Developing methods for model selection in causal health analyses: Literature Review Summary

University of Bristol

June 2024

Plan

Queries

Following fiddling / adjustment, we settle on querying the databases with the following :

- Web-of-Science:
  (TI=(confounder) OR TI=(confounding)) AND (ALL=(statistical model) OR ALL=(model selection) OR ALL=(variable selection)) AND (ALL=(epidemiology) OR ALL=(causal))
- PubMed:
  ("confounding"[Title/Abstract] OR "confounder"[Title/Abstract]) AND ("model selection"[Title/Abstract] OR "variable selection"[Title/Abstract])

Table of Papers

All papers returned by literature review query:

```
*** All Papers ***
                                                                                                                  Title  ... relevancy
0                    Familial Confounding of the Association Between Maternal Smoking During Pregnancy and Offspring Substance Use and Problems  ...         0
1            Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses  ...         2
2                                                  Confounding factors in determining causal soil moisture-precipitation feedback  ...         0
3      Concerns Regarding P Value-Based Variable Selection of Exposure Variables and Confounding Factors: Comment on the Article by Hawker et al  ...         0
4                                               Risk factor adjustment in marginal structural model estimation of optimal treatment regimes  ...         2
..                                                                                                                              ...  ...       ...
496                                        SENSITIVITY ANALYSIS FOR UNMEASURED CONFOUNDING IN COARSE STRUCTURAL NESTED MEAN MODELS  ...         1
497                                               SEMIPARAMETRIC INFERENCE OF CAUSAL EFFECT WITH NONIGNORABLE MISSING CONFOUNDERS  ...         1
498                                          SEMIPARAMETRIC CAUSAL MEDIATION ANALYSIS WITH UNMEASURED MEDIATOR-OUTCOME CONFOUNDING  ...         1
499                                          Reporting on statistical methods to adjust for confounding:: A cross-sectional survey  ...         0
500                                                 Lipid Peroxidation With Implication of Organic Pollution in Autistic Behaviors  ...         0

[501 rows x 7 columns]
```

We have 501 papers from these sources to search through.

Review Process

Starting with a set of paper names and DOI's, we assign to each a binary indicator for three categories:

- "Relevant_methods": contains any novel method or discussion of method behaviour
- "Relevant_simulation": good examples of simulation studies
- "Review_papers": papers which are themselves thorough reviews of the literature

