Three endpoints of radiobiology
R code documentation

Eugene Demidenko
Dartmouth Medical School, Hanover, NH
eugened@dartmouth.edu

January 27, 2010

1. Introduction

The aim of this notes is to describe the `R` code for statistical estimation of the three endpoints of tumor biology: doubling time (DT), tumor growth delay (TGD), and cancer cell surviving fraction (SF) using longitudinal tumor volume measurements in animal experiments based on the recently published paper Demidenko E (2010) "Three endpoints of in vivo tumour radiobiology and their statistical estimation," *International Journal of Radiation Biology*, vol. 86, No 2, pp. 164–173.

The following are the major program features and assumptions:

1. Several groups of animals are analyzed under one statistical mixed effect model. For example, the data example in the file `tumdat.csv` has 4 groups: Control (0), 3 Gy Radiation (1), 5 Gy Radiation (2), 10 Gy Radiation (3).

2. Tumor volume in control group grows exponentially starting from the time of the treatment (time=0)—straight line on the log scale. Tumor volume in the treated group regains exponential growth after time specified by the user (different for different treatment group).

3. The rates of regrowth in treatment groups equal to the rate of growth in control group. Thus, all groups have the same slope but different intercepts when the linear mixed model is applied.

4. Mixed model assumes that longitudinal observations from the same animal constitute a cluster. All slopes on the log scale are assumed to be the same but intercepts are random (animal-specific) with the mean equal to the group mean.

More complex/nonlinear regrowth models exist (Demidenko 2004, 2006).

2. Get started with R

R is the most popular public-domain statistical package and can be freely downloaded from

http://www.r-project.org/

You can download Windows, Mac or Linux version: (a) click at left on CRAN, (b) pick the site you want to download from, say http://cran.mtu.edu, (c) choose the type of the operating system Linux, MacOS X or Windows, (d) click on base and the package will start downloading as a zip file (remember where you store the zip file, usually it's on the desktop), (e) double click on the zip file and it will unpack and install R on your computer (we suggest accepting the default settings for the beginner). It should create an icon on your desktop. After double clicking on that the R icon should get a picture like this

