WEB MATERIAL

Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study

Til Stürmer, Michael Webster-Clark, Jennifer L. Lund, Richard Wyss, Alan R. Ellis, Mark Lunt, Kenneth J. Rothman, and Robert J. Glynn

Table of Contents:

Web Tables 1–6

Web Figures 1 and 2

Web Table 1. Mean incidence rate ratios, empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies without unmeasured confounding according to treatment prevalence[a]

	Treatment prevalence = 20%				Treatment prevalence = 50%				Treatment prevalence = 80%			
	RR	Var[b]	MSE[d]	RE[e]	RR	Var[b]	MSE[c]	RE[d]	RR	Var[b]	MSE[c]	RE[d]
Crude	1.77	0.0049	0.3300	**127**	1.83	0.0037	0.3695	**129**	2.05	0.0065	0.5188	**205**
True outcome model	1.00	0.0048	0.0049	**130**	1.00	0.0037	0.0037	**130**	1.00	0.0058	0.0058	**228**
IPTW[a]												
No trimming	1.00	0.0063	0.0063	**100**	1.00	0.0048	0.0048	**100**	1.01	0.0133	0.0133	**100**
Common range[e]	0.99	0.0062	0.0064	**101**	0.99	0.0047	0.0048	**101**	0.98	0.0128	0.0130	**104**
Stürmer[e]	0.99	0.0080	0.0081	**78**	1.00	0.0051	0.0051	**95**	1.00	0.0095	0.0095	**140**
Walker[e]	0.99	0.0097	0.0098	**65**	1.00	0.0060	0.0060	**79**	1.01	0.0111	0.0111	**120**
Crump[e]	1.00	0.0056	0.0056	**112**	1.00	0.0043	0.0043	**111**	1.01	0.0105	0.0105	**126**
SMR-treated weights												
No trimming	1.00	0.0063	0.0063	**100**	1.00	0.0065	0.0065	**74**	1.01	0.0150	0.0151	**88**
Common range	0.99	0.0062	0.0062	**102**	0.99	0.0063	0.0064	**76**	0.98	0.0145	0.0147	**92**
Stürmer	1.00	0.0075	0.0075	**84**	1.00	0.0054	0.0054	**88**	1.01	0.0101	0.0101	**132**
Walker	0.99	0.0092	0.0092	**69**	1.00	0.0064	0.0064	**75**	1.01	0.0116	0.0116	**114**
Crump	1.00	0.0063	0.0063	**99**	1.00	0.0053	0.0053	**91**	1.01	0.0113	0.0114	**117**
SMR-untreated weights												
No trimming	0.99	0.0080	0.0081	**78**	1.00	0.0047	0.0047	**102**	1.00	0.0065	0.0065	**205**
Common range	0.99	0.0079	0.0081	**80**	1.00	0.0047	0.0047	**103**	1.00	0.0065	0.0065	**206**
Stürmer	0.99	0.0088	0.0088	**72**	1.00	0.0054	0.0054	**89**	1.00	0.0082	0.0082	**161**
Walker	0.99	0.0104	0.0104	**61**	1.00	0.0063	0.0063	**76**	1.00	0.0103	0.0103	**129**
Crump	1.00	0.0064	0.0065	**98**	1.00	0.0046	0.0046	**104**	1.00	0.0100	0.0100	**133**
Overlap weights												
No trimming	1.00	0.0051	0.0051	**123**	1.00	0.0039	0.0039	**124**	1.00	0.0063	0.0063	**212**
Common range	1.00	0.0051	0.0051	**123**	1.00	0.0039	0.0039	**124**	1.00	0.0063	0.0063	**212**
Stürmer	1.00	0.0073	0.0074	**86**	1.00	0.0050	0.0050	**96**	1.00	0.0081	0.0081	**163**
Walker	0.99	0.0091	0.0091	**69**	1.00	0.0060	0.0060	**80**	1.00	0.0102	0.0102	**130**
Crump	1.00	0.0052	0.0052	**120**	1.00	0.0039	0.0039	**122**	1.00	0.0095	0.0095	**140**
Matching weights												
No trimming	1.00	0.0053	0.0053	**118**	1.00	0.0039	0.0039	**123**	1.00	0.0063	0.0063	**211**
Common range	1.00	0.0053	0.0053	**118**	1.00	0.0039	0.0039	**123**	1.00	0.0063	0.0063	**211**

