EUR. J. ONCOL.; Vol. 23, n. 2, pp. 86-98, 2018

The analysis of longitudinal data from life-span carcinogenicity bioassays on Sprague-Dawley rats

Daria Sgargi[1], Simona Panzacchi[1], Daniele Mandrioli[1], Fiorella Belpoggi[1], Rossella Miglio[2]

[1] Cesare Maltoni Cancer Research Center, Ramazzini Institute, Bentivoglio 40010, Bologna, Italy; [2] Department of Statistical Sciences, University of Bologna, Italy

Summary. *Background and aim of the work:* Long Term Carcinogenicity Bioassays (LTCB) are among the best instruments to strengthen the evidence on which regulatory agencies base their decision to classify harmful agents as human carcinogens, so they are fundamental to protect public health. The statistical analysis is essential to validate the results from cancer and non-cancer outcomes in carcinogenicity bioassay. This work proposes and applies some methodologies for the analysis of non-cancer outcomes, such as body weights. *Methods:* We use data from studies already concluded, evaluated and published: 4 bioassays aimed at testing the carcinogenic potential of Coca-Cola on Sprague-Dawley rats of different ages. The analysis of body weights of the second generation of rats was performed using mixed-effects models: linear models were fitted for nonlinear models we considered human non-linear growth functions. *Results:* Linear models were fitted using the log-transformation of time and polynomial term of third order for time. Sex and treatment influence body weight, age of dams during gestation doesn't. Growth models: Jenns-Bayley, Count and 1st order Berkey-Reed growth functions were evaluated; the latter best describes the data. Sex and treatment significantly influence all parameters. The direction, magnitude and significance of the effect variable is substantially similar in all models. The analysis of residuals highlights the same issues for all models: the extreme trends in the last part of life heavily affect the models' performance. *Conclusions:* Mixed-effects models allowed to account for the structural effect of covariates that act the same way on all individuals, and to add random effects that introduce a correlation among subjects if clustering happens; nonlinear human growth models added information about the whole growth process, therefore these may be useful methods in studies focused on development and sexual maturation.

Key words: longitudinal analysis, body weights, Sprague-Dawley rats, mixed-effects models, carcinogenicity studies

Introduction

Cancer is a major issue of public health and despite the progress achieved in the prevention and cure of the disease, it is still the second leading cause of death worldwide. The prevalence of risk factors is in fact increasing, including occupational, environmental factors or consisting of dangerous behaviours and lifestyles, such as pollution, smoking, alcohol consumption, obesity and hypertension (2, 3). In this framework, the importance of primary prevention is clear: in terms of public health, the experimental research on environmental and occupational agents is fundamental in order to identify carcinogens and give to national

and international public health agencies adequate data for the necessary regulation.

Epidemiological and experimental studies are the best source of evidence to identify the carcinogenic hazard of a substance and quantify the risk linked to exposure. The most predictive experimental model to anticipate human carcinogens are long term and life-span carcinogenicity bioassays (4-6).

The importance of the statistical analysis of the data obtained through long term studies is generally recognised: it is the necessary complement to establish and quantify whether the long term exposure to selected agents is associated with adverse effects, and it should always be regarded as an integral part of the studies (7). Despite this, not all guidelines explicitly treat in detail and depth the statistical analysis of data (Hothorn, 2014). Different guidelines, mostly from the Organization for Economic Cooperation and Development, illustrate and explain how to choose and perform the appropriate statistical tests, based on the kind of experiment, its objectives and the type of data; they also help to interpret the results and to understand their real meaning and relative importance (3, 5). To maintain coherence with the established methods in toxicology, the classical frequentist approach and the concept of hypothesis testing are adopted; the methods are systematically organized into a flowchart, proposing tests to verify the significance of differences between the treated and the control groups, according to the nature of the data. Consolidate and advanced methodologies exist for the direct assessment of carcinogenicity, specifically developed to handle peculiarities of these data and answer specific research questions. However, additional information that are routinely collected in experimental studies (such as body weights, feed and fluid consumption, the time of survival in life-span studies, etc…) are only used to monitor the conduct of the study and the health status of animals; no specific statistical method to treat them is suggested, so they are rarely analysed in depth.

Many methodologies nowadays exist to analyse these data, and are relatively easy to implement, thanks to different accessible statistical software: applying them would allow to fully use all the available data and to integrate them, reaching an overall more complete information on the effects of the tested compound on health.

