Supplementary materials

Table 1: Results of existing simulation studies regarding the three performance criteria. The first column indicates the study in which the results were found, the second column the outcome measure that was used, the third column the scenario(s) under which the results were found and the last column the order of of performance for the methods under study, where a>b indicates that method a performed better than method b, and a≈b indicates that the methods performed about equally well.

Study [1]	Outcome measure	Condition	Result
	Value of the objective function		
[Zhu et al., 2016]	Expected potential outcome	If tree-based regime	TR1>EE4>TR3>TR9
[Zhu et al., 2016]	Expected potential outcome	If linear regime	TR9>TR1≈EE4>TR3
[Laber and Zhao, 2015]	Expected potential outcome	If 2 treatments & tree-based regime	TR3>TR7
[Laber and Zhao, 2015]	Expected potential outcome	If 2 treatments & non tree-based regime	TR7>TR3
[Laber and Zhao, 2015]	Expected potential outcome	If continuous treatment	TR3>TR7
[Xu et al., 2015]	Expected potential outcome	-	TR4>TR10>EE4≈SE8≈TR9≈TR14
[Zhang et al., 2015]	Expected potential outcome	If decision list regime	TR5>TR7≈SE4>TR10
[Zhang et al., 2015]	Expected potential outcome	If linear regime	TR7≈SE4>TR5≈TR10
[Zhang et al., 2015]	Expected potential outcome	If non-linear regime	TR5>TR7≈SE4>TR10
[Zhao et al., 2015]	Expected potential outcome	If correct model for survival time	TR7>TR6
[Zhao et al., 2015]	Expected potential outcome	If incorrect model for survival time	TR6>TR7
[Huang and Fong, 2014]	Expected potential outcome	If linear regime	SE4>TR11b[2]>TR8a[3]>TR11a>TR8b>TR10b[4]>TR10a
[Huang and Fong, 2014]	Expected potential outcome	If linear regime & outliers	TR8a>TR11b>TR11a>TR8b>TR10b>TR7>TR10a
[Huang and Fong, 2014]	Expected potential outcome	If non-linear regime	TR11b>TR8b≈TR10b>TR8a≈TR11a>TR7>TR10a
[Huang and Fong, 2014]	Expected potential outcome	If non-linear regime & outliers	TR8b≈TR10b>TR11b>TR11a>TR8a>TR7>TR10a
[Huang and Fong, 2014]	Expected potential outcome	If highly non-linear regime	TR8b≈TR10b>TR11b>TR11a>TR8a>TR10a>TR7
[Foster et al., 2014]	Expected potential outcome	-	EE1>SE8>EE4
[Zhang et al., 2012b]	Expected potential outcome	If tree-based regime & sample size 200	TR12b[5]>TR11b>TR7>TR12a
[Zhang et al., 2012b]	Expected potential outcome	If tree-based regime & sample size 500	TR12b >TR11b>TR12a>TR7
[Zhang et al., 2012b]	Expected potential outcome	If tree-based regime & sample size 1000	TR12b>TR12a>TR11b>TR7
[Zhang et al., 2012b]	Expected potential outcome	If tree-based regime & incorrect prop. model	TR12b>TR11b>TR7>TR12a
[Zhang et al., 2012b]	Expected potential outcome	If linear regime & sample size $\leq$ 500	TR11b>TR7>TR12b>TR12a
[Zhang et al., 2012b]	Expected potential outcome	If linear regime & sample size $>$ 500	TR11b>TR12b>TR7>TR12a
[Zhang et al., 2012a]	Expected potential outcome	If linear regime & correct model	TR7≈TR11b>TR11a
[Zhang et al., 2012a]	Expected potential outcome	If linear regime & incorrect model	TR11b>TR11a>TR7
[Zhang et al., 2012a]	Expected potential outcome	If non-linear regime	TR11b>TR11a>TR7
[Qian and Murphy, 2011]	Expected potential outcome	If correctly modelled treatment effect	TR14≈TR7
[Qian and Murphy, 2011]	Expected potential outcome	If incorrectly modelled treatment effect	TR14>TR7
[Zhao et al., 2012]	MSE of expected potential outcome	If sample size $<$ 200	TR10>TR14>TR7
[Zhao et al., 2012]	MSE of expected potential outcome	If sample size $\geq$ 200	TR10≈TR14≈TR7
[Loh et al., 2015]	Accuracy estimated effect size	If only predictive covariates	SE10>SE2≈EE4>SE1>EE3≈SE8