Stürmer	1.00	0.0075	0.0075	**84**	1.00	0.0051	0.0051	**95**	1.00	0.0082	0.0082	**161**
Walker	0.99	0.0092	0.0092	**69**	1.00	0.0061	0.0061	**79**	1.00	0.0103	0.0103	**129**
Crump	1.00	0.0054	0.0054	**116**	1.00	0.0040	0.0040	**121**	1.00	0.0097	0.0097	**137**
Entropy weights												
No trimming	1.00	0.0051	0.0051	**123**	1.00	0.0039	0.0039	**122**	1.00	0.0066	0.0066	**202**
Common range	1.00	0.0051	0.0051	**123**	1.00	0.0039	0.0039	**122**	1.00	0.0066	0.0066	**202**
Stürmer	1.00	0.0074	0.0074	**85**	1.00	0.0050	0.0050	**96**	1.00	0.0082	0.0083	**161**
Walker	0.99	0.0091	0.0091	**69**	1.00	0.0060	0.0060	**80**	1.00	0.0103	0.0103	**129**
Crump	1.00	0.0053	0.0053	**120**	1.00	0.0040	0.0040	**121**	1.00	0.0096	0.0096	**138**

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: area under the receiver operating characteristics curve (c-statistic)

[a] True RR=1.0; AUC=0.75

[b] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[c] Mean squared error; mean of $[\log(RR) - \log(1.0)]^2$ over 5,000 simulated studies

[d] Relative efficiency; mean variance of untrimmed IPTW divided by mean variance of scenario times 100 – values above 100 indicate better precision of estimates

[e] No trimming: c-statistic=0.748, 0.750, and 0.765 for treatment prevalence=20, 50, and 80%, respectively; Common range: restricting to positivity (c-statistic=0.744, 0.747, and 0.762 for treatment prevalence=20, 50, and 80%, respectively); Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated (c-statistic=0.657, 0.659, and 0.671 for treatment prevalence=20, 50, and 80%, respectively); Walker: preference score trimming below 0.3 and above 0.7 (c-statistic=0.630, 0.630, and 0.632 for treatment prevalence=20, 50, and 80%, respectively); and Crump: propensity score trimming below 0.1 and above 0.9 (c-statistic=0.684, 0.733, and 0.713 for treatment prevalence=20, 50, and 80%, respectively)

Web Table 2. Mean incidence rate ratios, empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies without unmeasured confounding according to treatment prevalence[a]

	Treatment prevalence = 20%				Treatment prevalence = 50%				Treatment prevalence = 80%			
	RR	Var[c]	MSE[d]	RE[e]	RR	Var[c]	MSE[d]	RE[e]	RR	Var[c]	MSE[d]	RE[e]
Crude	1.92	0.0049	0.4295	**333**	1.94	0.0037	0.4455	**388**	2.20	0.0069	0.6286	**739**
True outcome model	1.00	0.0060	0.0060	**271**	1.00	0.0045	0.0045	**317**	1.00	0.0070	0.0070	**731**
IPTW[a]												
No trimming[b]	0.99	0.0162	0.0163	**100**	1.00	0.0142	0.0142	**100**	1.04	0.0509	0.0522	**100**
Common range	0.96	0.0157	0.0170	**104**	0.98	0.0138	0.0144	**103**	0.98	0.0487	0.0491	**105**
Stürmer	0.99	0.0120	0.0121	**135**	1.00	0.0069	0.0069	**205**	1.01	0.0136	0.0136	**375**
Walker	0.99	0.0161	0.0161	**101**	1.00	0.0095	0.0095	**149**	1.01	0.0171	0.0171	**298**
Crump	1.00	0.0076	0.0076	**214**	1.00	0.0059	0.0059	**241**	1.00	0.0124	0.0124	**409**
SMR-treated weights												
No trimming	1.00	0.0154	0.0154	**105**	1.01	0.0244	0.0244	**58**	1.04	0.0599	0.0617	**85**
Common range	0.99	0.0150	0.0152	**108**	0.98	0.0236	0.0239	**60**	0.98	0.0576	0.0578	**88**
Stürmer	1.00	0.0106	0.0106	**154**	1.00	0.0079	0.0079	**181**	1.01	0.0152	0.0152	**336**
Walker	1.00	0.0152	0.0152	**107**	1.00	0.0102	0.0102	**139**	1.01	0.0181	0.0181	**282**
Crump	1.00	0.0103	0.0103	**157**	1.00	0.0078	0.0078	**181**	1.00	0.0146	0.0146	**350**
SMR-untreated weights												
No trimming	0.99	0.0260	0.0261	**63**	1.00	0.0100	0.0100	**142**	1.00	0.0101	0.0101	**505**
Common range	0.96	0.0248	0.0265	**66**	0.98	0.0097	0.0100	**146**	0.99	0.0099	0.0100	**513**
Stürmer	0.99	0.0139	0.0140	**117**	1.00	0.0078	0.0078	**183**	1.00	0.0114	0.0114	**449**
Walker	0.99	0.0172	0.0173	**94**	1.00	0.0100	0.0100	**143**	1.01	0.0159	0.0159	**320**
Crump	1.00	0.0094	0.0094	**173**	1.00	0.0071	0.0071	**200**	1.00	0.0131	0.0131	**389**
Overlap weights												
No trimming	1.00	0.0065	0.0065	**249**	1.00	0.0049	0.0049	**292**	1.00	0.0079	0.0079	**644**
Common range	1.00	0.0065	0.0065	**249**	1.00	0.0049	0.0049	**292**	1.00	0.0079	0.0079	**644**
Stürmer	1.00	0.0102	0.0102	**160**	1.00	0.0068	0.0068	**208**	1.00	0.0111	0.0111	**459**
Walker	0.99	0.0149	0.0149	**109**	1.00	0.0095	0.0095	**149**	1.01	0.0157	0.0157	**324**
Crump	1.00	0.0070	0.0070	**231**	1.00	0.0053	0.0053	**268**	1.00	0.0111	0.0111	**460**
Matching weights												
No trimming	1.00	0.0068	0.0068	**240**	1.00	0.0050	0.0050	**284**	1.00	0.0081	0.0081	**629**
Common range	1.00	0.0068	0.0068	**240**	1.00	0.0050	0.0050	**284**	1.00	0.0081	0.0081	**630**