The aim of this work is to better exploit the potential of all available data on non-cancer outcomes from carcinogenicity studies to strengthen the knowledge of the tested substances. In particular, the objectives of this study are to examine one of the most common type of non-cancer outcomes, the body weight of experimental animals, to find appropriate methodologies to analyse its characteristics, and to apply them on some real data, in order to verify their suitability.

Materials and Methods

The data for this analysis were obtained from studies performed in 1986 at the Cesare Maltoni Cancer Research Centre (CMCRC) of the Ramazzini Institute, aimed at evaluating the possible association between continuous consumption of Coca-Cola and effects on tumour incidence in rodents. The soft drink was chosen as a test substance because of its widespread diffusion, the known effects of sweetened beverages on weight and the growing awareness of the importance of obesity as a risk factor for several types of tumours.

Four experiments were performed, each involving male and female Sprague-Dawley rats starting exposure at different ages (breeding rats of 30, 39 and 55 weeks of age; all their offspring of all litters, whose observation started at 8 weeks of age; and young non-breeding rats of 7 weeks of age). The experimental plan is schematically reported in Table 1. The soft drink was administered to rats ad libitum as a substitute of drinking water from the beginning of observation for the whole lifespan, until spontaneous death.

Here, we will focus on the second generation of rats: treated female and male breeders started to drink Coca-Cola one week before mating, while the control group was administered with tap water; dams continued the exposure during the whole period of the pregnancy and the weaning. After weaning, offspring continued to drink Coca-Cola ad libitum and from 8 weeks of age they were weighted and controlled for feed and beverages consumption until spontaneous death. All pups from all litters were included, in the same experimental group as their breeders, so rand-

Table 1. Experimental plan of the four bioassays performed for the project: treatments, age at beginning of observation and number of animals by sex for each experimental group.

Treatment	Age at start	M	F
Coca-Cola	7 weeks	80	80
Drinking water	7 weeks	100	100
Coca-Cola	55 weeks	70	70
Drinking water	55 weeks	70	70
Coca-Cola	Prenatal (offspring)	28	24
Drinking water	Prenatal (offspring)	32	24
Coca-Cola	30 weeks	55	55
Drinking water	30 weeks	55	55
Coca-Cola	Prenatal (offspring)	74	73
Drinking water	Prenatal (offspring)	110	98
Coca-Cola	39 weeks	110	110
Drinking water	39 weeks	110	110
Coca-Cola	Prenatal (offspring)	67	65
Drinking water	Prenatal (offspring)	49	55

omization was not used for the second generation; the data can therefore be considered clustered.

The experiments were planned and performed following the standard procedures of the CMCRC and in compliance with international guidelines; for a detailed presentation of the experimental plan, conduct and results of the analysis of tumour incidences, see the original publication from Belpoggi et al. (8).

This work is focused on the analysis of longitudinal measurements of body weight, one of the best indicator for the good conduct of chronic rodent bioassays; moreover for this particular test compound, a sugar-sweetened beverage, the body weight is an endpoint of particular interest, as well as a very important indicator of metabolic, hormonal and homeostatic functions, growth and sexual maturation (9).

Guidelines (5) suggest to graphically represent groups means to keep track of the indicatiors of the animals' well-being during the experiments; then, formal analysis should start checking the assumptions of normality, homogeneity of variance and absence of outliers, required for the subsequent analyses; in case the assumptions are not met, some solutions such as the log-transformation of data are suggested. Finally, several types of tests are proposed, to evaluate the differences between groups (Student's t-test or modified t-test with Satterthwaite's method, or ANOVA and pairwise comparisons). A concise representation of the suggested analyses can be seen in Figure 1.

This approach has some clear drawbacks: using overall summary measures instead of all available data for each individual in time may cause a loss of information. Furthermore, it is impossible to evaluate the effect of more variables at a time.

The methods that might be more suitable for analysing this kind of data (clustered longitudinal data) are based on mixed-effect models. Laird and Ware (10) were the first to propose a flexible class of mixed models for longitudinal data: it includes both growth and repeated-measures models as special cases, and it introduces population parameters, individual effects and within-subject variation, as well as between-subject variation (11, 12).