Review Process

A1 = Title

	A	B	C	D	E	F
1	Title	DOI	Relevant_methods	Relevant_simulation	Review_paper	Notes
2	Familial Confounding of the Association Between Maternal Smoking During Pregnancy and Offspring Substance Use and Problems	https://doi.org/10.1001/archgenpsychiatry.2011.2107	0	0	0	
3	Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses	https://doi.org/10.1002/1099-095X(200011/12)11:6<705::AI	1	1	0	Table 1
4	Confounding factors in determining causal soil moisture-precipitation feedback	https://doi.org/10.1002/2016WR019869	0	0	0	
5	Concerns Regarding P Value-Based Variable Selection of Exposure Variables and Confounding Factors: Comment on the Article by H	https://doi.org/10.1002/art.41771	0	0	0	
6	Risk factor adjustment in marginal structural model estimation of optimal treatment regimes	https://doi.org/10.1002/bimj.200800182	1	1	0	Table 1, Table 2
7	Correction of confounding bias in non-randomized studies by appropriate weighting	https://doi.org/10.1002/bimj.201000154	1	0	0	
8	Inverse-probability-of-treatment weighted estimation of causal parameters in the presence of error-contaminated and time-dependent c	https://doi.org/10.1002/bimj.201600228	1	1	0	Table 1, measurement error
9	Covariate selection strategies for causal inference: Classification and comparison	https://doi.org/10.1002/bimj.201700294	1	1	1	
10	Propensity score methods for time-dependent cluster confounding	https://doi.org/10.1002/bimj.201900277	1	1	0	
11	A comparison of full model specification and backward elimination of potential confounders when estimating marginal and conditional o	https://doi.org/10.1002/bimj.202100237	1	1	0	
12	Causal relationship and confounding in statistical models	https://doi.org/10.1002/bjs.10241	0	0	0	theory
13	Assessing spatial confounding in cancer disease mapping using R	https://doi.org/10.1002/cnr2.1263	1	0	0	spatial
14	Season-dependent impact of forage quality on stress in alpine chamois	https://doi.org/10.1002/ece3.10045	0	0	0	
15	Annual variation in breeding success in boreal forest grouse: Four decades of monitoring reveals bottom-up drivers to be more importa	https://doi.org/10.1002/ece3.9327	0	0	0	
16	Model selection in occupancy models: Inference versus prediction	https://doi.org/10.1002/ecy.3942	1	0	0	M-DAG
17	Mitigating spatial confounding by explicitly correlating Gaussian random fields	https://doi.org/10.1002/env.2727	1	0	0	spatial
18	Confounded by Sequencing Depth in Association Studies of Rare Alleles	https://doi.org/10.1002/gepi.20574	0	0	0	
19	Summary statistic analyses can mistake confounding bias for heritability	https://doi.org/10.1002/gepi.22259	0	0	0	
20	Intermediate confounding in trio relationships: The importance of complete data in effect size estimation	https://doi.org/10.1002/gepi.22294	0	0	0	
21	An enhanced machine learning tool forcis-eQTL mapping with regularization and confounder adjustments	https://doi.org/10.1002/gepi.22341	1	0	0	MI-based
22	Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from in	https://doi.org/10.1002/gepi.22556	0	0	0	
23	Stochastic dynamic causal modeling of working memory connections in cocaine dependence	https://doi.org/10.1002/hbm.22212	0	0	0	
24	Observing cognitive processes in time through functional MRI model comparison	https://doi.org/10.1002/hbm.26114	0	0	0	
25	Applying nonexperimental study approach to analyze historical batch data	https://doi.org/10.1002/jps.23066	0	0	0	
26	The role of the c-statistic in variable selection for propensity score models	https://doi.org/10.1002/pds.2074	1	0	0	
27	Measuring balance and model selection in propensity score methods	https://doi.org/10.1002/pds.2188	1	0	0	
28	Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system	https://doi.org/10.1002/pds.2328	1	0	0	
29	Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study	https://doi.org/10.1002/pds.3356	1	1	0	
30	A review of covariate selection for non-experimental comparative effectiveness research	https://doi.org/10.1002/pds.3506	1	0	1	

Results

Subsetting these Tables

Relevant papers:

*** Relevant Papers ***

	Title	...	relevancy
1	Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses	...	2
4	Risk factor adjustment in marginal structural model estimation of optimal treatment regimes	...	2
5	Correction of confounding bias in non-randomized studies by appropriate weighting	...	1
6	Inverse-probability-of-treatment weighted estimation of causal parameters in the presence of error-contaminated and time-dependent confounders	...	2
7	Covariate selection strategies for causal inference: Classification and comparison	...	3
..	...	...	...
490	Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS	...	1
491	Confounder detection in linear mediation models: Performance of kernel-based tests of independence	...	1
496	SENSITIVITY ANALYSIS FOR UNMEASURED CONFOUNDING IN COARSE STRUCTURAL NESTED MEAN MODELS	...	1
497	SEMIPARAMETRIC INFERENCE OF CAUSAL EFFECT WITH NONIGNORABLE MISSING CONFOUNDERS	...	1
498	SEMIPARAMETRIC CAUSAL MEDIATION ANALYSIS WITH UNMEASURED MEDIATOR-OUTCOME CONFOUNDING	...	1

[261 rows x 7 columns]

After removing applications of methods and other irrelevant articles, we have 261 papers left.

Subsetting these Tables

Relevant methods papers:

```
*** Relevant Methods Papers ***
                                                                                                              Title  ... relevancy
1             Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses  ...         2
4                                                       Risk factor adjustment in marginal structural model estimation of optimal treatment regimes  ...         2
5                                                                  Correction of confounding bias in non-randomized studies by appropriate weighting  ...         1
6    Inverse-probability-of-treatment weighted estimation of causal parameters in the presence of error-contaminated and time-dependent confounders  ...         2
7                                                                Covariate selection strategies for causal inference: Classification and comparison  ...         3
..                                                                                                              ...  ...       ...
490                  Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS  ...         1
491                                             Confounder detection in linear mediation models: Performance of kernel-based tests of independence  ...         1
496                                                      SENSITIVITY ANALYSIS FOR UNMEASURED CONFOUNDING IN COARSE STRUCTURAL NESTED MEAN MODELS  ...         1
497                                                              SEMIPARAMETRIC INFERENCE OF CAUSAL EFFECT WITH NONIGNORABLE MISSING CONFOUNDERS  ...         1
498                                                         SEMIPARAMETRIC CAUSAL MEDIATION ANALYSIS WITH UNMEASURED MEDIATOR-OUTCOME CONFOUNDING  ...         1

[254 rows x 7 columns]
```