Stürmer	1.00	0.0106	0.0106	**154**	1.00	0.0070	0.0070	**204**	1.00	0.0113	0.0114	**449**
Walker	1.00	0.0152	0.0152	**107**	1.00	0.0096	0.0096	**147**	1.01	0.0159	0.0159	**320**
Crump	1.00	0.0073	0.0073	**223**	1.00	0.0055	0.0054	**261**	1.00	0.0113	0.0113	**449**
Entropy weights												
No trimming	1.00	0.0066	0.0066	**246**	1.00	0.0050	0.0050	**285**	1.00	0.0085	0.0085	**598**
Common range	1.00	0.0066	0.0066	**245**	1.00	0.0050	0.0050	**285**	1.00	0.0085	0.0085	**599**
Stürmer	1.00	0.0102	0.0102	**159**	1.00	0.0069	0.0069	**207**	1.00	0.0113	0.0113	**453**
Walker	0.99	0.0149	0.0150	**109**	1.00	0.0095	0.0095	**149**	1.01	0.0158	0.0158	**323**
Crump	1.00	0.0070	0.0070	**231**	1.00	0.0053	0.0053	**266**	1.00	0.0112	0.0112	**456**

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: Area under the receiver operating characteristics curve (c-statistic)

[a] True RR=1.0; AUC=0.85

[b] Common range: restricting to positivity; Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated; Walker: preference score trimming below 0.3 and above 0.7; and Crump: propensity score trimming below 0.1 and above 0.9

[c] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[d] Mean squared error; mean of $[\log(RR) - \log(1.0)]^2$ over 5,000 simulated studies

[e] Relative efficiency; mean variance of untrimmed IPTW divided by mean variance of scenario times 100 – values above 100 indicate better precision of estimates

Web Table 3. Mean incidence rate ratios, empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies without unmeasured confounding according to treatment prevalence[a]