In their representation, the $n * 1$ vector of responses for the ith subject can be modelled as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \qquad i=1,\ldots N$$

where

- $\boldsymbol{X}_i$ is a $n_i * p$ design matrix of explanatory variables or fixed factors;
- $\boldsymbol{\beta}$ is a $p * 1$ vector of unknown population parameters, or fixed effects coefficients, describing the relationships between the outcome and the explanatory variables for groups defined by levels of a fixed factor (for example, describing the contrast between males and females);
- $\boldsymbol{Z}_i$ is a $n_i * q$ design matrix of variables of random factors;
- $\boldsymbol{b}_i$ is $q * 1$ a vector of unknown random effects specifically referred to a given level of a random factor, usually representing the deviations from the relationships described by fixed effects. Random effects can be set as random intercepts (random deviations for an individual or cluster from the overall fixed intercept), or as random coefficients (random deviations for an individual or cluster from the overall fixed effects). They are assumed to follow a multivariate normal distribution ~ $N(\mathbf{O}, \boldsymbol{D})$ with D being a $q * q$ symmetric, positive definite variance-covariance matrix;
- $\boldsymbol{\varepsilon}_i$ is a $n_i * 1$ vector of errors for the ith subject for each measurement occasion, whose terms do not

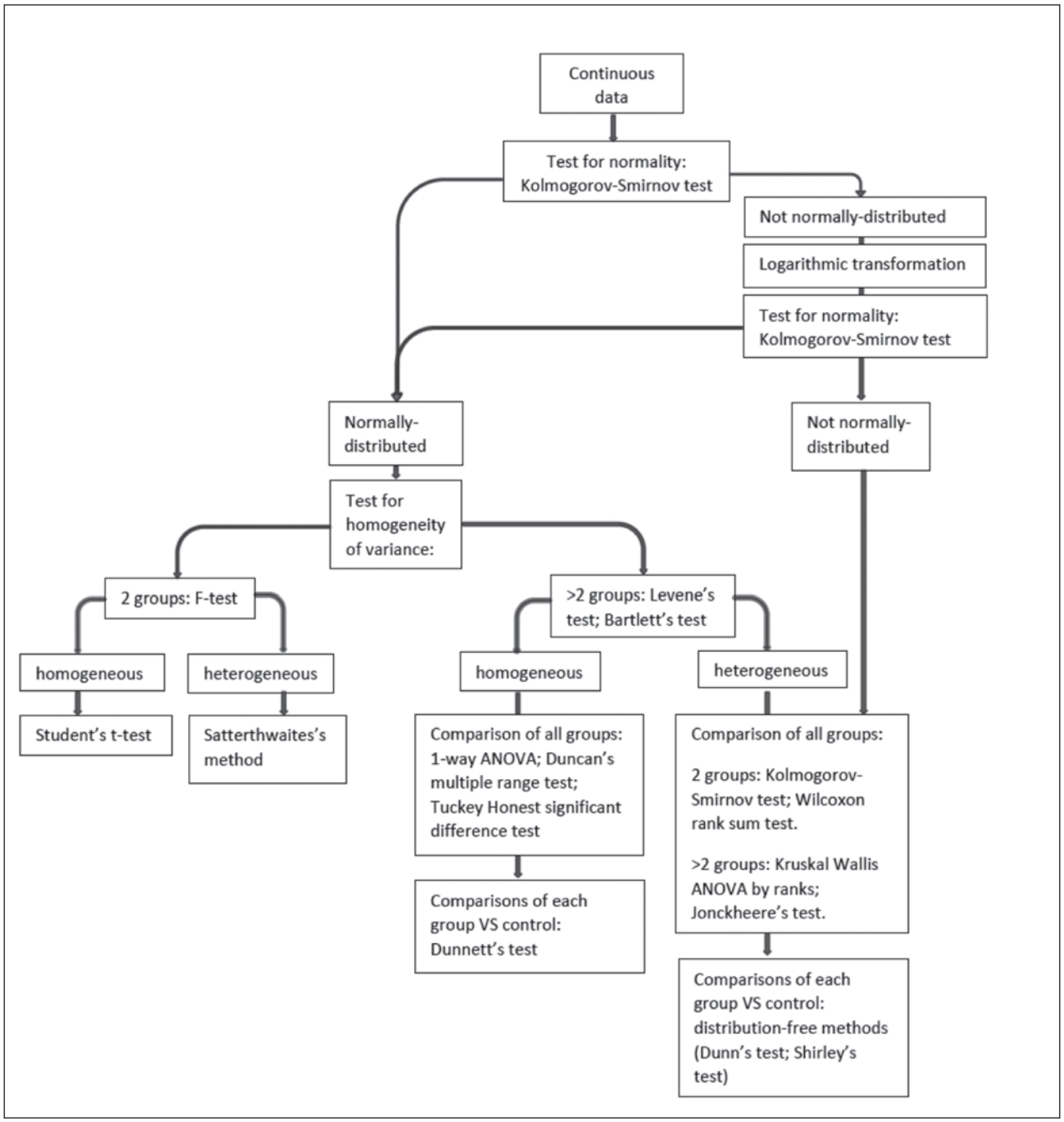

Figure 1. OECD statistical decision tree summarizing the suggested procedures for the analysis of continuous data. (Author's adaptation from OECD, 2012)

need to be independent but can be correlated within individuals. The residuals for each subject follow, again, a multivariate normal distribution ~ $N(\boldsymbol{O}, \boldsymbol{R}_i)$, with 0 mean and a positive definite $q * q$ variance-covariance matrix, $\boldsymbol{R}_i$.

Several covariance structures can be specified both for $\boldsymbol{D}$ and $\boldsymbol{R}_i$. Inference on the parameters' estimates can be based on least squares and maximum likelihood methods, or, formulating the model the appropriate way, using an empirical Bayesian method (13, 14).

The response variable is here assumed to present a linear and continuous trend, so the function included in the model to represent its relationship with independent variables through fixed and random effects is a linear one. Often, however, the trajectory of individual growth presents discontinuities or shows a nonlinear path. Several adaptations of the linear mixed effects model may be adopted to cope with these situations, the most common being:

- splitting the time of analysis into sub-periods, so that the linearity assumption is reasonable within each sub-model;
- identifying a suitable transformation of the outcome variable or the time scale;
- representing time as a polynomial function (15).

All these methods share a characteristic: they still imply a linear association that models the relationship between the outcome and the explicative variables. Often, however, the likelihood function depends on the parameters in a non-linear way: in such cases, the use of nonlinear models is justified by the possibility to obtain a more interpretable model and to use a smaller number of parameters.

Lindstrom and Bates (16) were the first to present a general, nonlinear mixed effects model for data in which the assumption of the normality of residual holds, but the expectation function is nonlinear. The model can be written as

$$y_{ij} = f(x_{ij}, \beta, u_i) + \varepsilon_{ij}, \qquad i = 1,\ldots,N$$

where f is a real-valued function x_{ij} is a vector of covariates containing both within- and between-subjects covariates, β is a $q * 1$ vector of unknown parameters of fixed effects, u_i is a vector of unobservable subjective random parameters following a multivariate normal distribution with 0 mean and variance-covariance matrix Σ, and ε_i is the error vector of dimension $n_i * 1$, following a multivariate normal distribution with 0 mean and variance-covariance matrix $\sigma^2 \Lambda$.

The two-stages representation of the model (17) helps to clarify how the non-linear function is used to express the individual trajectory of change at level 1

$$y_{ij} = m\left(x_{ij}^{w}, \varphi_i\right) + \varepsilon_{ij},$$

where m describes the behaviour of the individual growth as depending on individual-specific parameters φ_i and the vector of within-subject covariates x_{ij}^{w}, while the inter-individual variability can be expressed using a regular linear relationship at level 2:

$$\varphi_i = d\left(x_{ij}^{b}, \beta, u_i\right)$$

where d is a vector function that explains the variation of individual-specific parameters between subjects and incorporates β, the vector of parameters for the population, and x_{ij}^{b}, the set of between-subjects covariates. The assumptions underlying the non-linear mixed effects model are that the random effects u_i and the error terms ε_i are independent between each other and across individuals, that $\sigma^2 > 0$ and that matrix Σ is definite nonnegative.

Choosing the correct functional form to specify the relationships at level 1 is very important to obtain credible and accurate models. For this analysis a peculiar approach was experimented.