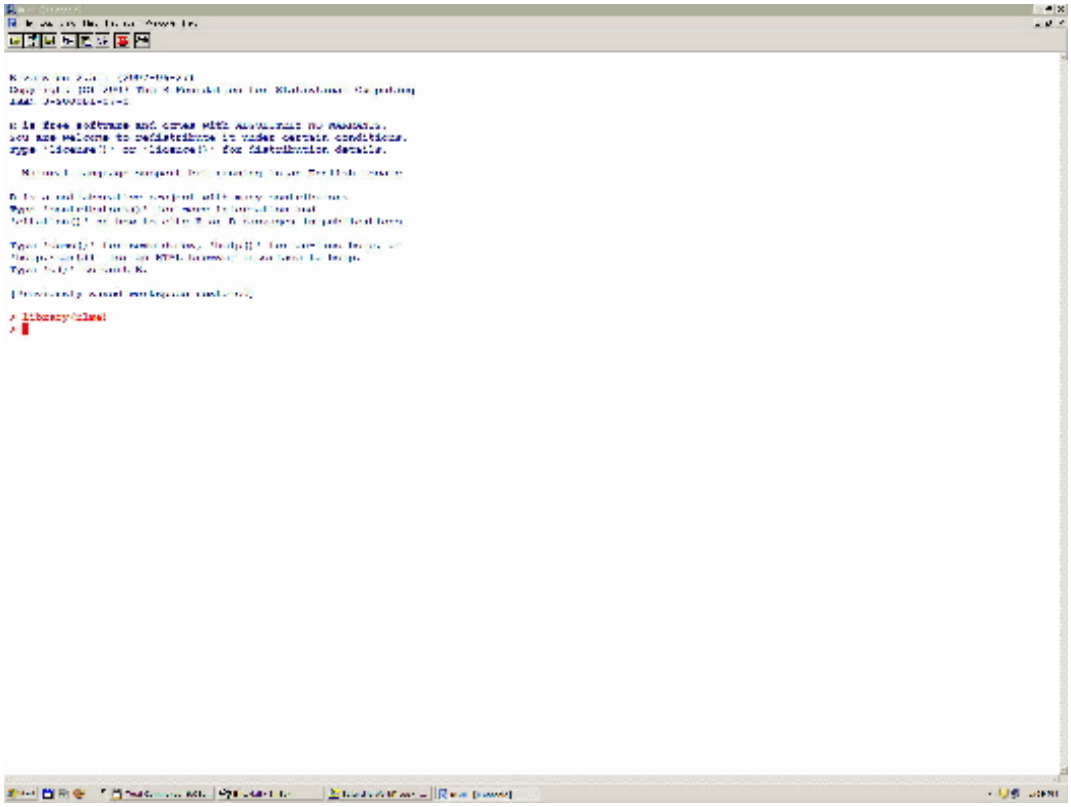

Figure 1: R command window with library nlme attached.

The interested reader may get more information about the R language and statistics with R in the recently published books, Crawley (2007) and Trosset (2009).

2.1. nlme library

In order to run our code `library(nlme)` should be attached, see figure 1. This library contains algorithm for linear mixed effects model used in our `R` code. The zip file nlme.zip should be unpacked into R subdirectory `library\nlme\` before attaching the library from `R`.

2.2. Data file

As was mentioned before, our code can handle several groups of animals, 0 indicates control and 1,2,.. indicate treatment groups. All data must be written in a 4-column table with column #1: Group, column #2: id, column #3: Time, column #4: TumVol. For example, the data example in the Excel file `tumdat.csv` has the form

Group	id	Time	TumVol
0	1	0	120.1
0	1	1	162.8
0	1	2	269.8
0	1	3	470.6
0	1	4	536.6
0	1	5	860.7
0	1	6	795.9
0	1	7	1013.3
.....			
3	28	1	230.8
3	28	2	238.5
3	28	3	273.2
3	28	4	321.4
3	28	5	387.3
3	28	6	601.2
3	28	7	354.7

The data can be prepared as an xls file and saved as csv. The data should have bare numbers (except the names of the columns) and cannot contain characters such NA, missing, space, etc. It is not required to have equal number of animals in each group or to measure tumor volume on the same day although the data in each group should be sufficient enough to run computations.

3. `R.growth` program

Before starting calculations `R.growth` program should be extracted from file `R.growth.r` by issuing

```
source("C:\\Projects\\Rgrowth\\R.growth.r")
```

in command window, assuming that the program `R.growth.r` (and the data) resides in directory

`C:\\Projects\\Rgrowth\\`

Of course, the user can save the files in another directory, then the directories should be changed accordingly. One can edit this program using `R.growth = edit(R.growth)`.

The arguments of `R.growth` program are as follows:

1. `job=1`: plots the tumor volume data on the log scale as a 2x2 panel plot (assuming 4 groups; if the number of groups is less than 4 there will be an empty space); see Figure 2 for our data example.

2. `job=2` plots the mean data and SE. These plots help determine where exponential growth regains; see Figure 3 for our data example.

3. `job=3` actually does statistical estimation using function `lme`. Plots on the log scale help to understand whether the assumption on equal exponential growth is appropriate; see Figure 4 for our data example

4. `regrtime` is a vector that specifies the times when exponential growth regains. The default times are 0 (controls group), 1,3 and 4. The dimension of this vector must be the same as the number of groups in the data file.

5. `groupNames` is a vector of group names, the default is `groupNames=c("Control","3 Gy Radiation","5Gy Radiation","10 Gy Radiation")`. The length of this vector must be the same as the number of groups.

Using the data example `tumdat.csv` saved in directory

`C:\\Projects\\Rgrowth\\`

program `R.growth()` produces graph below (this file and directory should be changed to where the user data reside). Note that the arguments of the program are not specified–the default values are used.

Call `R.growth(2)` creates Figure 3,
and finally call `R.growth(3)` creates Figure 4
with the output in the command window

```
>R.growth(3)
Read 1032 items
Linear mixed-effects model fit by REML
 Data:  NULL
 AIC BIC logLik
 129.9834 153.4795 -57.99169
Random effects:
```

