# Web Appendix B: Statistical Method

The following model is estimated with the data aggregated to the neonatal unit-month level:

(1)

where is the mortality rate for unit *j* in month *t,* is the measure of one to one nursing, is a vector of case-mix controls to represent health at admission, are unit fixed effects and are year and calendar month dummies, and is a random error term. is a vector of parameters to be estimated and is a scalar parameter of interest to be estimated.

## Estimation

We utilise two different estimators: (i) OLS that does not account for possible reverse causality running from 1:1 nursing to our outcome variable; (ii) Instrumental Variable Estimator (IVE) that accounts for this. In the former case, the estimator relies on the assumption of strict exogeneity, i.e. , , where . However, this assumption may not hold in this case since contains unobserved differences in average case-mix which are likely to be correlated with the treatment (such that . In the latter case we utilise a set of instrumental variables, , that satisfy for and which are correlated with .1 For these instruments we utilise lags of the one to one nursing variable, , .

It is necessary to identify the correct number of lags of one to one nursing to use as instruments for the contemporaneous level of one to one nursing. The simplest and most widely used method is sequential testing. Beginning with the maximum number of lags, which in this case is 12 as the data are monthly, the first stage model is estimated (using the within estimator) and a robust t-test conducted on the last lag; if this t-statistic is not above a certain threshold (in this case 1.96) then this lag is removed and the process repeated. This method leads to the first four lags being used as instruments.

## References

1 Cameron, A. C., & Trivedi, P. 22.2 GMM Estimation of Linear Panel Models. In *Microeconometrics: Methods and Applications* (1st Edition) 2000. Cambridge University Press.