The growth and maturation processes have long been studied in humans, and several models have been proposed to formalize their patterns during infancy, childhood and adolescence (18). It is well established that the pattern of growth of body dimensions of the "general type" (to be distinguished from those of lymphoid, neural and genital type) from birth to the adult age is increasing and S-shaped, since it progresses rapidly in the first years, then slows down, and accelerates again around the so-called pubertal spurt, and finally approaches a plateau when the approximate adult size is reached. Since many similarities can be recognized in the growth path of humans and rats, some of these human models have been translated, adapted and applied here to rats.

Several structural regression models based on an adequate parametric function were developed in time to represent the different phases of growth: these functions can be substituted to the generic function at the level 1 of the multilevel mixed effect model, since they describe the behaviour of each individual.

The first parametric model was elaborated already in 1937 by Jenss and Bayley (19): it was developed to describe growth from birth to approximately 8 years using 4 parameters combined in a function with a lin-

ear and an exponential part, accounting for growth and its decreasing rate:

$$y = a + b\,t - - e^{c+d\,t}$$

Another option is the Count model (20) proposed in 1943, that uses only 3 parameters combined in a linear way

$$y = a + b\,t + c\,\ln(t+1)$$

This model proved to perform slightly worse than the Jenss and Bayley, but both remain robust relative to the choice of starting values for the parameters. The Count model was later modified by Berkey and Reed (21) maintaining the simple, linear structure but adding one or two parameters:

$$1^{st}\ order:\ y = a + b\,t + c\,\ln(t+1) + \frac{d}{t}$$

$$2^{nd}\ order:\ y = a + b\,t + c\,\ln(t+1) + \frac{d_1}{t} + \frac{d_2}{t^2}$$

accommodating for one or two additional inflexion points and leading to a better fit, compared to the previous alternatives.

More complex models were developed to represent different phases of growth at the same time. It was showed (22) that rat's and human developmental phases are similar but growth rhythms differ, particularly in early phases, so it was chosen not to consider the models that were designed to account for the specific features and mechanisms of human growth, but rather to focus on those that could be used to describe a similar path, in terms of intensity and velocity, to the one of humans during young age, and adapt them to the available data from rats.

Results

All statistical and graphic analysis have been performed using the statistical software StataIC 15.

It was chosen to consider for analysis only the measurements taken until 114 weeks of age of rats, because after this timepoint the number of rats alive was considerably reduced (at 114 weeks of age, 52.5% of the animals were lost to follow-up; at 122 weeks of age, the following measurement, 91.4% were lost), and rats can be considered very old at this age (it is not possible to draw exact parallels, but it is well accepted (23) that 104 weeks of age in rats are comparable to around 65 years in humans).

The graphical analysis of individual and mean weights showed a high variability, both within and among subjects. All trends were quite similar in shape during the first period of growth, while during the adult/elderly period some peculiar patterns appeared: weights tended to decrease in the last part of life, because of diseases or the physiologic ageing process; some animals, on the other hand, experienced a rapid increase of weight due to the onset of mammary neoplastic lumps.

Given these characteristics that prevent the use of linear functions *tout court*, three options were evaluated for analysis: estimating linear mixed-effects models (I) using a mathematical transformation for the time variable, (II) using polynomial functions to represent time, to obtain an approximately linearized growth trajectory for each individual, and (III) fitting mixed-effects model using nonlinear "human" growth functions at the individual level. In all cases, the variables reflecting experimental conditions (such as sex, treatment regimen, and age of the dam at the beginning of gestation) were evaluated as fixed effect, while random intercept and slopes depending on time were introduced at the litter level, allowing correlation between the random slopes and intercepts.

I. Linear mixed-effect model, mathematical transformtation of time variable

Sex and the treatment regimen were found to significantly influence body weights; males ware sensibly heavier than females since the first observation; the treatment was responsible of a smaller but still relevant (p= 0.021) increase in weight since the first assessment. These results are summarised in Table 2; Figures 2 and 3 show a graphical representation of the average growth trajectories for treated and control animals, in male and female rats, and some example of the specific trajectories estimated for randomly selected litters, respectively.

Residuals were analysed to verify whether the underlying basic assumptions (linearity of the relationship between the outcome and the regressors, normality and homoscedasticity of residuals) were met: the unpredictable variability in the last phase of life was, as expected, very high.