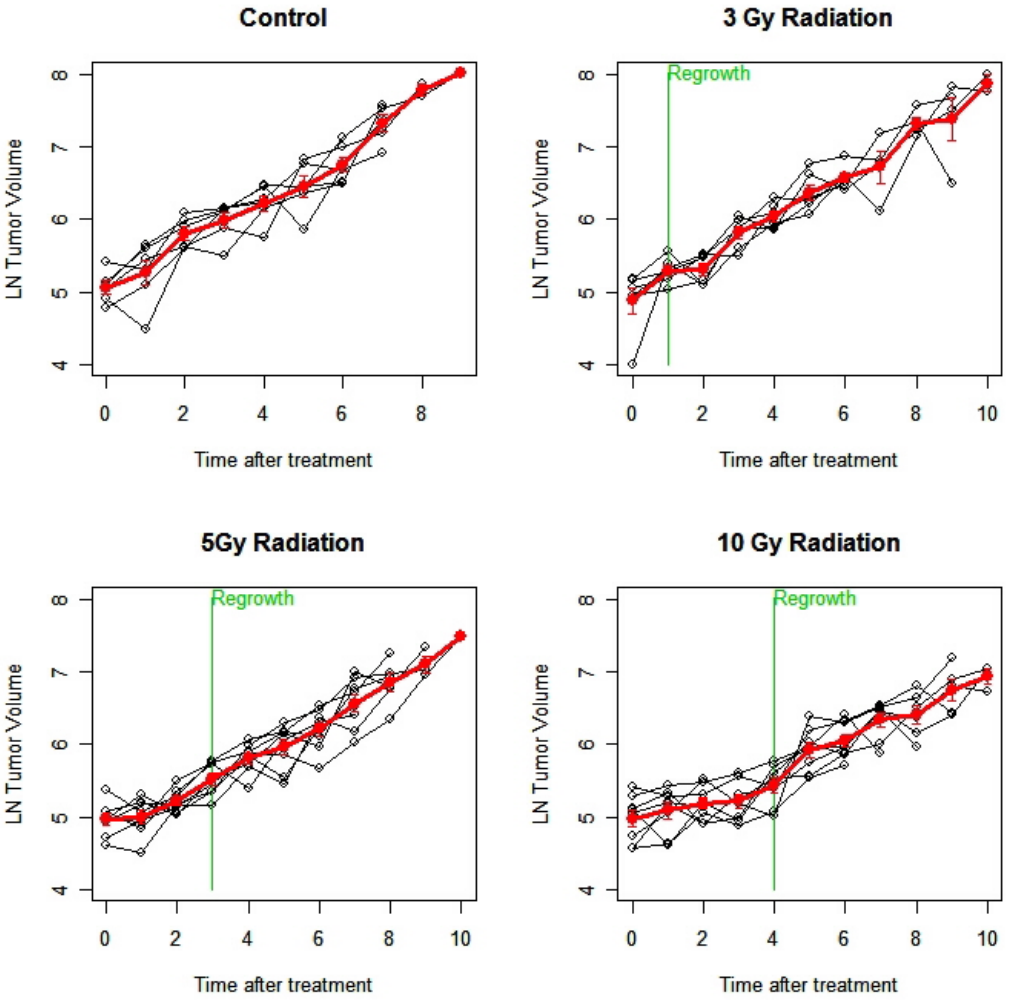

Figure 2: The graph created by `R.growth(1)`.

