

Interactive Shiny Apps for Flexible Modeling of Quantitative Predictors in Epidemiology

Alessio Crippa & Nicola Orsini

Department of Public Health Sciences, Karolinska Institutet
Stockholm, Sweden

alessio.crippa@ki.se



Abstract

Regression models (linear, logistic, Poisson, Cox, meta-analysis) are frequently used in epidemiology to investigate the dose-response relationships between quantitative predictors and the response variable. Flexible non-linear models involving transformations of the original predictors (e.g. quadratic, categories, splines terms) are common and widely used in medical research. However, their implementation and presentation may not be straightforward.

Therefore, we developed two interactive web applications in shiny, **flexmod** and **dosresmeta**, to facilitate flexible analysis of quantitative predictors. The shiny apps allow user to quickly explore different models without being familiar with any statistical software. They can be also useful for introducing the topic when teaching dose-response analysis.

Our aim is to present the developed web applications to estimate several dose-response models using common epidemiological studies of both individual (case-control, cohort) and aggregated data (meta-analysis). The main focus will be on the interpretation and presentation of the findings either in a tabular or graphical form.

Introduction

Many epidemiological studies aim to investigate the relationship between a quantitative predictor (also referred to as exposure, explanatory or independent variable) and an health outcome. The dose-response analysis can be carried out either on individual patient data or on aggregated data by using meta-analysis.

Regression models (linear, logistic, Poisson, Cox, meta-analysis) are typically used to relate the quantitative exposure to a transformation of the mean outcome (e.g. logit, log). Two common choices are to simply include the predictor in the model assuming a linear relationship, or to categorize the exposure and model it using dummy variables (assuming a step function). More flexible alternatives such as polynomials and regression splines (linear, quadratic, and cubic) [1] have been known for a long time. Their use, however, have been limited because of difficulties in the interpretation of the regression coefficients and in the graphical presentation of results [2].

Despite its importance, the dose-response analysis is oftentimes overlooked or poorly performed.

Main Objectives

Our aim is to present interactive web applications to facilitate a preliminary analysis of the dose-response association between the quantitative predictor and the health outcome based on either individual and aggregated data.

The developed web apps allow to

- estimate the dose-response association in most common regression models including linear, logistic, poisson, and Cox models, controlling for potential confounding;
- perform a dose-response analysis based on aggregated data (dose-response meta-analysis)
- present (and customize) graphically the predicted association, both in terms of absolute and relative measures;
- generate informative tables with point and interval estimates of the predicted outcome for selected exposure values;
- conduct hypothesis testing for the overall dose-response association and/or for testing departure from linearity.

Materials and Methods

Using the open source shiny R package, we developed two web apps for dose-response analysis of individual patient data, **flexmod** [4], and of aggregated data, **dosresmeta** [5]. Both web apps are hosted on a Shiny Server and available at <http://stats4life.se/webapp/flexmod> and <http://stats4life.se/webapp/dosresmeta>. The source code for the web apps is available on GitHub at <https://github.com/alecri/shiny-server>.