Of the 261, 252 contain a relevant discussion of methods.

Subsetting these Tables

Relevant simulation example papers:

*** Relevant Simulation Papers ***

	Title	...	relevancy
1	Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses	...	2
4	Risk factor adjustment in marginal structural model estimation of optimal treatment regimes	...	2
6	Inverse-probability-of-treatment weighted estimation of causal parameters in the presence of error-contaminated and time-dependent confounders	...	2
7	Covariate selection strategies for causal inference: Classification and comparison	...	3
8	Propensity score methods for time-dependent cluster confounding	...	2
9	A comparison of full model specification and backward elimination of potential confounders when estimating marginal and conditional causal effects...	...	2
27	Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study	...	2
32	A comparison of confounder selection and adjustment methods for estimating causal effects using large healthcare databases	...	3
34	Controlling for confounding via propensity score methods can result in biased estimation of the conditional AUC: A simulation study	...	1
38	Improved estimation of controlled direct effects in the presence of unmeasured confounding of intermediate variables	...	2
43	Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemio...	...	2
63	Confounder selection strategies targeting stable treatment effect estimators	...	2
65	Practical recommendations on double score matching for estimating causal effects	...	2
66	Identification and robust estimation of swapped direct and indirect effects: Mediation analysis with unmeasured mediator-outcome confounding and i...	...	2
67	Consistent inverse probability of treatment weighted estimation of the average treatment effect with mismeasured time-dependent confounders	...	2
78	Sensitivity analysis for unobserved confounding in causal mediation analysis allowing for effect modification, censoring and truncation	...	2
88	Causal modelling of heavy-tailed variables and confounders with application to river flow	...	2
97	High-Dimensional Propensity Score-Adjusted Case-Crossover for Discovering Adverse Drug Reactions from Computerized Administrative Healthcare Datab...	...	2
108	Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure	...	2
136	Comparison of high-dimensional confounder summary scores in comparative studies of newly marketed medications	...	2
211	Controlling for spatial confounding and spatial interference in causal inference: modelling insights from a computational experiment	...	2
219	Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders	...	1
228	Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution-A Simulat...	...	2
229	Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples	...	2
254	Investigation of the structure and magnitude of time-varying uncontrolled confounding in simulated cohort data analyzed using g-computation	...	2
264	SIMULATION STUDY OF CONFOUNDER-SELECTION STRATEGIES	...	2
267	Selection bias and confounding in case-crossover analyses of environmental time-series data	...	1
271	Studies with Many Covariates and Few Outcomes Selecting Covariates and Implementing Propensity-Score-Based Confounding Adjustment...	...	1
276	Using Super Learner Prediction Modeling to Improve High-dimensional Propensity Score Estimation	...	2
286	A Comparison of Methods to Estimate the Hazard Ratio Under Conditions of Time-varying Confounding and Nonpositivity	...	2
309	Data-driven confounder selection via Markov and Bayesian networks	...	2
391	Propensity score specification for optimal estimation of average treatment effect with binary response	...	2
393	The change in estimate method for selecting confounders: A simulation study	...	2
414	Comparing g-computation, propensity score-based weighting, and targeted maximum likelihood estimation for analyzing externally controlled trials w...	...	3
472	How to Apply Multiple Imputation in Propensity Score Matching with Partially Observed Confounders: A Simulation Study and Practical Recommendations	...	2

[35 rows x 7 columns]

Of the 261, 35 are good simulation examples worth emulating.