	Treatment prevalence = 20%				Treatment prevalence = 50%				Treatment prevalence = 80%			
	RR	Var[b]	MSE[c]	RE[d]	RR	Var[b]	MSE[c]	RE[d]	RR	Var[b]	MSE[c]	RE[d]
Crude	1.33	0.0055	0.0855	**98**	1.34	0.0043	0.0914	**98**	1.38	0.0079	0.1102	**115**
True outcome model	1.00	0.0049	0.0049	**109**	1.00	0.0039	0.0039	**107**	1.00	0.0071	0.0071	**128**
IPTW												
No trimming	1.00	0.0054	0.0054	**100**	1.00	0.0042	0.0042	**100**	1.00	0.0091	0.0091	**100**
Common range[e]	1.00	0.0054	0.0054	**100**	1.00	0.0042	0.0042	**100**	1.00	0.0090	0.0090	**101**
Stürmer[e]	1.00	0.0067	0.0067	**80**	1.00	0.0049	0.0049	**86**	1.00	0.0094	0.0094	**97**
Walker[e]	1.00	0.0063	0.0063	**85**	1.00	0.0046	0.0046	**91**	1.00	0.0090	0.0090	**101**
Crump[e]	1.00	0.0054	0.0054	**100**	1.00	0.0042	0.0042	**100**	1.00	0.0091	0.0091	**99**
SMR-treated weights												
No trimming	1.00	0.0054	0.0054	**100**	1.00	0.0047	0.0047	**89**	1.00	0.0097	0.0097	**93**
Common range	1.00	0.0054	0.0054	**100**	1.00	0.0047	0.0047	**89**	1.00	0.0096	0.0097	**94**
Stürmer	1.00	0.0066	0.0066	**81**	1.00	0.0051	0.0051	**82**	1.00	0.0097	0.0097	**94**
Walker	1.00	0.0062	0.0062	**86**	1.00	0.0048	0.0048	**87**	1.00	0.0093	0.0093	**97**
Crump	1.00	0.0054	0.0054	**99**	1.00	0.0047	0.0047	**89**	1.00	0.0095	0.0095	**96**
SMR-untreated weights												
No trimming	1.00	0.0057	0.0057	**94**	1.00	0.0041	0.0041	**101**	1.00	0.0073	0.0073	**124**
Common range	1.00	0.0057	0.0057	**94**	1.00	0.0041	0.0041	**101**	1.00	0.0073	0.0073	**124**
Stürmer	1.00	0.0069	0.0069	**78**	1.00	0.0049	0.0049	**85**	1.00	0.0087	0.0087	**105**
Walker	1.00	0.0065	0.0066	**82**	1.00	0.0046	0.0046	**90**	1.00	0.0082	0.0082	**111**
Crump	1.00	0.0057	0.0057	**94**	1.00	0.0041	0.0041	**101**	1.00	0.0083	0.0083	**109**
Overlap weights												
No trimming	1.00	0.0052	0.0052	**104**	1.00	0.0041	0.0041	**103**	1.00	0.0074	0.0074	**123**
Common range	1.00	0.0052	0.0052	**104**	1.00	0.0041	0.0041	**103**	1.00	0.0074	0.0074	**123**
Stürmer	1.00	0.0065	0.0066	**82**	1.00	0.0048	0.0048	**86**	1.00	0.0087	0.0087	**104**
Walker	1.00	0.0061	0.0061	**88**	1.00	0.0045	0.0045	**92**	1.00	0.0082	0.0082	**110**
Crump	1.00	0.0052	0.0052	**103**	1.00	0.0041	0.0041	**103**	1.00	0.0084	0.0084	**108**
Matching weights												
No trimming	1.00	0.0053	0.0053	**101**	1.00	0.0041	0.0041	**102**	1.00	0.0073	0.0073	**124**
Common range	1.00	0.0053	0.0053	**101**	1.00	0.0041	0.0041	**102**	1.00	0.0073	0.0073	**124**

Stürmer	1.00	0.0066	0.0066	**81**	1.00	0.0049	0.0049	**86**	1.00	0.0087	0.0087	**105**
Walker	1.00	0.0062	0.0062	**86**	1.00	0.0046	0.0046	**91**	1.00	0.0082	0.0082	**111**
Crump	1.00	0.0054	0.0054	**100**	1.00	0.0041	0.0041	**102**	1.00	0.0083	0.0083	**109**
Entropy weights												
No trimming	1.00	0.0051	0.0051	**104**	1.00	0.0041	0.0041	**102**	1.00	0.008	0.008	**120**
Common range	1.00	0.0051	0.0051	**104**	1.00	0.0041	0.0041	**102**	1.00	0.008	0.008	**120**
Stürmer	1.00	0.0065	0.0065	**82**	1.00	0.0048	0.0048	**86**	1.00	0.009	0.009	**103**
Walker	1.00	0.0061	0.0061	**88**	1.00	0.0045	0.0045	**92**	1.00	0.008	0.008	**109**
Crump	1.00	0.0052	0.0052	**103**	1.00	0.0041	0.0041	**102**	1.00	0.008	0.008	**107**

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: Area under the receiver operating characteristics curve (c-statistic)

[a] True RR=1.0; AUC=0.65

[b] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[c] Mean squared error; mean of $[\log(RR) - \log(1.0)]^2$ over 5,000 simulated studies

[d] Relative efficiency; mean variance of untrimmed IPTW divided by mean variance of scenario times 100 – values above 100 indicate better precision of estimates

[e] Common range: restricting to positivity; Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated; Walker: preference score trimming below 0.3 and above 0.7; and Crump: propensity score trimming below 0.1 and above 0.9