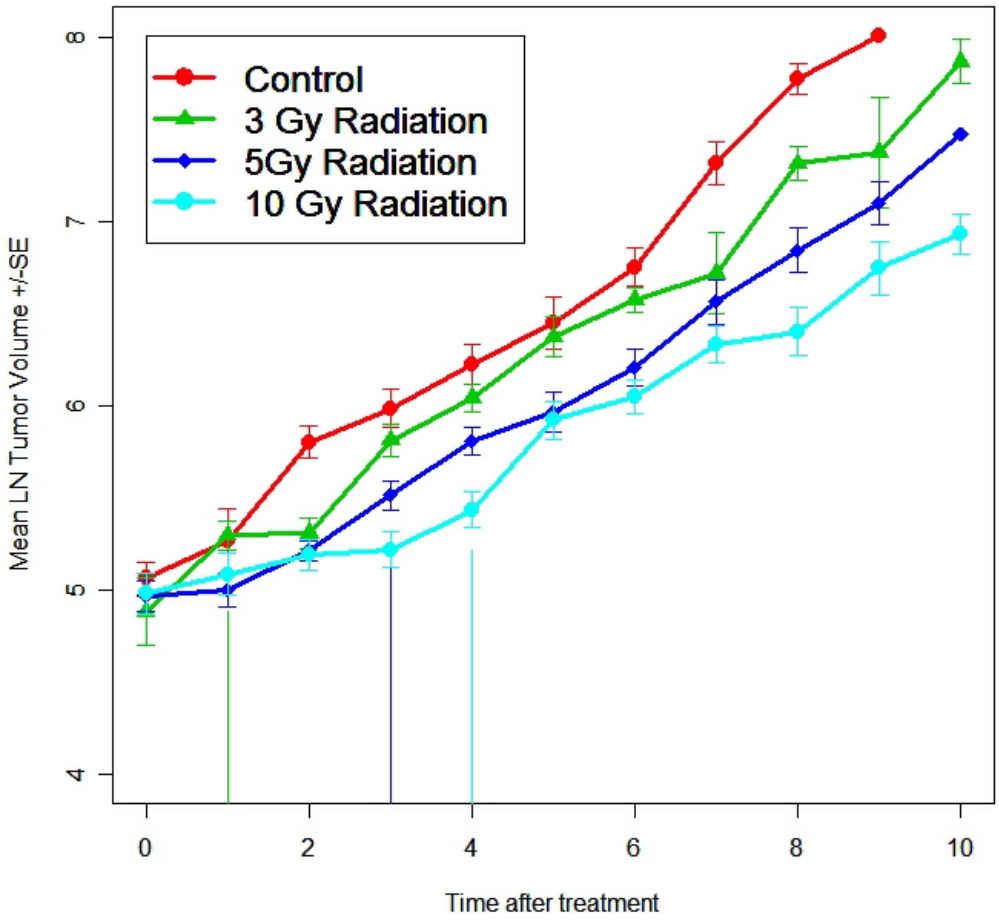

Figure 3: The graph created by `R.growth(2)`. As follows from this graph, empirical curves in each group are fairly parallel and straight that complies with assumption on the exponential regrowth with equal rate of growth.

