Zapiszmy poszczególne warunki, wprowadzając celowo pewną redundancję, aby łatwiej było korzystać z nich bezpośrednio.

DEMOGRAPHIC PARITY:

$$(1) \begin{cases} \mathbb{P}(\hat{Y}=1 \mid A=a) = \mathbb{P}(\hat{Y}=1 \mid A=b) \\ \mathbb{P}(\hat{Y}=0 \mid A=a) = \mathbb{P}(\hat{Y}=0 \mid A=b) \end{cases}$$

EQUAL OPPORTUNITY:

$$(2) \begin{cases} \mathbb{P}(\hat{Y}=1 \mid Y=1, A=a) = \mathbb{P}(\hat{Y}=1 \mid Y=1, A=b) \\ \mathbb{P}(\hat{Y}=0 \mid Y=1, A=a) = \mathbb{P}(\hat{Y}=0 \mid Y=1, A=b) \\ \mathbb{P}(\hat{Y}=1 \mid Y=0, A=a) = \mathbb{P}(\hat{Y}=1 \mid Y=0, A=b) \\ \mathbb{P}(\hat{Y}=0 \mid Y=0, A=a) = \mathbb{P}(\hat{Y}=0 \mid Y=0, A=b) \end{cases}$$

PREDICTIVE PARITY:

$$(3) \begin{cases} \mathbb{P}(Y=1 \mid \hat{Y}=1, A=a) = \mathbb{P}(Y=1 \mid \hat{Y}=1, A=b) \\ \mathbb{P}(Y=0 \mid \hat{Y}=1, A=a) = \mathbb{P}(Y=0 \mid \hat{Y}=1, A=b) \\ \mathbb{P}(Y=1 \mid \hat{Y}=0, A=a) = \mathbb{P}(Y=1 \mid \hat{Y}=0, A=b) \\ \mathbb{P}(Y=0 \mid \hat{Y}=0, A=a) = \mathbb{P}(Y=0 \mid \hat{Y}=0, A=b) \end{cases}$$

Natomiast trywialna sytuacja, czyli niezależność Y i A wyraża się równaniem:

$$(4) \quad \mathbb{P}(Y=1 \mid A=a) = \mathbb{P}(Y=1 \mid A=b)$$

Czyli trzeba do rozwiązania wykazać następujące implikacje (teza zadania jest równoważna stwierdzeniu, że z dowolnych dwóch warunków fairness wynika trywialność):

~~[illegible]~~

- $(1) \wedge (2) \Rightarrow (4)$
- $(1) \wedge (3) \Rightarrow (4)$
- $(2) \wedge (3) \Rightarrow (4)$

Pierwsza implikacja okazuje się nie zachodzić. Istotnie, rozważmy rozkład łączny, dla $\varepsilon < \frac{1}{8}$:

$$\mathbb{P}(\hat{Y}=1, Y=0, A=a) = \mathbb{P}(\hat{Y}=1, Y=1, A=a) = \mathbb{P}(\hat{Y}=0, Y=0, A=a) =$$
$$= \mathbb{P}(\hat{Y}=0, Y=1, A=a) = \frac{1}{8}$$
$$\mathbb{P}(\hat{Y}=1, Y=1, A=b) = \mathbb{P}(\hat{Y}=0, Y=1, A=b) = \frac{1}{8} - \varepsilon$$
$$\mathbb{P}(\hat{Y}=1, Y=0, A=b) = \mathbb{P}(\hat{Y}=0, Y=0, A=b) = \frac{1}{8} + \varepsilon$$

(1) zachodzi, bo mamy $\mathbb{P}(\hat{Y}=1 | A=a) = \mathbb{P}(\hat{Y}=1 | A=b) = \frac{1}{2}$ itd.

