Looking at knock-on effects: univariable, multivariable linear regression

CHRISTINA EASTER, MSc AND KARLA HEMMING, INSTITUTE OF APPLIED HEALTH RESEARCH, UNIVERSITY OF BIRMINGHAM, BIRMINGHAM, UK

Learning points
- Linear regression is used to model the association between independent variables (predictors) and dependent variables (continuous outcomes).
- Multivariable linear regression is used to account for multiple characteristics.
- Coefficients represent an average increase/decrease in the outcome for each unit increase or change in category of the predictors.

Linear regression is used to describe the association between a continuous dependent variable of interest (outcome) and other patient characteristics or predictors (independent continuous, categorical or dichotomous variables). Fitting a linear regression model estimates a constant, predictor coefficients and their corresponding 95% confidence intervals (95% CI).

In a univariable analysis, only one independent variable is included in the model and the model shows the association between that variable and the outcome, without any consideration of other characteristics. In a multivariable analysis, more than one predictor is included in the model.

Multivariable models depict the relation between multiple independent variables and the outcome, allowing for the impact of other characteristics. This is sometimes referred to as an adjusted association. For example, Waage et al. (2016) used multivariable linear regression to assess the association between ethnicity and other maternal characteristics (independent variables) with the outcome postpartum weight retention (dependent variable). Individual associations for each predictor and outcome have also been modelled (univariable analysis). Tables 1 and 2 give the coefficients and corresponding 95% CI for each independent variable for univariable and multivariable analysis, respectively. The constant (or intercept) tells us the expected outcome value when we set the categorical independent variables to the reference group and set the continuous independent variables to zero. For example, the multivariable analysis (Table 2) constant tells us the estimated postpartum weight retention (the outcome) for a Western European (reference group) with an education level of ≥12 years (reference group) and a self-reported gestational weight gain (GWG) of 0 kg (set to zero, as this is a continuous variable) is 5.2 kg (95% CI 1.2-8.2 kg) increase in postpartum weight retention compared with Western European women (the reference category). The association between ethnicity and weight gain retention can be seen in Figure 2.

Multivariable linear regression is used when there is interest in multiple characteristics on the outcome. We interpret the coefficients (Table 2) in the same way, although acknowledging adjustment for other predictors in the model.

Table 1. Univariable analysis results using linear regression for the outcome of postpartum weight retention (kg) (Waage et al. BJOG 123: 699–708)

| Maternal characteristics          | Coefficients | 95% confidence interval |
|-----------------------------------|--------------|-------------------------|
| **Ethnicity**                     |              |                         |
| Western Europe                    | Reference value |                         |
| South Asia                        | 2.80         | 1.80–3.80               |
| Middle East                       | 2.80         | 1.60–4.00               |
| Africa                            | 4.00         | 2.30–5.70               |
| East Asia                         | 1.40         | –0.40 to 3.20           |
| Eastern Europe                    | 2.90         | 1.10–4.80               |
| Self-reported gestational weight gain (kg) | 0.51         | 0.50–0.60               |
| **Education level**               |              |                         |
| ≥12 years                         | Reference value |                         |
| <12 years                         | 2.00         | 1.20–2.80               |
variables), ethnicity and self-reported gestational weight gain.

Useful resources
A simple linear regression paper to aid understanding.
- http://www.bmj.com/content/346/bmj.f2340
Multivariable linear regression furthering understanding on interpretation.
- http://www.bmj.com/content/349/bmj.g4887

Acknowledgements
CATCH ME: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 633196.

Disclosure of interests
None declared by CE. KH receives fees as a statistical reviewer for BJOG outside of this work. Completed disclosure of interest forms are available to view online as supporting information.

Data availability statement
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Supporting Information
Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table 2. Simplified version of the results found using multivariable linear regression analysis for ethnicity, self-reported weight gain (kg) and education level (adjusted for age and parity) on postpartum weight retention (kg). The constant value was not reported in the paper and so for illustrative purposes a reasonable value of 5.2 will be used here (Waage et al. BJOG 123:699–708)

| Maternal characteristics                  | Coefficients | 95% confidence interval |
|-------------------------------------------|--------------|-------------------------|
| Ethnicity                                 |              |                         |
| Western Europe                            | Reference value |                         |
| South Asia                                | 2.80         | 1.90–3.60               |
| Middle East                               | 2.80         | 1.00–3.00               |
| Africa                                    | 4.40         | 3.00–5.80               |
| East Asia                                 | 0.91         | -0.49 to 2.30           |
| Eastern Europe                            | 0.45         | -1.00 to 1.90           |
| Self-reported gestational weight gain (kg)| 0.54         | 0.49–0.59               |
| Education level                           |              |                         |
| ≥12 years                                 | Reference value |                         |
| <12 years                                 | 0.74         | 0.02–1.46               |

Figure 1. Linear relationship between postpartum weight retention (kg) and self-reported gestational weight gain (kg) using the univariable analysis coefficients; a reasonable constant value of 2.6 has been used for the purpose of this illustration, as the constant value was not reported in the paper.
Figure 2. Association between postpartum weight retention (kg) and ethnicity (Western Europe, South Asia, Middle East, Africa, East Asia and Eastern Europe) using the univariable analysis coefficients; a reasonable constant value of 2.6 has been used for the purpose of this illustration, as the constant value was not reported in the paper. The dashed line depicts the value of the reference category.

Figure 3. The association between self-reported gestation weight gain (kg), ethnicity and the outcome of postpartum weight retention (kg). Each line depicts the relation between self-reported gestation weight gain and postpartum weight retention for the corresponding ethnicity (Western Europe, South Asia, Middle East, Africa, East Asia and Eastern Europe). Using the results from the multivariable linear regression analysis (Table 2) and a reasonable constant value of 5.2, which has been used for the purpose of this illustration, as the constant value was not reported in the paper; adjusting for education level, age (years) and parity.