Web Table 4. Mean incidence rate ratios (RR), empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies with unmeasured confounding leading to treatment contrary to prediction according unmeasured confounding pattern[a]

	Last Resort Treatment				Treatment Withheld				Last Resort and Withheld			
	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]
Crude	3.60	0.0031	0.3484	0	2.95	0.0068	0.1579	0	2.91	0.0067	0.1465	0
Residual confounding	2.06	0.0030	0.0037	92	1.36	0.0065	0.1548	0	1.40	0.0065	0.1352	0
True outcome model	2.00	0.0028	0.0028	95	2.00	0.0032	0.0032	95	2.00	0.0032	0.0032	95
IPTW												
No trimming	2.09	0.0041	0.0061	90	1.08	0.0167	0.4019	0	1.15	0.0154	0.3198	0
Common range[e]	2.07	0.0040	0.0054	92	1.06	0.0162	0.4145	0	1.14	0.0150	0.3304	0
Stürmer[e]	2.00	0.0039	0.0039	96	1.97	0.0061	0.0064	95	1.97	0.0060	0.0062	95
Walker[e]	2.00	0.0045	0.0045	96	2.00	0.0068	0.0068	96	2.00	0.0065	0.0065	96
Crump[e]	2.07	0.0037	0.0050	93	1.20	0.0112	0.2761	0	1.23	0.0113	0.2467	0
SMR-treated weights												
No trimming	2.06	0.0056	0.0064	93	0.95	0.0226	0.5807	0	1.00	0.0212	0.4987	0
Common range	2.04	0.0055	0.0058	94	0.94	0.0218	0.5997	0	0.99	0.0204	0.5172	0
Stürmer	2.00	0.0044	0.0044	96	1.96	0.0071	0.0075	95	1.96	0.0070	0.0074	95
Walker	2.00	0.0049	0.0049	96	2.00	0.0074	0.0074	96	2.00	0.0071	0.0071	96
Crump	2.04	0.0047	0.0050	95	1.07	0.0147	0.4063	0	1.08	0.0150	0.3925	0
SMR-untreated weights												
No trimming	2.16	0.0041	0.0100	81	1.49	0.0061	0.0911	5	1.61	0.0066	0.0536	28
Common range	2.16	0.0041	0.0097	82	1.49	0.0061	0.0937	4	1.61	0.0066	0.0549	27
Stürmer	2.00	0.0040	0.0040	96	1.98	0.0058	0.0059	96	1.98	0.0058	0.0058	96
Walker	2.00	0.0045	0.0045	96	2.00	0.0069	0.0069	96	2.00	0.0066	0.0066	96
Crump	2.14	0.0038	0.0081	84	1.54	0.0059	0.0730	10	1.64	0.0063	0.0467	32
Overlap weights												
No trimming	2.07	0.0031	0.0044	92	1.31	0.0077	0.1847	0	1.37	0.0076	0.1501	1
Common range	2.07	0.0031	0.0043	93	1.31	0.0076	0.1858	0	1.37	0.0075	0.1515	1
Stürmer	2.00	0.0038	0.0038	96	1.98	0.0059	0.0060	95	1.98	0.0058	0.0059	95
Walker	2.00	0.0045	0.0045	96	2.00	0.0068	0.0068	96	2.00	0.0064	0.0064	96
Crump	2.06	0.0032	0.0039	94	1.38	0.0074	0.1442	1	1.41	0.0073	0.1289	2
Matching weights												
No trimming	2.06	0.0031	0.0040	93	1.42	0.0063	0.1238	1	1.47	0.0062	0.0999	3

Common range	2.06	0.0031	0.0040	93	1.42	0.0063	0.1243	1	1.47	0.0062	0.1006	3
Stürmer	2.00	0.0039	0.0039	96	1.98	0.0058	0.0059	96	1.98	0.0057	0.0058	95
Walker	2.00	0.0045	0.0045	96	2.00	0.0068	0.0068	96	2.00	0.0065	0.0065	96
Crump	2.05	0.0031	0.0037	94	1.48	0.0062	0.0975	3	1.51	0.0062	0.0866	5
Entropy weights												
No trimming	2.08	0.0032	0.0047	92	1.25	0.0089	0.2270	0	1.32	0.0087	0.1833	1
Common range	2.07	0.0032	0.0045	92	1.25	0.0088	0.2292	0	1.31	0.0086	0.1858	1
Stürmer	2.00	0.0039	0.0039	96	1.97	0.0059	0.0061	95	1.98	0.0058	0.0060	95
Walker	2.00	0.0045	0.0045	96	2.00	0.0068	0.0068	96	2.00	0.0065	0.0065	96
Crump	2.06	0.0032	0.0041	94	1.33	0.0081	0.1725	1	1.36	0.0081	0.1541	1

