The effect of machine learning tools for evidence synthesis on resource use and time-to-completion

Chris Rose, Norwegian Institute of Public Health ( 2 Mar 2023)

Generated using git revision: f12b622

# Methods

Analyses were performed as specified in the protocol using Stata 16 (StataCorp LLC, College Station, Texas, USA). Briefly, we analyzed resource use (person-hours) on the log scale using extended interval regression (eintreg) and used a likelihood-adjusted-censoring inverse-probability-weighted regression adjustment (LAC-IPWRA; stteffects) model to estimate mean difference in time-to-completion. Ongoing reviews were right censored at the end of data collection (31 January 2023). All analyses accounted for right-censored outcomes and for nonrandom endogenous treatment allocation, which was modelled in terms of review field (welfare or healthcare) and whether any evidence synthesis (quantitative or qualitative) was planned. We had no reason to suspect informative (nonrandom) censoring, so did not model a censoring mechanism. We re-expressed the estimates as ratios (relative resource use and relative time-to-completion) to aid generalization to other institutions. We present two-sided 95% confidence intervals and p-values where appropriate and use a prespecified p < 0.05 significance criterion throughout. We also present the time-to-completion data using Kaplan-Meier estimates of survivor functions.

# Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type of ML Use** | **Reviews** | **Sample Mean¹** | **Effect Estimate²** | **p-value** |
| **Resource Use** |  | Person-hours |  |  |
| G | 19 | 569 | 0.27 (0.03 to 2.75) | 0.269 |
| H | 11 | 140 |  |  |
| J | 19 | 569 | 1.49 (0.78 to 2.85) | 0.223 |
| K | 5 | 888 |  |  |
| Q | 24 | 625 | 1.54 (0.52 to 4.55) | 0.439 |
| R | 11 | 140 |  |  |
| **Time-to-completion** |  | Weeks |  |  |
| G | 19 | 27.4 | 1.08 (0.64 to 1.84) | 0.767 |
| H | 12 | 28.2 |  |  |
| J | 19 | 27.4 | 0.89 (0.55 to 1.46) | 0.647 |
| K | 5 | 36.2 |  |  |
| Q | 24 | 29.3 | 1.06 (0.66 to 1.71) | 0.800 |
| R | 12 | 28.2 |  |  |
| ¹Data are means of samples restricted to completed reviews and may underestimate resource use (person-hours) and time-to-completion (weeks) due to right-censoring of ongoing projects. ²Estimates are relative resource use and relative time-to-completion, account for right-censored outcomes and nonrandom endogenous treatment allocation, and are adjusted for planned meta-analysis. | | | | |

## Kaplan-Meier estimates for Recommended vs No ML Use

## Kaplan-Meier estimates for Recommended vs Ron-recommended ML Use

## Kaplan-Meier estimates for Any vs No ML Use

# References

TODO: Add references.

# Appendix 1 — Protocol Deviations

We updated the preprint version of the protocol during data extraction but before starting the analysis or unblinding the statistician (CJR) to redefine the comparisons in terms of under- and overuse of machine learning (TODO: Cite revision). However, only two reviews were judged to have under- or overused machine learning, so it was not possible to perform the revised analyses. We therefore performed and report the analyses as originally planned.

# Appendix 2 — Full Regression Results

## Regression results for rec\_vs\_none with respect to resource

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| meta\_analysis\_planned |  |  |  |  |  |  |
| Yes | 0.97 | 0.37 | 2.63 | 0.01 | 0.25 | 1.70 |
|  |  |  |  |  |  |  |
| rec\_vs\_none |  |  |  |  |  |  |
| H | -1.31 | 1.19 | -1.11 | 0.27 | -3.64 | 1.01 |
| \_cons | 6.00 | 0.37 | 16.42 | 0.00 | 5.28 | 6.71 |
| rec\_vs\_none |  |  |  |  |  |  |
| field |  |  |  |  |  |  |
| Welface | -0.90 | 0.48 | -1.89 | 0.06 | -1.83 | 0.03 |
| /rec\_vs\_none |  |  |  |  |  |  |
| cut1 | -0.10 | 0.35 |  |  | -0.78 | 0.59 |
| var(e.log\_resource1) | 0.58 | 0.15 |  |  | 0.35 | 0.96 |
| corr(e.rec\_vs\_none,e.log\_resource1) | -0.01 | 1.07 | -0.01 | 0.99 | -0.97 | 0.97 |

