filling in
  - results of simulation study encouraging
  - choice of \( H \) is a delicate matter (chose \( K = 20 \))
  - in preliminary results,
    - the estimated values of the rules were quite close to one another
    - across the validation data, the recommendation rule's suggestions differed from the actual interventions

Thank you very much for your attention. Any question?
Discussion

- Work still in progress:
  - theory ✓
  - coding ✓
  - simulation study ✓
  - real data cleaning ✓
  - real data application ongoing

- A few insights:
  - the empirical probability of survival equals $\sim 90\%$: difficult to perform as well
  - in ABC (more severe) patients, the empirical probability of survival equals $\sim 70\%$: perhaps less challenging to do as well or even better
  - we rely on $\sim 20 \text{ Algo}_{\text{regression}}$, $\sim 5 \text{ Algo}_{\text{screening}}$, 3 \text{ Algo}_{\text{filling in}}$
  - results of simulation study encouraging
  - choice of $\mathcal{H}$ is a delicate matter (chose $K = 20$)
  - in preliminary results,
    - the estimated values of the rules were quite close to one another
    - across the validation data, the recommendation rule’s suggestions differed in law from the actual interventions
Discussion

- Work still in progress:
  - theory ✓
  - coding ✓
  - simulation study ✓
  - real data cleaning ✓
  - real data application ongoing

- A few insights:
  - the empirical probability of survival equals $\sim 90\%$: difficult to perform as well
  - in ABC (more severe) patients, the empirical probability of survival equals $\sim 70\%$: perhaps less
  - challenging to do as well or even better
  - we rely on $\sim 20$ $\text{Algo}_{\text{regression}}$, $\sim 5$ $\text{Algo}_{\text{screening}}$, 3 $\text{Algo}_{\text{filling in}}$
  - results of simulation study encouraging
  - choice of $\mathcal{H}$ is a delicate matter (chose $K = 20$)
  - in preliminary results,
    - the estimated values of the rules were quite close to one another
    - across the validation data, the recommendation rule's suggestions differed in law from the actual interventions

- Thank you very much for your attention. Any question?
Discussion
References (a few among many)

- Montoya et al., *Estimators for the value of the optimal dynamic treatment rule with application to criminal justice interventions*, Int J Biostat, to appear, 2023
- Spahn et al., *The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition*, Critical Care, 23:98, 2019
- Williamson et al., *A general framework for inference on algorithm-agnostic variable importance*, JASA, pages 1–14, to appear, 2022