



HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

Epidemiology. 2020 July ; 31(4): e33–e34. doi:10.1097/EDE.0000000000001189.

Re. Selecting Optimal Subgroups for Treatment Using Many Covariates

Bas B. L. Penning de Vries,

Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

Rolf H. H. Groenwold,

Departments of Clinical Epidemiology and Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

Alex Luedtke

Department of Statistics, University of Washington, Seattle, Washington

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

To the Editor:

In a recent publication, VanderWeele et al.¹ considered the task of finding a treatment subgroup that maximizes the mean potential outcome. They showed that the task can sometimes be considerably simplified by deriving optimal treatment assignment rules of a simple form: assign treatment in a greedy fashion to all individuals with the next largest benefit (i.e., the difference in potential outcome means given covariates) or the next highest benefit–cost ratio (with cost being a positive function of baseline covariates) until the resource or cost constraint, respectively, is exceeded. As they state in their eAppendix; <http://links.lww.com/EDE/B655>, the optimality of the rules relies critically on the assumption that there are no ties between individuals. Although tied treatment effects or benefit–cost ratios may occur with many covariates, they are perhaps more realistic when few and only discrete baseline variables are considered to define treatment rules.

Consider for example the setting of the Table and suppose that the total cost may not exceed 130. According to the rule of VanderWeele et al.,¹ individuals in the first stratum should be assigned treatment. Because the presented rules assign treatment to either all or no individuals in any given stratum, no more individuals can be selected without violating the cost constraint. This rule yields a mean potential outcome of 2.3. However, because of ties, a better rule that likewise selects either all or no individuals of a stratum, does exist: assign treatment to strata 2 and 3 (with a mean potential outcome of 2.5). Thus, in the presence of ties, the optimal rule need not be greedy (see also the literature on the classic knapsack problem; e.g., Korte and Vygen²). We note that a better rule may be obtained by augmenting our data with a sequence of independent, possibly unfair, coin tosses. As shown in the

b.b.l.penning_de_vries@lumc.nl.

The authors report no conflicts of interest.

eAppendix; <http://links.lww.com/EDE/B655> (but see also Luedtke and van der Laan³), maximizing the mean potential outcome across rules of this kind is achieved in the cost-constrained setting by treating those with a benefit–cost ratio strictly greater than some positive constant and a random selection of those with a benefit–cost ratio that equals that constant. For our example, this means treating all members of stratum 1 as well as those members of strata 2 and 3 whose independent coin toss, with probability 3/13 of showing heads, results in heads (mean potential outcome: 3.5).

It seems unlikely that these treatment rules would be implemented via biased coin tosses in real-world settings. If resources are made available in a single batch, one could calculate the amount of resources that would need to be allocated to the “always-treat” portion of the population, reserve this portion of resources for always-treat individuals, and then allocate the remainder to the “sometimes-treat” portion of the population on a first-come, first-serve basis until that portion of resources runs out. Bias could however be introduced by doing this, for example, when sometimes-treat individuals who visit the clinic more frequently are systematically less (or more) likely to benefit from treatment. However, there may be ways to account for this (e.g., by including frequency of visits as a covariate).

Finally, we add that with multiple treatment levels and cost constraints, mean potential outcomes need not be optimized by the greedy approach of assigning to subjects the treatment level with the highest benefit–cost ratio above or at treatment level-specific thresholds (to satisfy cost constraints), even if the observed data are augmented with a sequence of independent coin tosses (eAppendix; <http://links.lww.com/EDE/B655>). Regardless of the form the rule should take, however, we encourage researchers to follow VanderWeele et al.¹ in taking a more formal approach to “precision medicine” with clearly specified objectives, so that the optimal rule form may be derived and estimation strategies be evaluated.

Acknowledgments

R.H.H.G. was funded by the Netherlands Organization for Scientific Research (NWO-Vidi project 917.16.430). A.L. was funded by the National Institutes of Health through award number DP2-LM013340. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies.

REFERENCES

1. VanderWeele TJ, Luedtke AR, van der Laan MJ, Kessler RC. Selecting optimal sub-groups for treatment using many covariates. *Epidemiology*. 2019;300:334–341.
2. Korte B, Vygen J. *Combinatorial Optimization: Theory and Algorithms*. 4th ed Heidelberg: Springer; 2008.
3. Luedtke AR, van der Laan M.J. Optimal dynamic treatments in resource-limited settings. *Int J Biostat*. 2016;120:283–303.

Characteristics of Hypothetical Population of Size 100 with Baseline Covariates Forming Five Strata

TABLE.

	Stratum				
	1	2	3	4	5
Number of individuals	25	20	10	15	30
Conditional mean potential outcome					
Under no treatment	-5	4	0	-5	-5
Under treatment	15	20	20	5	-15
Cost of treatment per individual	4	4	5	10	10
Benefit-cost ratio	5	4	4	1	-1

If those and only those in stratum 1 are treated, the total cost is 25×4= 100 and the mean potential outcome is. If those and only those patients in strata 2 and 3 are treated, the total cost is, and the mean potential outcome is. If patients in stratum 1 are treated with probability 1, patients in strata 2 and 3 with probability 3/13, and the rest with probability 0, the expected total cost is and the mean potential outcome is.