**English for Medical Students: Writing research articles in English**

**03 Developing academic literacies: The Methods section**

These activities focus on the Methods section of a typical research article. Research suggests that typically the Methods section falls into up to 3 moves with possible steps, such as the following:

1. *Describing Data-Collection Procedure:*
	1. Indicating source of data.
	2. Indicating data size.
	3. Indicating criteria for data collection.
2. *Describing Experimental Procedures:*
	1. Identification of main research apparatus.
	2. Recounting experimental process.
	3. Indicating criteria for success.
3. *Describing Data-Analysis Procedures:*
	1. Defining terminologies.
	2. Indicating process of data classification.
	3. Identifying analytical instrument/procedure.
	4. Indicating modification to instrument/procedure.

Not all of these moves and steps will be present in any given article, but they very often are. We will first look at how they are realised in the article whose Introduction we read for the last session:

Rasella, D., Harhay, M. O., Pamponet, M. L., Aquino, R., & Barreto, M. L. (2014). Impact of primary health care on mortality from heart and cerebrovascular diseases in Brazil: a nationwide analysis of longitudinal data. *BMJ,* 349, g4014*.*

[http://www.bmj.com/content/349/bmj.g4014.full.pdf+html](http://www.bmj.com/content/349/bmj.g4014.full.pdf%2Bhtml)

1. Read the full section quickly and answer the following questions:
	1. How many of the three moves are present in this article? In what order are they presented?
	2. Colour-code each move and step that you find, labelling each one.
	3. Note down any useful language. Add any particularly useful vocabulary to **Your Vocabulary** List on moodle.

**Methods**

This study employs a mixed ecological design by combining an ecological multiple-group and longitudinal time-trend design.19 The unit of analysis is the municipality (county). Descriptive analyses of the trend of hospitalisation and mortality rates for the selected cardiovascular disease causes and for all the variables along the period under study were performed; subsequently regression models to assess the association between FHP coverage and cardiovascular disease were fitted; and finally several complementary analyses, including difference-in-difference and propensity score matching, were performed to verify the robustness of the results.

From the 5507 Brazilian municipalities in 2000, we selected those that had adequate vital information according to criteria based on five indicators20: (1) age standardised mortality rate; (2) ratio of reported live births to estimated live births; (3) average annual deviation from the three year period mean of the mortality rate; (4) average annual deviation from the three year period mean of the birth rate; and (5) proportion of ill defined deaths (ICD-10 (international classiﬁcation of diseases, 10th revision) codes R95-R99). For the period under study, 1622 municipalities met the inclusion criteria for adequacy of vital information, representing 30% of the municipalities in Brazil. (Figure 1S in the supplementary data on bmj.com shows the distribution of the selected municipalities in the country.)

We calculated age standardised mortality rates for ages 20-74 years by cause of death for each municipality by the method of direct standardisation, using the Brazilian population structure of the year 2000 as the standard population. With the intention of selecting specific causes of cardiovascular disease that could be reduced by the FHP, we considered the Brazilian List of Ambulatory Care Sensitive Hospitalisations (ICSAP),15 a set of conditions for which access to effective primary care can decrease the rate of hospitalisation in Brazil. Among the causes of death listed within the ICD-10, we selected the ones from the groups of ischaemic heart diseases and other forms of heart diseases (codes I20 (angina pectoris), I24 (other acute ischaemic heart diseases), I50 (heart failure, including pulmonary oedema J81)) and cerebrovascular diseases (codes I63 (cerebral infarction), I64 (stroke, not specified as haemorrhage or infarction), I65 (occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction), I66 (occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction), I67 (other cerebrovascular diseases), I69 (sequelae of cerebrovascular disease, including transient cerebral ischaemic attacks and vascular syndromes of brain in cerebrovascular diseases G45 and G46)). Mortality from accidents (codes V01-X59) was included as a negative control, because the FHP was not expected to influence this and does not include actions related to accident prevention. Thus, we hypothesised an effect of FHP on mortality associated with cardiovascular disease but not mortality from accidents.

Two different metrics for FHP coverage, our exposure to treatment, were considered: (1) annual FHP coverage, derived by dividing the number of persons in FHP catchment areas by the total population of the municipality; and (2) the average FHP coverage in the previous four, six, and eight years, derived by the sum of the annual coverage of the previous four, six, and eight years divided by the number of respective years. The average FHP coverage measures the effect of duration of FHP implementation and the level of exposure. For example, we hypothesised that the effect of consistent high population coverage for a period of eight consecutive years would yield greater outcomes than high annual coverage only recently attained in the municipality. To aid comparisons and to obtain meaningful risk ratio estimates of the FHP effect, we categorised FHP coverage, both annual and average, using cut-offs used in previous studies: without FHP (0), incipient (coverage of <30% of the population), intermediate (coverage >30% and <70%), and consolidated (coverage ≥70%).12 13 21