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: Area under the receiver operating characteristics curve (c-statistic)

[a] True RR=2.0; AUC=0.75; treatment prevalence = 50%

[b] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[c] Mean squared error; mean of $[\log(RR) - \log(2.0)]^2$ over 5,000 simulated studies

[d] Percent of simulated studies in which the 95% confidence interval (CI) includes the true value (RR=2.0)

[e] Common range: restricting to positivity; Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated; Walker: preference score trimming below 0.3 and above 0.7; and Crump: propensity score trimming below 0.1 and above 0.9

Web Table 5. Mean incidence rate ratios, empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies with unmeasured confounding leading to treatment contrary to prediction according to weighting and trimming method[a]

	Last Resort Treatment				Treatment Withheld				Last Resort and Withheld			
	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]
Crude	3.97	0.0062	0.4761	0	3.07	0.0137	0.1976	0	2.98	0.0141	0.1738	1
Residual confounding	2.06	0.0054	0.0063	94	1.35	0.0121	0.1637	2	1.39	0.0127	0.1457	4
True outcome model	2.01	0.0053	0.0053	95	2.01	0.0057	0.0057	95	2.01	0.0059	0.0060	95
IPTW												
No trimming	2.12	0.0118	0.0149	90	0.98	0.0421	0.5569	2	1.06	0.0396	0.4423	5
Common range[e]	2.07	0.0113	0.0126	93	0.95	0.0404	0.5886	2	1.04	0.0381	0.4712	4
Stürmer[e]	2.01	0.0091	0.0091	95	1.70	0.0202	0.0459	78	1.69	0.0209	0.0494	77
Walker[e]	2.01	0.0101	0.0101	96	2.01	0.0164	0.0164	96	1.97	0.0171	0.0173	95
Crump[e]	2.12	0.0092	0.0127	93	2.01	0.0150	0.0151	95	2.11	0.0151	0.0180	94
SMR-treated weights												
No trimming	2.11	0.0133	0.0160	91	0.95	0.0455	0.6057	2	1.02	0.0430	0.4927	4
Common range	2.06	0.0128	0.0136	94	0.92	0.0436	0.6418	1	1.00	0.0413	0.5266	3
Stürmer	2.01	0.0097	0.0098	95	1.68	0.0220	0.0517	76	1.67	0.0228	0.0557	76
Walker	2.01	0.0107	0.0107	95	2.01	0.0173	0.0173	95	1.97	0.0181	0.0184	95
Crump	2.09	0.0099	0.0119	94	2.01	0.0163	0.0163	95	2.08	0.0161	0.0177	95
SMR-untreated weights												
No trimming	2.19	0.0061	0.0147	83	1.40	0.0123	0.1385	12	1.53	0.0133	0.0863	35
Common range	2.19	0.0061	0.0142	84	1.40	0.0123	0.1412	12	1.52	0.0133	0.0881	34
Stürmer	2.01	0.0074	0.0074	96	1.84	0.0128	0.0200	88	1.83	0.0131	0.0209	88
Walker	2.01	0.0088	0.0088	96	2.01	0.0145	0.0145	95	1.99	0.0145	0.0146	95
Crump	2.25	0.0087	0.0229	80	2.01	0.0132	0.0132	96	2.24	0.0144	0.0276	86
Overlap weights												
No trimming	2.14	0.0059	0.0104	89	1.33	0.0139	0.1806	7	1.42	0.0144	0.1304	19
Common range	2.13	0.0059	0.0100	90	1.33	0.0138	0.1824	6	1.42	0.0144	0.1327	18
Stürmer	2.01	0.0075	0.0075	96	1.82	0.0135	0.0227	86	1.81	0.0138	0.0240	86
Walker	2.01	0.0089	0.0089	96	2.01	0.0145	0.0146	95	1.99	0.0146	0.0147	95
Crump	2.17	0.0082	0.0145	89	2.01	0.0131	0.0132	95	2.16	0.0135	0.0193	91
Matching weights												
No trimming	2.17	0.0058	0.0123	86	1.39	0.0125	0.1441	11	1.51	0.0131	0.0935	30

