References

1. Mittinty MN, Lynch J. Reflection on modern methods: risk ratio regression—simple concept yet complex computation. *Int J Epidemiol* 2023;**52**:309–14.
2. Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC Press, 2020.

International Journal of Epidemiology, 2023, **52**, 1991–1992
https://doi.org/10.1093/ije/dyad077
Advance Access Publication Date: 27 May 2023

OXFORD

Risk ratio regression—simple concept yet complex computation

Murthy N Mittinty[1]* and John Lynch[2,3]

[1]College of Medicine and Public Health, Flinders University, Adelaide, South Australia, Australia, [2]School of Public Health, The University of Adelaide, Adelaide, Australia and [3]Population Health Sciences, University of Bristol, Bristol, UK

*Corresponding author. College of Medicine and Public Health, Flinders University, Adelaide, SA 5042, Australia.
E-mail: murthy.mittinty@flinders.edu.au

In our article we state that if a log-binomial regression with default computation algorithm (Fisher Scoring matrix) converges then it is straightforward to read the risk ratio from the beta-coefficient. In such situations one does not need to use logistic regression or any other regression methods described in our manuscript to estimate the risk ratio. For this case, we agree with Popham that the risk ratio is a simple concept and a simple computation. However, when the log-binomial regression does not converge then one needs to use alternative methods as suggested in the manuscript. The process we refer to as complex computation involves choosing the appropriate regression/computation methods for estimating the maximum likelihood and in turn the risk ratio, standard error and confidence intervals to perform a test of null hypothesis.

The suggestion proposed by Popham to use the predicted probabilities from the confounder-adjusted logistic regression requires care. There are now many model-based estimates for risk ratios as discussed in earlier publications.[1,2] One needs to be clear on how to obtain the predicted probabilities. Is it through (i) marginal standardization or (ii) prediction to modes or (iii) prediction to means? These estimates will correspond to three different target populations.[3] When marginal standardization is chosen, then the estimated probabilities will allow inference suitable for any specific target population. Method (ii) would yield estimates corresponding to a specific stratum of individuals defined by the most common confounder values. Method (iii) is good when the covariate distributions are continuous. In the presence of dichotomous covariates, Method (iii) will yield results that are not meaningful to any real-world data. Even though one may be clear on the target population, the computation of confidence intervals becomes tedious as one needs to use a bootstrapping or delta method.

Like log-binomial, logistic regression also suffers from issues such as sparse data bias. Sparse data bias occurs when there are few case numbers for a combination of covariates and outcome levels. To overcome this issue, one may need to use methods such as Frith regression or Exact logistic regressions.[4] Once again, the estimated probabilities from these regressions will correspond to different target populations. Moreover, the standardized summary may mask the heterogeneity of the effect and it is for this reason that one needs to examine the stratum-specific estimates too. Additionally, the standardized estimate should not be combined with its standard error to perform a test of null hypothesis, as the resulting test could be inefficient.[1] Further, model-based estimates of risk can be sensitive to model misspecifications and influential data points.

If the model fits well, the model-based estimates tend to have smaller mean-squared error compared with raw covariate-specific estimates.[5] However, computation of the risk ratio and its confidence intervals becomes involved. Overall, we think that when the log-binomial does not converge, the process of estimating the risk ratio and its confidence intervals is complex.

Author contributions

M.N.M. wrote the initial draft; J.L. read and approved the draft. Both authors approved the final version.

Conflict of interest

None declared.

References

1. Greenland S, Holland P. Estimating standardized risk differences from odds ratios. *Biometrics* 1991;**47**:319–22.
2. Stijnen T, Van Houwelingen HC. Relative risk, risk difference and rate difference models for sparse stratified data: a pseudo likelihood approach. *Stat Med* 1993;**12**:2285–303.
3. Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol* 2014;**43**: 962–70.
4. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016;**27**:i1981.
5. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

International Journal of Epidemiology, 2023, **52**, 1992–1994
https://doi.org/10.1093/ije/dyad131
Advance Access Publication Date: 3 October 2023

OXFORD

Solid cancer mortality among US radiation workers

Richard Wakeford

Centre for Occupational and Environmental Health, The University of Manchester, Manchester, UK

Centre for Occupational and Environmental Health, The University of Manchester, Ellen Wilkinson Building, Oxford Road, Manchester M13 9PL, UK. E-mail: Richard.Wakeford@manchester.ac.uk

The paper by Kelly-Reif *et al.*[1] updating the findings for rates of solid cancer mortality in a pooled cohort of 101 363 workers monitored for exposure to ionizing radiation while employed for ≥1 year at five sites in the USA is a welcome addition to the epidemiological evidence on the risk of cancer following exposure to a series of many small doses of radiation received at a low dose rate.[2] As the authors note, such exposure is a major consideration in contemporary routine radiological protection and the evidence derived from protracted occupational exposures will play an important role in the current deliberations of the International Commission on Radiological Protection when assessing the need to revise its recommendations.[3,4]

Employment records at the five US sites were available from various years during the period 1944–52, the initial year depending on the site, so that the study included the early years of employment when radiation exposures tended to be higher than in recent years. Kelly-Reif *et al.*[1] modelled the relationship between solid cancer mortality and radiation exposure in the cohort as the linear excess relative rate per unit equivalent dose of radiation accumulated occupationally (ERR/Sv) with a dose lag of 10 years to account for the minimum latent period. Estimates of ERR/Sv were presented for a number of types, and groups of types, of solid cancer with follow-up to the end of 2016; there were a total of 12 069 solid cancer deaths. The cumulative dose used in the analysis was that received from penetrating low linear energy transfer radiations (x-rays and gamma-rays) as measured and recorded for the purposes of regulation and radiological protection; doses from neutrons and intakes of radionuclides were not included.

Estimates of ERR/Sv were presented for the full cohort, and also for the sub-cohort of 56 260 workers who were first employed in 1960 or later (among whom there were 3712 solid cancer deaths), because the technology for measuring photon doses had improved by 1960 and recorded doses would be expected to be more accurate during this later period. Table 1 shows the ERR/Sv estimates for the full cohort and the sub-cohort for those types, and groups of types, of solid cancers (and chronic obstructive pulmonary disease) with >1000 deaths in the full cohort.[1]

What is striking in Table 1 is that the ERR/Sv estimates for the sub-cohort of workers hired after 1959 are consistently and notably larger than those for the entire cohort. Unfortunately, Kelly-Reif *et al.*[1] do not present ERR/Sv estimates for the sub-cohort of 45 103 workers first employed before 1960 (among whom there were there were 8357 solid cancer deaths), so it is not possible to make a proper assessment of the differences in ERR/Sv estimates between the two sub-cohorts but, on the face of it, these differences would certainly appear to be worthy of further investigation.

The mean cumulative photon dose for workers in the whole cohort is 26.5 mSv whereas that for the sub-cohort hired after 1959 is 14.5 mSv,[1] so workers first employed before 1960 accumulated, on average, substantially more