II. Linear mixed-effect model, polynomial representation of time

The best option to represent time in this context is a third order polynomial. Results regarding the direction and magnitude of the effect of each variable on body weight and its overall relevance, are substantially assimilable to those obtained in the previous analysis. Again, the unconstrained consumption of Coca-Cola was associated with a significant (p= 0.016) increase in body weight, as illustrated in Table 3 and in Figures 4 and 5. According to the goodness of fit measures, anyway, the previous specification should be preferred. The analysis of residuals highlighted again that the problems arising from the extreme values of weights of elderly or ill rats remained quite evident: residuals had some issues concerning normality, and appeared quite heteroscedastic in relation to time, too.

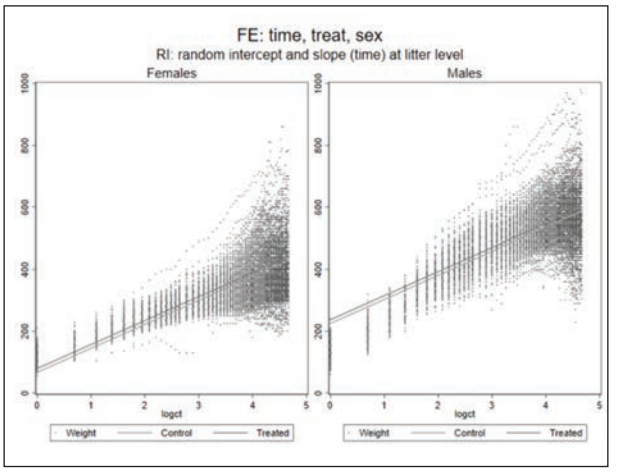

Figure 2. Plot of body weights (in grams) and estimated average predictions obtained using a linear multilevel mixed effects model (fixed-effects predictors: natural logarithm of time since first observation (in weeks), sex and treatment).

III. Nonlinear mixed-effects models, human growth functions

After comparing the performances of models built with the Jenns and Bayley, the Count and the 1st order Berkey and Reed growth functions, the Berkey

Table 2. Linear multilevel mixed effects model using transformed variables; results from the regression of body weights (in grams) on the natural logarithm of time since first observation (in weeks), sex and treatment

Mixed-effects REML regression of body weight					
Fixed-effects Parameters	Coef.	Std. Err.	P>|z|	[95% Conf. Interval]	
log(time)	77.69	1.98	0.000	73.81	81.57
Treatment	12.17	5.29	0.021	1.80	22.54
Sex	158.16	0.54	0.000	157.10	159.22
_constant	67.01	4.51	0.000	58.16	75.86
Random-effects Parameters	Estimate	Std. Err.		[95% Conf. Interval]	
Litter: Unstructured					
sd(log(time))	19.4	1.43		16.79	22.42
sd(_cons)	35.88	3.19		30.14	42.72
corr(log(time),_cons)	-0.71	0.06		-0.81	-0.57
sd(Residual)	47.55	0.18		47.21	47.9

Goodness-of-fit Measures			
Log-likelihood	-193723.6	*df*	8
AIC	387463.1		
BIC	387531.2		

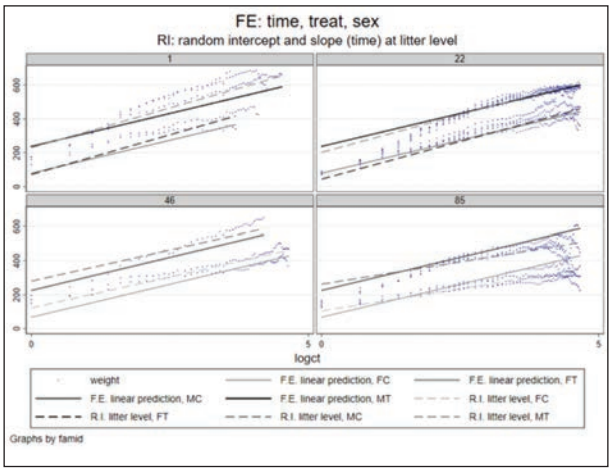

Figure 3. Plot of body weights (in grams) and linear predictions with fixed and random part in four randomly selected litters; results obtained using a linear multilevel mixed effects model (fixed-effects predictors: natural logarithm of time since first observation (in weeks), sex and treatment).