Common range	2.16	0.0058	0.0121	87	1.39	0.0124	0.1451	11	1.50	0.0130	0.0949	30
Stürmer	2.01	0.0074	0.0074	96	1.84	0.0128	0.0200	88	1.83	0.0131	0.0209	88
Walker	2.01	0.0088	0.0088	96	2.01	0.0145	0.0145	95	1.99	0.0145	0.0146	95
Crump	2.20	0.0082	0.0177	84	2.01	0.0131	0.0132	96	2.20	0.0136	0.0226	89
Entropy weights												
No trimming	2.13	0.0062	0.0102	90	1.24	0.0171	0.2462	4	1.33	0.0175	0.1851	11
Common range	2.12	0.0062	0.0096	91	1.23	0.0169	0.2500	4	1.32	0.0173	0.1895	10
Stürmer	2.01	0.0077	0.0077	96	1.79	0.0146	0.0269	84	1.78	0.0149	0.0285	84
Walker	2.01	0.0090	0.0090	96	2.01	0.0148	0.0148	95	1.98	0.0149	0.0150	95
Crump	2.15	0.0083	0.0138	90	2.01	0.0134	0.0134	95	2.14	0.0137	0.0186	92

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: Area under the receiver operating characteristics curve (c-statistic)

[a] True RR=2.0; AUC=0.75; treatment prevalence = 80%

[b] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[c] Mean squared error; mean of $[\log(RR) - \log(2.0)]^2$ over 5,000 simulated studies

[d] Percent of simulated studies in which the 95% confidence interval (CI) includes the true value (RR=2.0)

[e] Common range: restricting to positivity; Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated; Walker: preference score trimming below 0.3 and above 0.7; and Crump: propensity score trimming below 0.1 and above 0.9

Web Table 6. Mean incidence rate ratios, empirical variance, mean square error, and percent coverage of 95% confidence intervals from 5,000 simulated studies with unmeasured confounding leading to treatment withheld according to c-statistic of propensity score model[a]

	AUC = 0.65				AUC = 0.75 (see table S4)				AUC = 0.85			
	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]	RR	Var[b]	MSE[c]	Cov[d]
Crude	2.16	0.0057	0.0120	62	2.95	0.0068	0.1579	0	4.03	0.0064	0.4989	0
Residual confounding	1.52	0.0051	0.0801	1	1.36	0.0065	0.1548	0	1.54	0.0076	0.0762	7
True outcome model	2.00	0.0032	0.0032	95	2.00	0.0032	0.0032	95	2.00	0.0045	0.0045	95
IPTW												
No trimming	1.40	0.0075	0.1363	1	1.08	0.0167	0.4019	0	1.04	0.0635	0.4930	8
Common range[e]	1.39	0.0074	0.1377	1	1.06	0.0162	0.4145	0	1.00	0.0603	0.5384	5
Stürmer[e]	1.90	0.0052	0.0078	90	1.97	0.0061	0.0064	95	2.01	0.0085	0.0085	95
Walker[e]	1.77	0.0057	0.0212	62	2.00	0.0068	0.0068	96	2.00	0.0110	0.0110	96
Crump[e]	1.40	0.0075	0.1362	1	1.20	0.0112	0.2761	0	1.86	0.0106	0.0156	90
SMR-treated weights												
No trimming	1.29	0.0100	0.2029	0	0.95	0.0226	0.5807	0	0.90	0.0828	0.7181	6
Common range	1.28	0.0099	0.2055	0	0.94	0.0218	0.5997	0	0.87	0.0783	0.7798	4
Stürmer	1.88	0.0058	0.0098	88	1.96	0.0071	0.0075	95	2.01	0.0103	0.0103	95
Walker	1.72	0.0069	0.0297	54	2.00	0.0074	0.0074	96	2.01	0.0122	0.0122	95
Crump	1.29	0.0100	0.2028	0	1.07	0.0147	0.4063	0	1.80	0.0170	0.0276	87
SMR-untreated weights												
No trimming	1.59	0.0048	0.0588	10	1.49	0.0061	0.0911	5	1.78	0.0093	0.0234	77
Common range	1.58	0.0048	0.0592	10	1.49	0.0061	0.0937	4	1.75	0.0091	0.0268	72
Stürmer	1.93	0.0047	0.0062	92	1.98	0.0058	0.0059	96	2.00	0.0086	0.0086	95
Walker	1.83	0.0048	0.0127	77	2.00	0.0069	0.0069	96	2.00	0.0110	0.0110	96
Crump	1.59	0.0048	0.0588	10	1.54	0.0059	0.0730	10	1.97	0.0075	0.0078	95
Overlap weights												
No trimming	1.48	0.0059	0.0964	2	1.31	0.0077	0.1847	0	1.56	0.0094	0.0696	27
Common range	1.48	0.0059	0.0968	2	1.31	0.0076	0.1858	0	1.56	0.0094	0.0699	27
Stürmer	1.91	0.0050	0.0072	91	1.98	0.0059	0.0060	95	2.01	0.0082	0.0082	95
Walker	1.79	0.0054	0.0179	68	2.00	0.0068	0.0068	96	2.00	0.0109	0.0109	96
Crump	1.48	0.0059	0.0964	2	1.38	0.0074	0.1442	1	1.93	0.0073	0.0085	93
Matching weights												
No trimming	1.56	0.0050	0.0678	7	1.42	0.0063	0.1238	1	1.66	0.0078	0.0422	44

