WHAT SHOULD BE THE FOCUS OF STATISTICAL PERSONALISED MEDICINE?

Philip Dawid
University of Cambridge

Joint work with Mats Stensrud, Aaron Sarvet and Stephen Senn

OVERVIEW

• Scott Mueller and Judea Pearl have recently been promoting a new statistical approach to personalised medicine. In contrast to the traditional statistical approach based on decision theory, their idea is that medical decisions should focus on the “counterfactual contrast” between a patient’s actual response to the treatment they get, and the (necessarily unobservable) “potential response” to the treatment they do not get.

• Mueller and Pearl further claim that combining observational and experimental data can lead to improved decisions.

• We deconstruct and critique their arguments, and reaffirm the appropriateness of the traditional approach.

Medical Decisions

$L$: background characteristics of patient
$X$: give treatment?
$Y$: response

• Data available on $(L, X, Y)$ for many past patients

• How should doctor use $L$ to decide $X$?
  – or to prioritize patients for treatment?

BIG Assumption

$L$: background characteristics of patient
$X$: give treatment?
$Y$: response

• Data available on $(L, X, Y)$ for many past patients

• We assume that new patient is exchangeable with previous cases
  – typically not so for clinical trials

Decision-Theoretic (DT) Approach

- $L$ absent
- Value outcome $Y=y$ with utility $U = u(y)$
- From data, estimate distribution of $Y$ for each hypothesised intervention $X ← x$
- Choose $x$ to maximise expected utility
  \[ E \{ U(Y) \mid X ← x \} \]
  If $Y$ is binary, maximise $P(Y = 1 \mid X ← x)$

Use of Covariate Information

- For current patient, can observe $L=l$
- From data, can estimate conditional distributions of $Y$, given $L, X ← x$
  – statistical/ML techniques
- Choose $x$ to maximise “personalised” expected utility
  \[ E \{ U(Y) \mid L = l, X ← x \} \]
Use of Covariate Information

- For current patient, can’t observe $L$
- From data, can estimate conditional distributions of $Y$, given $L, X \leftarrow x$
  
  — but can’t use this conditional distribution
- Revert: Still choose $x$ to maximise overall expected utility
  
  $E\{U(Y) | X \leftarrow x\}$

Mueller and Pearl (MP) Approach

- Introduce $Y(x)$, “potential response” to intervention $X \leftarrow x$
- Consider $(Y(0), Y(1))$ as having a joint distribution, unaffected by decision $X \leftarrow x$ taken
- Compare outcome $Y(1)$ if treated with outcome $Y(0)$ if not
  
  - $Y(1) = Y(0)$ – no effect
  - $Y(1) > Y(0)$ – benefit
  - $Y(1) < Y(0)$ – harm
- Treat accounting for expected benefit/harm

MP Approach

Three strands to their approach:

1. Assess value of treatment in terms of pair of potential outcomes $(Y(0), Y(1))$
   
   e.g., $P(\text{benefit}), P(\text{harm})$

2. Use data to set bounds on their probabilities

3. Combine experimental and observational data

1. Utility Function

  - Assign utility $U$ depending on both actual and counterfactual response (takes $L = (Y(0), Y(1))$
    
    e.g., $1$ if benefit, $-3$ if harm, $0$ if untreated
    
    - $P(\text{benefit}) - 3 \times P(\text{harm})$ vs. $0$

BUT—

- Impossible ever to observe both $Y(0)$ and $Y(1)$ together
- So typically can not identify joint distribution of $(Y(0), Y(1))$ from data
  
  - Likewise, $P(\text{benefit}), P(\text{harm})$
2. Probability Bounds

*Can not identify* joint distribution of \((Y(0), Y(1))\) from data

**HOWEVER:**

From *experimental data*, yielding the marginal distributions of each of \(Y(0), Y(1)\) separately, we can *set bounds* on their joint distribution (and so on \(P(\text{benefit}), P(\text{harm})\),...)

---

2. Probability Bounds

- From *experimental data*, yielding the marginal distributions of each of \(Y(0), Y(1)\) separately, we can *set bounds* on their joint distribution (and so on \(P(\text{benefit}), P(\text{harm})\),...)

\[
\begin{array}{c|ccc}
   Y(1) & 1 & 0 \\
\hline
   Y(0) & x & .21 - x & .21 \\
   0 & .49 - x & 3 + x & .79 \\
   .49 & .51 & 1 \\
\end{array}
\]

- \(P(Y=1 | X←1) = 0.49\)
- \(P(Y=1 | X←0) = 0.21\)

\(0 \leq x \leq .21\)

---

2. Probability Bounds

- From *experimental data*, yielding the marginal distributions of each of \(Y(0), Y(1)\) separately, we can *set bounds* on their joint distribution (and so on \(P(\text{benefit}), P(\text{harm})\),...)

- This joint distribution is *fully identified* just when either \(Y(0)\) or \(Y(1)\) is degenerate (at 0 or 1)
  - In this case either \(P(\text{benefit}) = 0\), or \(P(\text{harm}) = 0\)

- Then clear what to do
- DT and MP agree

---

Use of Covariate

- From *experimental data*, yielding the marginal distributions of each of \(Y(0), Y(1)\) separately, we can *set bounds* on their joint distribution (and so on \(P(\text{benefit}), P(\text{harm})\),...)