Study [1]	Outcome measure	Condition	Result
[Loh et al., 2015]	Accuracy estimated effect size	If predictive & prognostic covariates	SE1>EE3>EE4>SE8>SE2≈SE11
[Foster et al., 2011]	Accuracy estimated enhanced effect	-	EE4>TR7
[Foster et al., 2014]	Strength enhanced treatment effect	-	EE4 >SE8>EE1
[Tian et al., 2014]	Agreement estimated vs. true effect	-	SE4>TR7
	Recovery		
[Loh et al., 2015]	Correct split variables selected	If only predictive covariates	SE2≈SE10≈EE4>SE1≈SE8>EE3
[Loh et al., 2015]	Correct split variables selected	If predictive & prognostic covariates	SE1≈SE8>EE3>EE43>SE10≈SE2
[Foster et al., 2014]	Correct split variables selected	-	EE4>SE8>EE1
[Foster et al., 2011]	Correct split variables selected	-	EE4>TR7
[Tian et al., 2014]	Number of correctly selected covariates	-	SE4>TR7
[Kehl and Ulm, 2006]	Correct subgroups identified	If weak interaction	EE5>TR7
[Kehl and Ulm, 2006]	Correct subgroups identified	If strong interaction	EE5≈TR7
[Zhu et al., 2016]	Correct treatment assignment	If tree-based regime	TR1>EE4>TR9>TR3
[Zhu et al., 2016]	Correct treatment assignment	If linear regime	TR9>TR1>EE4>TR3
[Xu et al., 2015]	Correct treatment assignment	-	TR4≈TR10>EE4≈SE8≈TR9≈TR14
[Imai et al., 2013]	Correct treatment assignment	-	TR9>TR14
[Zhao et al., 2012]	MSE of misclassification rate	-	TR10>TR7>TR14
[Imai et al., 2013]	False discovery rate largest effect	If sample size $\leq$ 1000	TR9>TR14
[Imai et al., 2013]	False discovery rate largest effect	If sample size 5000	TR14>TR9
[Imai et al., 2013]	False discovery rate four largest effects	-	TR9>TR14
[Imai et al., 2013]	Discovery rate largest effect	-	TR14>TR9
[Imai et al., 2013]	Discovery rate four largest effects	If sample size $\leq$ 1000	TR14>TR9
[Imai et al., 2013]	Discovery rate four largest effects	If sample size 5000	TR9>TR14
	Inferential errors		
[Loh et al., 2015]	Type I error	-	SE8>SE1>EE3>SE2≈SE10≈EE4
[Foster et al., 2011]	Type I error	-	TR7>EE4
[Loh et al., 2015]	Type II error	-	SE1≈EE4≈EE3≈SE2≈SE10>SE8
[Foster et al., 2011]	Type II error	-	EE4>TR7
[Kehl and Ulm, 2006]	Type II error	If weak or medium interaction	EE5>TR7
[Kehl and Ulm, 2006]	Type II error	If strong interaction	EE5≈TR7

[1] The abbreviations were introduced in Table 1.
[2] TR11a refers to the method with the maximization of the expected potential outcome based on the IPWE, TR11b refers to the method with the maximization of the expected potential outcome based on the AIPWE.
[3] TR8a refers to the method with linear kernel, TR8b to the method with Radial Basis Function (RBF) kernel.
[4] TR10a refers to the method with linear kernel, TR10b to the method with RBF kernel.
[5] TR12a refers to the method with estimation of the contrasts based on the IPWE, TR12b to estimation of the contrasts based on the AIPWE.

Table 2: Overview of available software for the methods for detecting subgroups involved in treatment-subgroup interactions from the data in a post-hoc manner. The first column indicates the method, the second column indicates the name of the package, the third column the type of software or code, and the fourth column where it can be found. The fifth column indicates whether the software or code is documented or not, and the last column indicates whether the authors that proposed the method where involved in creating the software or code.