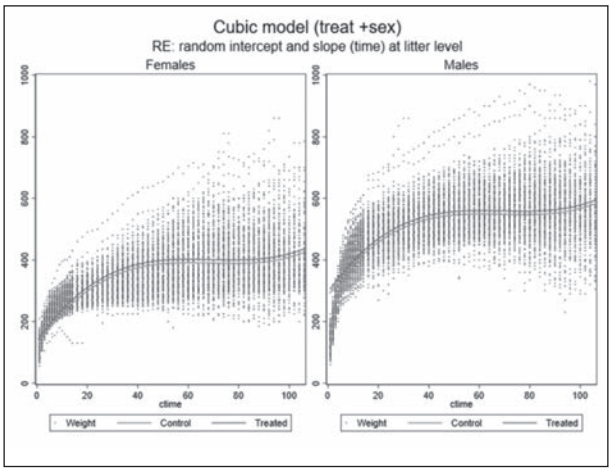

Figure 4. Plot of body weight (in grams) measurements and estimated average predictions, linear multilevel mixed effects model using a third degree polynomial term to represent time since first observation (in weeks)

Table 3. Linear multilevel mixed effects model using polynomial representation of time; results from the regression of weight (in grams) on third degree polynomial term for time since first observation (in weeks), sex and treatment.

Mixed-effects REML regression					
Fixed-effects Parameters	Coef.	Std. Err.	P>|z|	[95% Conf. Interval]	
Time	11.92	0.19	0.000	11.56	12.29
time^2	-0.17	0.00	0.000	-0.18	-0.17
time^3	0.00	0.00	0.000	0.00	0.00
Treatment	13.42	5.59	0.016	2.46	24.38
Sex	158.13	0.57	0.000	157.01	159.25
_constant	123.43	3.88	0.000	115.83	131.03
Random-effects Parameters	Estimate	Std. Err.		[95% Conf. Interval]	
Litter: Independent					
sd(time)	1.64	0.15		1.37	1.95
sd(time2)	0.03	0.003		0.02	0.03
sd(time3)	0.0002	0.00002		0.0001	0.0002
sd(_cons)	26.47	2.16		22.55	31.07
sd(Residual)	49.76	0.18		49.39	50.12

Goodness-of-fit Measures			
Log-likelihood	-195757.7	*df*	11
AIC	391537.4		
BIC	391631		

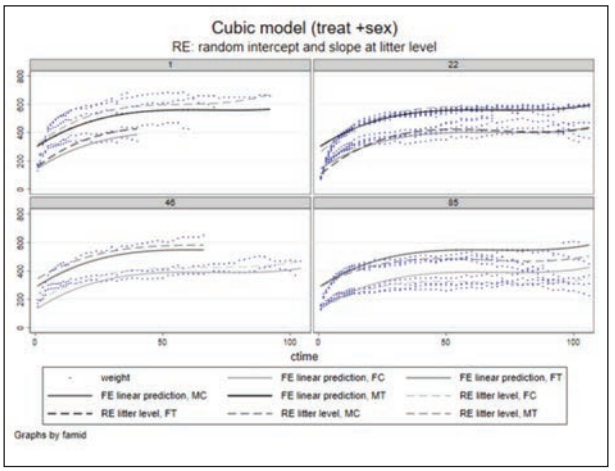

Figure 5. Plot of body weights (in grams) and linear predictions with fixed and random part in four randomly selected litters, linear multilevel mixed effects model using third degree polynomial to represent time since first observation (in weeks).

and Reed was chosen as the most appropriate for these data. The model building in this case was quite different from the previous ones: it uses four parameters to describe the specific functional form of the individual growth curve

$$y = a + b\,t + c\,\ln(t+1) + \frac{d}{t}$$

that may respectively represent the starting point, growth rate, acceleration and deceleration of growth. Here, the function was built so that each parameter would be dependent on sex and treatment regimen; a random effect at the litter level was introduced and evaluated for all of them, so that each litter is not constrained to have, for example, the same intercept or inflexion points. The estimated parameters are not as easy to interpret, as in the previous cases; we can resume, anyway, that the previous results are confirmed (the individual change depended on age, while the inter-individual change was described using sex and the treatment regimen as fixed effects, and a random effect at the litter level, to explain the variations in each parameter of the model). The estimates and their graphical representation are displayed in Table 4, and in Figures 6 and 7. Nevertheless, the problems of excessive variability in the tail of the growth trajectories remained relevant, as expected.