(2) zachodzi, bo mamy $\mathbb{P}(\hat{Y}=1 | Y=1, A=a) = \mathbb{P}(\hat{Y}=1 | Y=1, A=b) =$
$= \mathbb{P}(\hat{Y}=1 | Y=0, A=a) = \mathbb{P}(\hat{Y}=1 | Y=0, A=b) = \frac{1}{2}$ itd.

ale $\mathbb{P}(Y=1 | A=a) = \frac{1}{2} \neq \frac{1}{2} - 4\varepsilon = \mathbb{P}(Y=1 | A=b)$,
czyli (4) nie zachodzi. Oznacza to, że w treści zadania jest jakaś nieścisłość i definicja sytuacji trywialnej powinna być szersza (np. nie dopuszczać $\hat{Y} \perp Y$). □

Dowód implikacji $(1) \wedge (3) \Rightarrow (4)$ to prosty rachunek:

$$\mathbb{P}(Y=1 | A=a) = \mathbb{P}(Y=1 | \hat{Y}=1, A=a)\, \mathbb{P}(\hat{Y}=1 | A=a) + \mathbb{P}(Y=1 | \hat{Y}=0, A=a)\, \mathbb{P}(\hat{Y}=0 | A=a) \overset{(1)(3)}{=} \mathbb{P}(Y=1 | \hat{Y}=1, A=b)\, \mathbb{P}(\hat{Y}=1 | A=b) + \mathbb{P}(Y=1 | \hat{Y}=0, A=b)\, \mathbb{P}(\hat{Y}=0 | A=b) = \mathbb{P}(Y=1 | A=b)$$ □.

Dla dowodu $(2) \wedge (3) \Rightarrow (4)$ zauważmy, że dzieląc przez siebie odpowiednie równości wyznaczymy, że stosunek

$$\mathbb{P}(\hat{Y}=1, Y=1 | A=a) : \mathbb{P}(\hat{Y}=1, Y=0 | A=a) : \mathbb{P}(\hat{Y}=0, Y=0 | A=a) : \mathbb{P}(\hat{Y}=1, Y=0 | A=a)$$

jest taki sam, jak

$$\mathbb{P}(\hat{Y}=1, Y=1 | A=b) : \mathbb{P}(\hat{Y}=1, Y=0 | A=b) : \mathbb{P}(\hat{Y}=0, Y=0 | A=b) : \mathbb{P}(\hat{Y}=1, Y=0 | A=b)$$

Przykładowo, równość $\frac{\mathbb{P}(\hat{Y}=1, Y=1 | A=a)}{\mathbb{P}(\hat{Y}=1, Y=0 | A=a)} = \frac{\mathbb{P}(\hat{Y}=1, Y=1 | A=b)}{\mathbb{P}(\hat{Y}=1, Y=0 | A=b)}$ wyznaczymy dzieląc (3.1) przez (3.2) (pierwsza równość z (3) przez drugą) itd.

Z równości tych stosunków wynika, że iloraz

$$\frac{\mathbb{P}(\hat{Y}=i, Y=j \mid A=a)}{\mathbb{P}(\hat{Y}=i, Y=j \mid A=b)}$$

, dla $i,j \in \{0,1\}$ jest stały.

Nazwijmy go c. Zatem dla $i,j \in \{0,1\}$ mamy

$$\mathbb{P}(\hat{Y}=i, Y=j \mid A=a) = c \cdot \mathbb{P}(\hat{Y}=i, Y=j \mid A=b).$$

Dodając te 4 równości stronami uzyskujemy $c=1$.

Ostatecznie:

$$\mathbb{P}(Y=1 \mid A=a) = \mathbb{P}(Y=1, \hat{Y}=1 \mid A=a) + \mathbb{P}(Y=1, \hat{Y}=0 \mid A=a) =$$
$$= \mathbb{P}(Y=1, \hat{Y}=1 \mid A=b) + \mathbb{P}(Y=1, \hat{Y}=0 \mid A=b) =$$
$$= \mathbb{P}(Y=1 \mid A=b) \quad \square$$