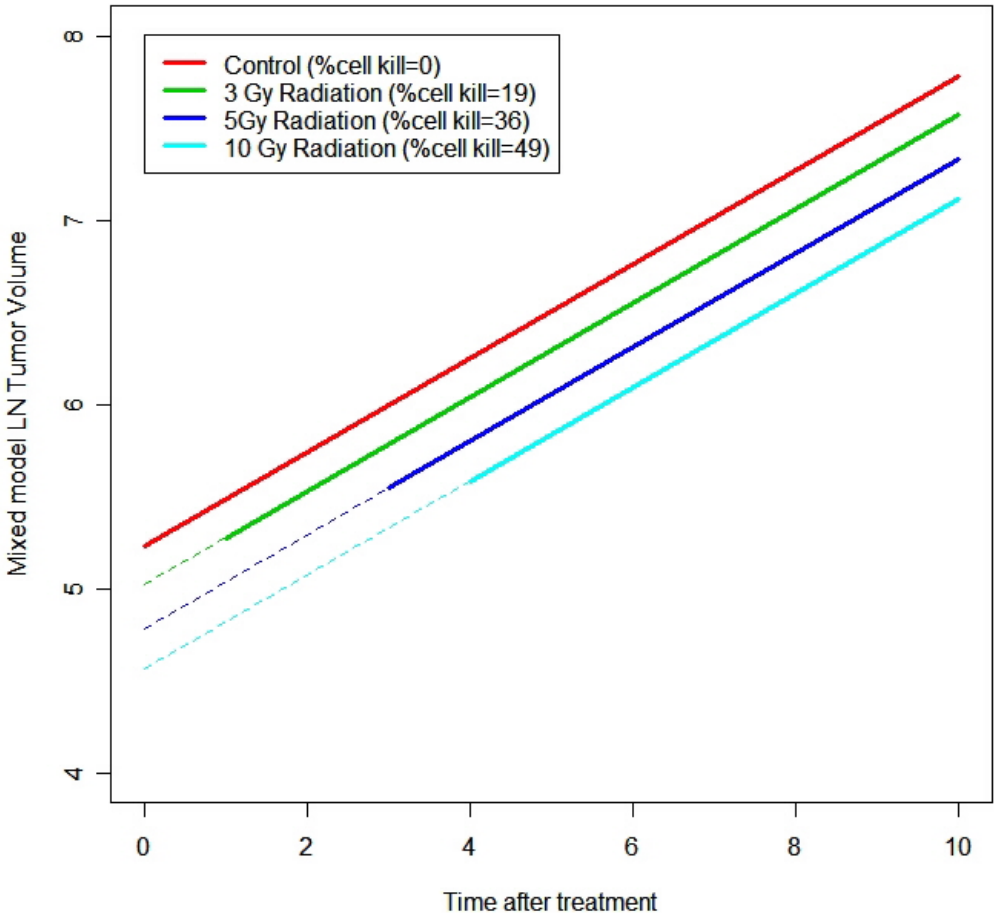

Figure 4: The graph created by `R.growth(3)`. Linear models for 4 groups with the same slope but group-specific intercepts. The distance between lines on the x-axis determines Tumor Growth Delay (TGD), the distance between lines on the y-axis determines ln surviving fraction.

```
 Formula:  ~1 | ind
 (Intercept) Residual
StdDev:  0.1120651 0.2883045
Fixed effects:  y ~X + rate - 1
 Value Std.Error DF t-value p-value
X1 5.233894 0.06695001 24 78.17616 0
X2 5.022340 0.06746347 24 74.44532 0
X3 4.783711 0.06531236 24 73.24358 0
X4 4.562924 0.06875463 24 66.36534 0
rate 0.255225 0.00724770 189 35.21464 0
 Correlation:
 X1 X2 X3 X4
X2 0.182
X3 0.216 0.246
X4 0.229 0.261 0.310
rate -0.399 -0.455 -0.541 -0.573
Standardized Within-Group Residuals:
 Min Q1 Med Q3 Max
-3.3230572 -0.4882690 0.0742184 0.5985161 2.6093493
Number of Observations:  217
Number of Groups:  28
 DT TGD LN SF %SF %KF
Control 2.7158 0.0000 0.0000 100.0000 0.0000
SE Control 0.2421 0.2623 0.0670 6.6950 6.6950
3 Gy Radiation 3.5447 0.8289 -0.2116 80.9325 19.0675
SE 3 Gy Radiation 0.3471 0.3366 0.0860 6.9580 6.9580
5Gy Radiation 4.4797 1.7639 -0.4502 63.7511 36.2489
SE 5Gy Radiation 0.3360 0.3231 0.0828 5.2792 5.2792
10 Gy Radiation 5.3448 2.6289 -0.6710 51.1212 48.8788
SE 10 Gy Radiation 0.3422 0.3275 0.0843 4.3088 4.3088
```

A few comments on the output:

1. `y ~X + rate - 1` specifies the `lme` model, where matrix `X` is the dummy variable matrix to estimate the group-specific intercepts and the common rate as the slope at the time variable. For example, $5.02 - 5.23 = -0.21$ is ln Surviving Fraction in group 3 Gy.

2. Formula `~1 | ind` specifies the random effect.

3. The table at the end of the output takes the results of `lme` estimation and computes endpoints and their standard errors.

Below we present the data in the form close to that in the paper.

Table. Results of `lme` estimation, common rate of tumor volume growth=26% per day.

Group	DT (SE)	TGD (SE)	$\ln SF$ (SE)	$SF\%$ (SE)	$KF\%$ (SE)
Control	2.72 (.24)	0		100	0
3 Gy	3.54 (0.34)	0.83 (0.34)	−0.21 (0.09)	80.9 (7.0)	19.1 (7.0)
5 Gy	4.47 (0.34)	1.76 (0.32)	−0.45 (0.08)	63.8 (5.3)	36.2 (5.3)
10 Gy	5.34 (0.34)	2.62 (0.33)	−0.67 (0.08)	51.1 (4.3)	48.9 (4.3)

These estimates along with standard errors can be used for group comparison. For example to compare 3 Gy group with 5 Gy group in terms of tumor growth delay (TGD) we compute

$$Z = \frac{1.76 - 0.83}{\sqrt{0.34^2 + 0.32^2}} = 1.99$$

with the p-value=0.047.

References

[1] Crawley MJ (2007). *The R Book*. New York, Wiley.

[2] Demidenko E (2004). *Mixed Models: Theory and Applications*. New York, Wiley.

[3] Demidenko E (2006). The assessment of tumour response to treatment, *Applied Statistics*, vol. 55, Part 3, pp. 365–377.

[4] Trosset MW (2009). *An Introduction to Statistical Inference and Its Applications with R*. Boca Raton: CRS Press.