Common range	1.56	0.0050	0.0680	7	1.42	0.0063	0.1243	1	1.66	0.0078	0.0423	44
Stürmer	1.92	0.0048	0.0065	92	1.98	0.0058	0.0059	96	2.00	0.0082	0.0083	95
Walker	1.82	0.0050	0.0141	74	2.00	0.0068	0.0068	96	2.00	0.0110	0.0110	96
Crump	1.56	0.0051	0.0678	7	1.48	0.0062	0.0975	3	1.95	0.0069	0.0076	94
Entropy weights												
No trimming	1.46	0.0062	0.1058	2	1.25	0.0089	0.2270	0	1.45	0.0126	0.1139	17
Common range	1.46	0.0062	0.1064	2	1.25	0.0088	0.2292	0	1.45	0.0125	0.1152	16
Stürmer	1.91	0.0050	0.0074	91	1.97	0.0059	0.0061	95	2.01	0.0083	0.0083	95
Walker	1.78	0.0055	0.0188	67	2.00	0.0068	0.0068	96	2.00	0.0109	0.0109	96
Crump	1.46	0.0062	0.1057	2	1.33	0.0081	0.1725	1	1.92	0.0077	0.0096	92

IPTW: Inverse probability of treatment weights; average treatment effect in the overall population; SMR-treated weights: Standardized mortality ratio weights; average treatment effect in the treated; SMR-untreated weights: Standardized mortality ratio weights; average treatment effect in the untreated; RR: incidence rate ratio; AUC: Area under the receiver operating characteristics curve (c-statistic)

[a] True RR=2.0; treatment prevalence = 50%

[b] Variance of treatment effect estimates [log(RR)] over 5,000 simulated studies

[c] Mean squared error; mean of $[\log(RR) - \log(2.0)]^2$ over 5,000 simulated studies

[d] Percent of simulated studies in which the 95% confidence interval (CI) includes the true value (RR=2.0)

[e] Common range: restricting to positivity; Stürmer: asymmetric trimming at 5^{th} percentile of treated and 95^{th} percentile of untreated; Walker: preference score trimming below 0.3 and above 0.7; and Crump: propensity score trimming below 0.1 and above 0.9

Web Figure 1. Simulation setup from Stürmer et al. (10).

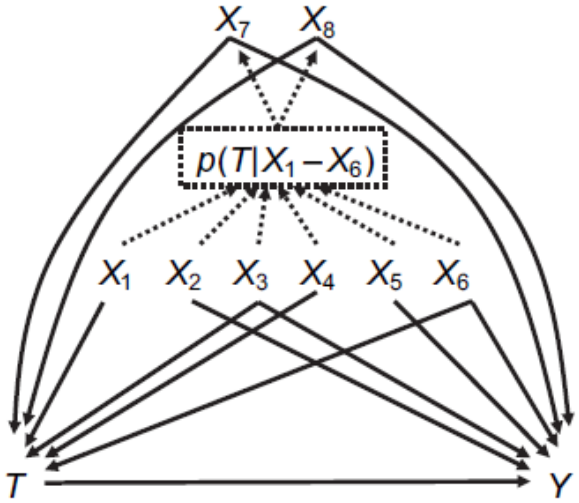

Figure 1. Conceptual diagram of a simulation study depicting treatment (T) and outcome (Y) as a function of measured covariates (X_1–X_6) and unmeasured covariates (X_7 and X_8). The solid lines represent causal associations, and the dashed lines represent noncausal associations used in the 2-step simulation process to mimic treatment contrary to prediction by measured covariates (X_1–X_6).

Web Figure 2. Density plots of propensity score distributions in treated (blue) and untreated (red) in the basic scenario without unmeasured confounding according to the c-statistic of propensity score model; left panel: c=0.75; middle panel: c=0.85; right panel: c=0.65; all based on a single imputation of a population with N=1,000,000.