## Regression results for rec\_vs\_none with respect to time

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Robust |  |  |  |  |
| \_t | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| ATE |  |  |  |  |  |  |
| rec\_vs\_none |  |  |  |  |  |  |
| (H vs G) | 2.74 | 9.30 | 0.29 | 0.77 | -15.48 | 20.97 |
| POmean |  |  |  |  |  |  |
| rec\_vs\_none |  |  |  |  |  |  |
| G | 32.94 | 6.40 | 5.15 | 0.00 | 20.40 | 45.48 |

## Regression results for rec\_vs\_nonrec with respect to resource

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| rec\_vs\_nonrec |  |  |  |  |  |  |
| K | 0.40 | 0.33 | 1.22 | 0.22 | -0.24 | 1.05 |
|  |  |  |  |  |  |  |
| meta\_analysis\_planned |  |  |  |  |  |  |
| Yes | 0.33 | 0.27 | 1.21 | 0.23 | -0.20 | 0.86 |
| \_cons | 6.31 | 0.20 | 31.63 | 0.00 | 5.91 | 6.70 |
| /lnsigma | -0.38 | 0.15 | -2.55 | 0.01 | -0.66 | -0.09 |
| sigma | 0.69 | 0.10 |  |  | 0.51 | 0.92 |

## Regression results for rec\_vs\_nonrec with respect to time

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Robust |  |  |  |  |
| \_t | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| ATE |  |  |  |  |  |  |
| rec\_vs\_nonrec |  |  |  |  |  |  |
| (K vs J) | -3.72 | 8.29 | -0.45 | 0.65 | -19.96 | 12.52 |
| POmean |  |  |  |  |  |  |
| rec\_vs\_nonrec |  |  |  |  |  |  |
| J | 34.31 | 7.26 | 4.73 | 0.00 | 20.09 | 48.54 |

## Regression results for any\_vs\_none with respect to resource

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| meta\_analysis\_planned |  |  |  |  |  |  |
| Yes | 0.99 | 0.31 | 3.17 | 0.00 | 0.38 | 1.61 |
|  |  |  |  |  |  |  |
| any\_vs\_none |  |  |  |  |  |  |
| R | 0.43 | 0.55 | 0.77 | 0.44 | -0.66 | 1.52 |
| \_cons | 5.54 | 0.31 | 17.88 | 0.00 | 4.94 | 6.15 |
| any\_vs\_none |  |  |  |  |  |  |
| field |  |  |  |  |  |  |
| Welface | -0.59 | 0.31 | -1.89 | 0.06 | -1.20 | 0.02 |
| /any\_vs\_none |  |  |  |  |  |  |
| cut1 | 0.28 | 0.25 |  |  | -0.21 | 0.76 |
| var(e.log\_resource1) | 1.34 | 0.55 |  |  | 0.61 | 2.98 |
| corr(e.any\_vs\_none,e.log\_resource1) | -0.95 | 0.08 | -12.47 | 0.00 | -1.00 | -0.25 |

## Regression results for any\_vs\_none with respect to time

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Robust |  |  |  |  |
| \_t | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] | |
| ATE |  |  |  |  |  |  |
| any\_vs\_none |  |  |  |  |  |  |
| (R vs Q) | 2.22 | 8.89 | 0.25 | 0.80 | -15.19 | 19.64 |
| POmean |  |  |  |  |  |  |
| any\_vs\_none |  |  |  |  |  |  |
| Q | 35.15 | 5.45 | 6.45 | 0.00 | 24.47 | 45.84 |