- We can *improve* these bounds if there is a pre-treatment covariate \(L\) such that we know the joint distribution of \((L, Y)\) under either intervention – even if we don’t observe \(L\) for current patient

- Now we have full identification just when, *given* any value of \(L\), either \(Y(0)\) or \(Y(1)\) is degenerate

- DM and MP agree if \(L\) observed

---

3. Combining experimental and observational data

- One possible covariate is the *intention to treat* (ITT) variable \(X^*\), which determines the actual treatment \(X\) in the observational case, but is overridden by the *applied* \(X\) in an experiment

- \(X^*\) sometimes observable in an experiment

- *But even if not, by combining experimental and observational data we can still compute the experimental joint distributions of \((X^*, Y)\)*

- In either case this gives improved MP bounds

---

Another BIG Assumption

- From *experimental data*, yielding the marginal distributions of each of \(Y(0), Y(1)\) separately, we can *set bounds* on their joint distribution (and so on \(P(\text{benefit}), P(\text{harm})\),...)

- If we also have *observational data*, we can improve these bounds

- We must assume observational cases are exchangeable with experimental cases – *typically not so for clinical trials*
EXAMPLE 1

- **Experimental data:**
  \( P(\text{survive} \mid \text{drug}) = 0.49, P(\text{survive} \mid \text{no drug}) = 0.21 \)

**DT:** Better to give treatment

**MP:** We have

\[
0.28 \leq P(\text{benefit}) \leq 0.49 \\
0 \leq P(\text{harm}) \leq 0.21
\]

What to do?

- **Adding Observational data:**
  \( P(\text{drug}) = 0.7, P(\text{survive} \mid \text{drug}) = 0.7, P(\text{survive} \mid \text{no drug}) = 0.7 \)
  we get

\[
P(\text{benefit}) = 0.28 \\
P(\text{harm}) = 0
\]

—so give treatment

---

EXAMPLE 2

- **Experimental data:**
  \( P(\text{survive} \mid \text{drug}) = 0.49, P(\text{survive} \mid \text{no drug}) = 0.21 \)

**DT:** Better to give treatment

**MP:** We have

\[
0.28 \leq P(\text{benefit}) \leq 0.49 \\
0 \leq P(\text{harm}) \leq 0.21
\]

What to do?

- **Adding Observational data:**
  \( P(\text{drug}) = 0.7, P(\text{survive} \mid \text{drug}) = 0.7, P(\text{survive} \mid \text{no drug}) = 0.7 \)
  we get

\[
P(\text{benefit}) = 0.49 \\
P(\text{harm}) = 0.21
\]

\[
0.49 - 3 \times 0.21 < 0. \text{ So don’t treat.}
\]

EXAMPLE 2

- **Experimental data:**
  \( P(\text{survive} \mid \text{drug}) = 0.49, P(\text{survive} \mid \text{no drug}) = 0.21 \)

**DT:** Better to give treatment

**MP:**

\[
P(\text{benefit}) = 0.49 \\
P(\text{harm}) = 0.21
\]

\[
0.49 - 3 \times 0.21 < 0. \text{ So don’t treat.}
\]

In a population of similar individuals, treating everyone (DT) would lead to 49% surviving, whereas treating no one (MP) would lead to 21% surviving!

---

EXAMPLE 3

- **Experimental data:**
  \( P(\text{survive} \mid \text{drug}) = 0.49, P(\text{survive} \mid \text{no drug}) = 0.21 \)

**DT:** Better to give treatment

**MP:** We have

\[
0.28 \leq P(\text{benefit}) \leq 0.49 \\
0 \leq P(\text{harm}) \leq 0.21
\]

What to do?

- **Adding Observational data:**
  \( P(\text{drug}) = 0.7, P(\text{survive} \mid \text{drug}) = 0.29, P(\text{survive} \mid \text{no drug}) = 0.33 \)
  we get

\[
0.28 \leq P(\text{benefit}) \leq 0.40 \\
0 \leq P(\text{harm}) \leq 0.12
\]

What to do?

---

3. Combining experimental and observational data

- In Examples 1 & 2, after combining data, the joint distribution of \((Y(0), Y(1))\) was fully identified

- This happens just when, for either intended treatment \(X^* = x^*\), there is a setting \(X \leftarrow x^\dagger\) under which \(Y\) is determined (= 0 or 1)

- Ideally, observe \(x^*\) for new patient, and choose \(X \leftarrow x^\dagger\) if that yields \(Y = 1\); else, other setting

- DT and MP agree

- More generally, observe \(X^*\) to improve DT
3. Combining experimental and observational data

- In Example 1, after combining data, the joint distribution of \((Y(0), Y(1))\) was fully identified:
  \(P(\text{benefit}) = 0.29, P(\text{harm}) = 0\)

- We find:
  - For \(X^* = 0\), \(X \leftarrow 1 \Rightarrow Y = 1\). So treat
  - For \(X^* = 1\), \(X \leftarrow 0 \Rightarrow Y = 0\). So treat

- So treat anyway! DT and MP agree

Messages?

- Reason for combining observational and experimental data is to give us a handle on unobserved ITT variable \(X^*\)
- If we can observe \(X^*\) in a new patient, this can improve decision-making
- If we can’t, too bad
  - unless utility depends on unobserved features
  - as it does for MP
- Is such a utility reasonable?

REFERENCES

Ang Li and Judea Pearl (2019).
Unit selection based on counterfactual logic.

Scott Mueller and Judea Pearl (2022).
Personalized decision making – a conceptual introduction.