Covariates recognised in the literature as determinants of mortality from cardiovascular disease were selected a priori for inclusion in the mathematical models.3 22 23 These covariates were stratified according to the median or, when available, reference values for Brazil. They included percentage of people below the poverty line in the municipality (stratiﬁed as ≤15.9% and >15.9%, median of the distribution), per capita income (monthly, stratiﬁed as ≤R$525 and >R$525, median of the distribution), percentage of individuals having basic household appliances (stratiﬁed as ≤48.8% and >48.8%, median of the distribution), percentage of individuals living in households with inadequate sanitation (stratiﬁed as ≤13.8% and >13.8%, median of the distribution), percentage of illiterates individuals over 15 years old (stratiﬁed as ≤11.0% and >11.0%, median of the distribution), presence of local hospital beds, number physicians per 1000 inhabitants (stratiﬁed as ≤0.55 and >0.55, median of the distribution), urbanisation rate (stratiﬁed as ≤76.6% and >76.6%, median of the distribution), percentage of highly educated (university graduated) among those >25 years old (stratiﬁed as ≤4.8% and >4.8%, median of the distribution), and presence of tomography and ultrasonography in the municipality.

**Data sources**

The data used in this study were collected from different national information systems. We obtained statistics made available by the Brazilian Ministry of Health from the Mortality Information System, Primary Care Information System, and Outpatient Information System.11 Socioeconomic and demographic variables were retrieved from the Brazilian Institute of Geography and Statistics.24 Because these covariates were obtained from the 2000 and 2010 national censuses databases, the annual values from 2001-09 were calculated by linear interpolation.

**Statistical analyses**

We employed multivariable conditional negative binomial regression models for panel data with ﬁxed effects speciﬁcation. The appendix on bmj.com gives a detailed explanation on the advantages of these models in impact evaluations, which are increasingly used in the recent literature,12 13 21 25 together with the regression formula used. Negative binomial regression is used when the outcome to be analysed is a count (such as deaths during a calendar year) and the Poisson model assumption that the mean is equal to variance is not met, usually because the data present greater dispersion.26 Longitudinal, also termed panel data, models include a disturbance or error term and a second term to control for unobserved time-invariant characteristics of the unit of analysis. In our analysis, these unobserved time-invariant characteristics would be the unmeasured geographical, historical, or sociocultural characteristics of each municipality.27

The choice of ﬁxed effect or random effect models was based on the Hausman speciﬁcation test and on the fact that fixed effects models permit correlations between the unobserved time-invariant term and the explanatory variables, such as the overall coverage of the programmes, allowing to control for possible selection bias in the implementation of FHP.28 As explained in the appendix, in order to verify if conditional models were removing the fixed effects of each panel,26 we fit models with different specifications, such as unconditional negative binomial regression models and conditional Poisson regressions with robust standard errors. Conditional, fixed effects, negative binomial regression models seemed to be the most appropriate for our analysis. These regression models were selected to estimate mortality rate ratios, both crude and adjusted, using municipalities with no coverage of FHP as reference category.

To elucidate the mechanism of action of the FHP, we analysed the association of FHP coverage with process variables such as the number of health education activities performed in the community, the number of domiciliary visits or consultations of health professionals of any level over the population of the municipality, and the number of medical consultations in primary care. These activities constitute an important part of the FHP health prevention and promotion strategy.10 In order to strengthen the plausibility of an effect of the FHP on cardiovascular disease mortality, multivariate models with the same independent variables were fitted using as outcome hospitalisation rates for the same ICD codes of the mortality under study. A variable representing time was not included in the models because the mortality rate ratio, comparing two groups of coverage exposed to the same mortality time trend, allowed us to control for secular trends. A detailed explanation of the theoretical aspects and sensitivity tests we performed to reasonably exclude the possibility of relevant uncontrolled secular trends is outlined in the appendix.

In order to assess the robustness of the fixed effect, multivariate regression models, we conducted a difference-in-difference analysis using the municipalities with no FHP coverage in 2000 (n=849), the years 2000 and 2009, and the same FHP coverage levels and covariates used in the fixed effects models. As further analysis with the same database, we conducted a propensity score matching analysis and fixed effects, multivariate regressions with propensity score weighting.

For database processing and analysis, Stata software version 12.0 was used.

1. Now return to the article that you chose on a topic that interests you.
	1. Read the Methods section and do the same activities that you did for (1) above.
	2. Upload your colour-coded analysis to moodle by **19th August.**
2. Think of a piece of research that you have been involved with.
	1. Draft a brief Methods section for an article about it.
	2. Upload your draft to the moodle site by **19th August**.
3. **By 19th August**, in the learning blog on moodle, write around 150-200 words in your learning blog, e.g. on the following topics:
* Useful language that you have particularly noticed that is useful in:
	+ Describing data collection
	+ Describing experimental procedures
	+ Describing data analysis procedures
* Questions/comments about how a Methods section should be drafted