Method [1]	Name package	Form [2]	Location	Documentation publicly available?	Authors method involved?
EE3	SIDES	dedicated R-package	CRAN	Yes	No
	SIDESxl	excel add-in	BioPharmNet	No	Yes
EE4	–	R-code	BioPharmNet	No	Yes
SE1	GUIDE	executable	author's page [3]	Yes	Yes
SE2	GUIDE	executable	author's page [3]	Yes	Yes
SE5	QUINT	dedicated R-package	CRAN	Yes	Yes
SE7	STIMA	dedicated R-package	author's page [4]	Yes	Yes
SE8	–	R-code	BioPharmNet	No	Yes
SE10	PARTY	broad R-package	CRAN	Yes	Yes
SE13	DSBayes	R-package	CRAN	Yes	No
TR1	–	R-code	supp. mat.	No	Yes
TR2	–	R-code	appendix	No	Yes
TR4	–	R-code	supp. mat.	No	Yes
TR5	–	R-code	author's page [5]	No	Yes
TR7	DTRlearn;	broad R-package	CRAN	Yes	Yes
	DynTxRegime	broad R-package	CRAN	Yes	Yes
TR8	–	R-code	supp. mat.	No	Yes
TR9	FindIt	dedicated R-package	CRAN	Yes	Yes
TR10	DTRlearn;	broad R-package	CRAN	Yes	Yes
	DynTxRegime	broad R-package	CRAN	Yes	No
TR11	DynTxRegime	broad R-package	CRAN	Yes	Yes
TR12	DynTxRegime	broad R-package	CRAN	Yes	Yes
	e.g., RPART [6]	general R-package	CRAN	Yes	No
TR13	LARS [7]	general R-package	CRAN	Yes	No

Note. A dash means that the software or source code does not have a name.

[1] The abbreviations were introduced in Table 1.

[2] *Dedicated R-package*: An R-package that is only dedicated to the method under study.
Excel add-in: Software that can be ran in Microsoft Excel.
R-code: R-code that has not been organized into a package.
Executable: A program that can be ran by a computer, but for which the source cannot be read by the user.
Broad R-package: An R-package that includes both functions related to the method under study and functions not related to the method under study.
General R-package: An R-package that was developed for another (broader) purpose, that can be used (with some adjustments or preprocessing steps) to apply the method under study.

[3] http://www.stat.wisc.edu/~loh/guide.html

[4] http://www.elisedusseldorp.nl/Index.php?type=6

[5] http://www4.ncsu.edu/~yzhang52/

[6] Several existing general packages may be used to minimize the weighted misclassification error, depending on the methodology one wants to use for it.

[7] Examples for how to apply the method using lars included in supplementary materials

References

[Foster et al., 2014] Foster, J., Taylor, J., Kaciroti, N., and Nan, B. (2014). Simple subgroup approximations to optimal treatment regimes from randomized clinical trial data. *Biostatistics*, 16(2):368–382.

[Foster et al., 2011] Foster, J., Taylor, J., and Ruberg, S. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24):2867–2880.

[Huang and Fong, 2014] Huang, Y. and Fong, Y. (2014). Identifying optimal biomarker combinations for treatment selection via a robust kernel method. *Biometrics*, 70(4):891–901.

[Imai et al., 2013] Imai, K., Ratkovic, M., et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

[Kehl and Ulm, 2006] Kehl, V. and Ulm, K. (2006). Responder identification in clinical trials with censored data. *Computational Statistics & Data Analysis*, 50(5):1338–1355.

[Laber and Zhao, 2015] Laber, E. and Zhao, Y. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

[Loh et al., 2015] Loh, W., He, X., and Man, M. (2015). A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in Medicine*, 34(11):1818–1833.

[Qian and Murphy, 2011] Qian, M. and Murphy, S. (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics*, 39(2):1180.

[Tian et al., 2014] Tian, L., Alizadeh, A., Gentles, A., and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532.

[Xu et al., 2015] Xu, Y., Yu, M., Zhao, Y., Li, Q., Wang, S., and Shao, J. (2015). Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics*, 71(3):645–653.

[Zhang et al., 2012a] Zhang, B., Tsiatis, A., Davidian, M., Zhang, M., and Laber, E. (2012a). Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114.

[Zhang et al., 2012b] Zhang, B., Tsiatis, A., Laber, E., and Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

[Zhang et al., 2015] Zhang, Y., Laber, E., Tsiatis, A., and Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904.

[Zhao et al., 2015] Zhao, Y., Zeng, D., Laber, E., Song, R., Yuan, M., and Kosorok, M. (2015). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168.

[Zhao et al., 2012] Zhao, Y., Zeng, D., Rush, A., and Kosorok, M. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

[Zhu et al., 2016] Zhu, R., Zhao, Y., Chen, G., Ma, S., and Zhao, H. (2016). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*.