Subsetting these Tables

Key review papers:

*** Relevant Review Papers ***

	Title	...	relevancy
7	Covariate selection strategies for causal inference: Classification and comparison	...	3
28	A review of covariate selection for non-experimental comparative effectiveness research	...	2
31	Addressing unmeasured confounding in comparative observational research	...	2
32	A comparison of confounder selection and adjustment methods for estimating causal effects using large healthcare databases	...	3
33	Machine learning for improving high-dimensional proxy confounder adjustment in healthcare database studies: An overview of the current literature	...	2
79	Variable selection: current practice in epidemiological studies	...	2
84	Principles of confounder selection	...	2
85	A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement	...	2
92	Low-dimensional confounder adjustment and high-dimensional penalized estimation for survival analysis	...	2
94	A Review of the Techniques Used to Control Confounding Bias and How Spatiotemporal Variation Can Be Controlled in Environmental Impact Studies	...	2
106	Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics	...	2
134	Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review	...	2
138	Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review	...	2
146	Nonlinear association criterion, nonlinear Granger causality and related issues with applications to neuroimage studies	...	1
217	Recent Methodological Trends in Epidemiology: No Need for Data-Driven Variable Selection?	...	2
286	A Comparison of Methods to Estimate the Hazard Ratio Under Conditions of Time-varying Confounding and Nonpositivity	...	2
378	Estimating controlled direct effects in the presence of intermediate confounding of the mediator-outcome relationship: Comparison of five differen...	...	2
414	Comparing g-computation, propensity score-based weighting, and targeted maximum likelihood estimation for analyzing externally controlled trials w...	...	3
421	Model selection, confounder control, and marginal structural models: Review and new applications	...	2
426	Regularization and Confounding in Linear Regression for Treatment Effect Estimation	...	1
471	Meta-Analysis of Observational Studies with Unmeasured Confounders	...	2

[21 rows x 7 columns]

Of the 261, 21 are themselves key reviews of the literature.

Subsetting these Tables

"Best" papers:

```
*** Relevant Best Papers ***
                                                                                                                                Title  ... relevancy
7                                                                     Covariate selection strategies for causal inference: Classification and comparison  ...         3
32                                  A comparison of confounder selection and adjustment methods for estimating causal effects using large healthcare databases  ...         3
414  Comparing g-computation, propensity score-based weighting, and targeted maximum likelihood estimation for analyzing externally controlled trials w...  ...         3

[3 rows x 7 columns]
```

Of the 261, 3 papers were relevant in all three categories simultaneously.

Methods List

From this, we move towards a list of all 115 methods encountered in the literature with claims around confounder handling:

	A	B
1	Method_name	Method_name_long
2	Positive Control Analysis	Positive Control Analysis
3	Negative Control Analysis	Negative Control Analysis
4	MSM	Marginal Structural Models
5	Risk Scores	Risk Scores
6	Propensity Scores (standard)	Propensity Scores (standard)
7	Propensity Scores (clusters)	Propensity Scores (clusters)
8	Propensity Scores (c statistic)	Propensity Scores (c statistic)
9	Propensity Scores (high dimensional, typical)	Propensity Scores (high dimensional, typical)
10	Propensity Scores (high dimensional, adjusted)	Propensity Scores (high dimensional, adjusted)
11	Propensity Scores (high dimensional, large scale) (LSPS)	Propensity Scores (high dimensional, large scale) (LSPS)
12	C-Bias Analysis	
13	IV Analysis	Instrumental variable analysis
14	IPTW (naive)	Inverse probability of treatment weighting (naive)
15	IPTW (corrected)	Inverse probability of treatment weighting (corrected)
16	Correction-based	Correction-based
17	AIC	Akaike information criterion
18	BIC	Bayesian information criterion
19	LDSC	Linkage disequilibrium score regression
20	LASSO (standard)	Least absolute shrinkage and selection operator (standard)
21	LASSO (outcome-adaptive)	Least absolute shrinkage and selection operator (outcome-adaptive)
22	LASSO (LSKM)	Least absolute shrinkage and selection operator (least squares kernel machine)
23	LASSO (group)	Least absolute shrinkage and selection operator (group)
24	LASSO (least-angle)	Least absolute shrinkage and selection operator (least-angle)
25	LASSO (infinitesimal stagewise)	Least absolute shrinkage and selection operator (infinitesimal stagewise)
26	Forward Selection	Forward Selection
27	Backward Selection	Backward Selection