**Supplementary Note A**

**Algorithms**

To solve the optimization problems (4) and (7), we first need to use sampling formulas to approximate the expectations. We first discuss implementation of the imputation block.

*Imputation block optimization*

Assume that individuals are sampled. Sampling approximation of is given by

, (A1)

where

To enforce that the estimated factual outcome should be as close to the observed factual outcome as possible (2), we post the following restriction:

. (A2)

The optimization problem (4) can be implemented by

, (A3)

. (A4)

Optimization problems (A3) and (A4) can be solved by backpropagation (stochastic gradient decent) algorithms (4). The details for the algorithms are given in supplementary note B.

**ITE block optimization**

ITE block intends to estimate the counterfactual outcomes using the observed outcomes and imputed counterfactual outcomes. Its performance metrics are defined

for (binary treatments):

, (A5)

for :

. (A6)

Sampling formula for is

. (A7)

The optimization problem (7) for ITE can be reformulated as

, (A8)

. (A9).

Again, stochastic gradient descent methods can be used to solve optimization problems (A8) and (A9). Algorithms for their numerical implementation are similar to algorithms for the imputation block.