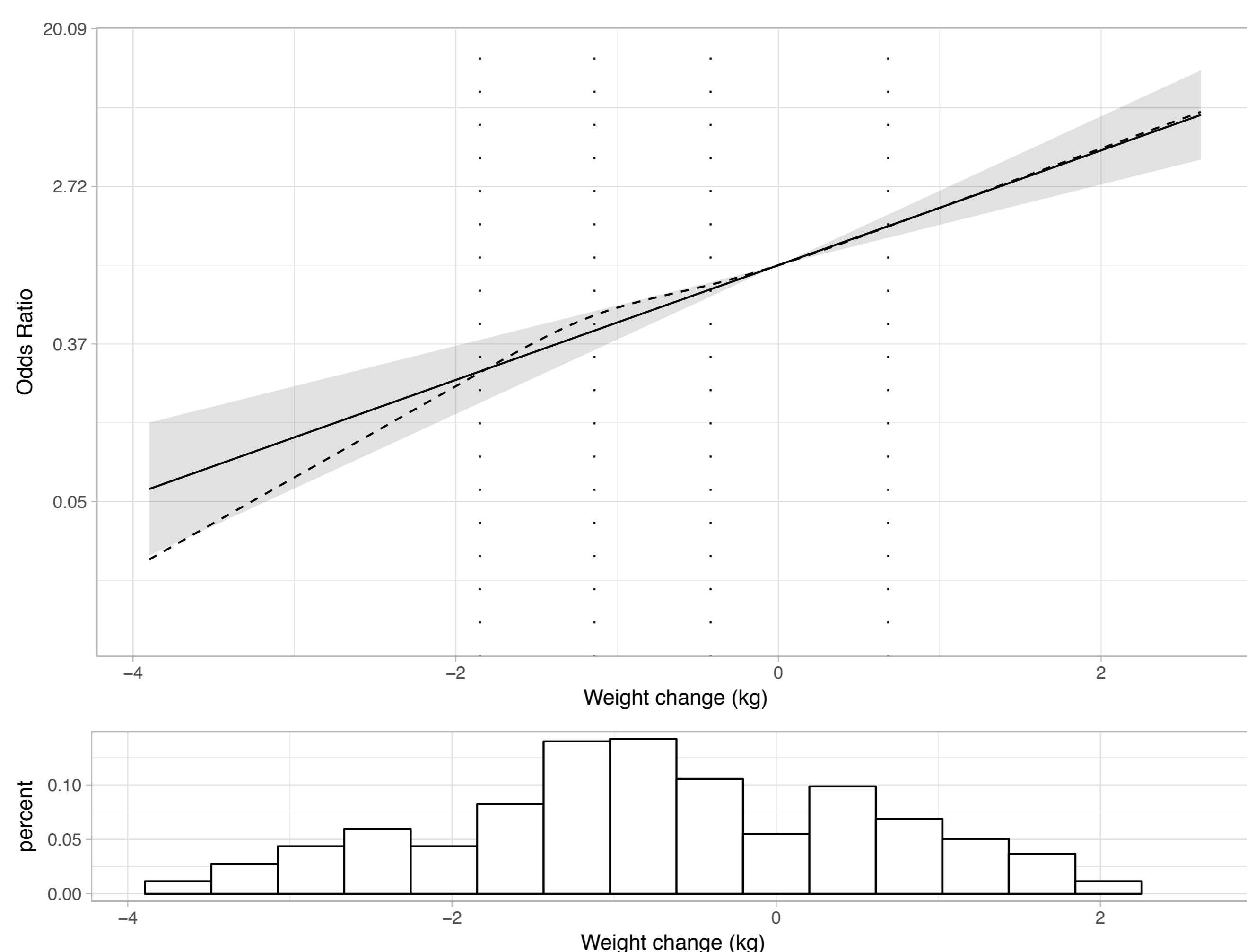


Figure 1: Example of a customizable figure build using the **flexmod** web app.

The regression models in **flexmod** are estimated using standard functions in the base (*lm*, *glm*) and survival (*Surv*) packages. The *Epi* package is employed to easily derive absolute and relative predictions. The meta-analytical models in **dosresmeta** are estimated using the *dosresmeta* package [3]. Modeling of the quantitative predictor is done using standard functions for dividing continuous variables in intervals, and for creating polynomials. The *rms* and *splines* packages are instead used for creating splines transformations. All the graphs are generated using *ggplot2*.

The **flexmod** uses data from a cohort of marathon runners to identify the principal risk factors of hyponatremia (NEJM 2005). The **dosresmeta** app, instead, provides four different aggregated data, using the association between BMI and renal cancer as default example (BJC 2001).

Results

The **flexmod** app is structured in four main tabs: data, modeling strategies, predictions, and R output. In the sidebar of the data tab it is possible to upload your own data (from an excel format) or to use the example data. In this tab, the user needs to define the type of outcome (continuous, binary, counts, or survival) and select the names for the outcome, the exposure, and the eventual confounder variables. The output in the main panel presents a snapshot of the data.

Different strategies for modeling the dose-response association of interest are presented in the sidebar of the modeling strategies tab. These include: categories (with customizable number and cut-points), linear trend, splines (linear, quadratic, cubic, and restricted cubic) with customizable number and location of knots, and polynomials. The main panel reports the graph of the estimated dose-response curves in terms of relative and absolute measures depending on the outcome type (Figure 1). The reference value for relative measures can be selected in the sidebar. It is possible to customize the graph by changing many graphical parameters (labels, line-type, axes, and themes) and download it in different formats (pdf, eps, jpeg).

Point and interval estimates for the predicted outcome (both relative and absolute) are reported in the predictions tab. It is possible to choose for which exposure values calculate the predicted outcome as well as the reference value in case of relative measures. The results in the table can be downloaded in a csv file. The last tab reports the R output with the summaries of the different dose-response models. The first table reports information regarding the significance of the different models, and *p*-values for testing departure from the linearity assumption.

The **dosresmeta** app is structured similarly to the **flexmod** app with four tabs for the results (Table, Summary, Dose-response, Prediction) but with a single sidebar to define the different inputs.

Specific to the **dosresmeta** app, the Summary tab presents descriptive plots and tables describing the aggregated data displayed in the Table tab. Similar dose-response models can be selected in the sidebar, with the analytical results displayed in the Dose-response tab, while graphical and tabular predictions are reported in the Prediction tab. It is important to notice that only relative measures are available, given the nature of aggregated data [3].

Conclusions

- In epidemiological studies it is crucial to decide how to model a quantitative predictor in a regression models.
- Any strategy has its own advantages and limitations. A graphical comparison of alternative models can help to choose which model to adopt.
- The interactive web apps can greatly facilitate teaching dose-response analysis and prevent common difficulties encountered by researchers.
- Both **flexmod** and **dosresmeta** are practical tools to perform a preliminary analysis or to run regression models without being familiar with any statistical software.
- Other shiny web app can be found at <http://stats4life.se/webapp>

References

- [1] Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*. 1995 Jul 1:356-65.
- [2] Orsini, N. and Greenland, S., 2011. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal*, 11(1), p.1.
- [3] Crippa A, Orsini N. Multivariate dose-response meta-analysis: the *dosresmeta* R package. *J Stat Softw*. 2016 Aug 1;72:1-5.
- [4] Crippa A, Orsini N: Flexible modeling. [<http://stats4life.se/webapp/flexmod>]
- [5] Crippa A, Orsini N: Performing dose-response meta-analysis. [<http://stats4life.se/webapp/dosresmeta>]



Figure 2: QR code for the collection of developed web apps.