Discussion

The aim of this work was to go beyond the standard statistical techniques that are routinely used in experimental carcinogenicity studies to analyse non-cancer endpoints. We proposed and applied some methodologies that encompass the use of the different approaches, instead of summary measures, in order to answer the research questions in a more comprehensive way.

It is difficult to directly compare the results with those of the original publication, since different approaches were adopted. The use of mixed-effects models allowed to use every measurement available from each individual animal: this was important given the features of the data, that presented a consistent random variability, mostly in the last part of the animals' lives. Furthermore, it allowed to account for the structural effect of covariates that act the same way on all individuals, and to add random effects that introduce a correlation among subjects, accounting for clustered data. Finally, applying nonlinear human growth functions allowed to consider the change of body weight in time as a process, instead of a generic series of measurements: this can be useful in studies whose aim is to characterise possible variations in the development and the sexual maturation linked to the exposures under analysis.

In this study, where the tested compound was a highly caloric and sweetened beverage, body weights are of primary interest, and it would be even more interesting to deepen the analyses evaluating them in association with tumour incidence. Indeed, the unconstrained consumption of Coca-Cola in this experiment was associated with a significant increase in body weight (the statistically significant result emerges from all estimated models: linear mixed-effect model, logarithmic function of time: p= 0.021; linear mixed-effect model, cubic function of time: p= 0.016; non-linear mixed-effect model, Berkey-Reed growth function: p= 0.000 for all parameters) and it is well established that overweight and obesity are positively associated with the increase of the risk of many types of cancer (24-26). They were not explicitly considered here, but these analyses are suitable for rats observed from an adult age, as well; for this purpose, a linear model or the in-

Table 4. Nonlinear multilevel mixed effects model using Berkey-Reed function to represent individual growth; results from the regression of weight (in grams) on age (in weeks), sex and treatment.

Mixed-effects ML non-linear regression						
Fixed-effects Parameters		Coef.	Std. Err.	P>|z|	[95% Conf. Interval]	
a	sex	565.79	33.21	0.000	500.69	630.89
	treatment	-380.53	33.3	0.000	-445.8	-315.25
	_constant	655.28	28.32	0.000	599.77	710.79
b	sex	-0.15	0.12	0.221	-0.39	0.09
	treatment	-0.84	0.19	0.000	-1.23	-0.46
	_constant	1.57	0.15	0.000	1.28	1.86
c	sex	78.15	9.41	0.000	59.7	69.6
	treatment	-118.26	9.44	0.000	-136.77	-99.75
	_constant	89.06	8.01	0.000	73.35	104.77
d	sex	-3381.29	137.82	0.000	-3651.41	-3111.17
	treatment	1330.01	138.13	0.000	1059.28	1600.74
	_constant	-3125.87	117.84	0.000	-3356.83	-2894.92
Random-effects Parameters			Estimate	Std. Err.	[95% Conf. Interval]	
Litter: Identity						
	var(U0)	0.57	0.08		0.43	0.75
	var(Residual)	1966.77	14.56		1938.44	1995.51

Goodness-of-fit M**easures**			
Log-likelihood	-191030	*Df*	14
AIC	382088		
BIC	382207.1		

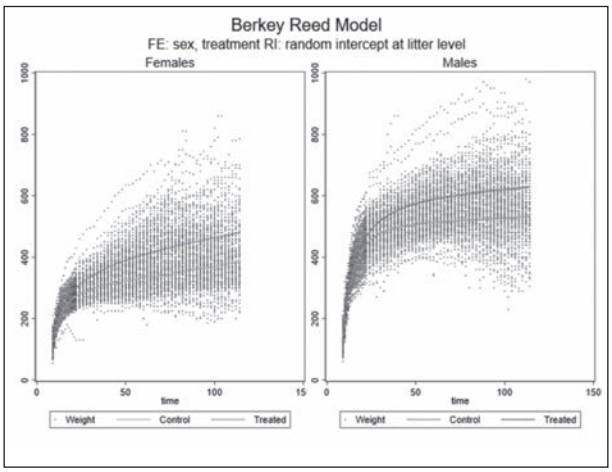

Figure 6. Plot of body weights (in grams) and estimated average predictions, nonlinear multilevel mixed effects model using Berkey-Reed function to represent individual growth

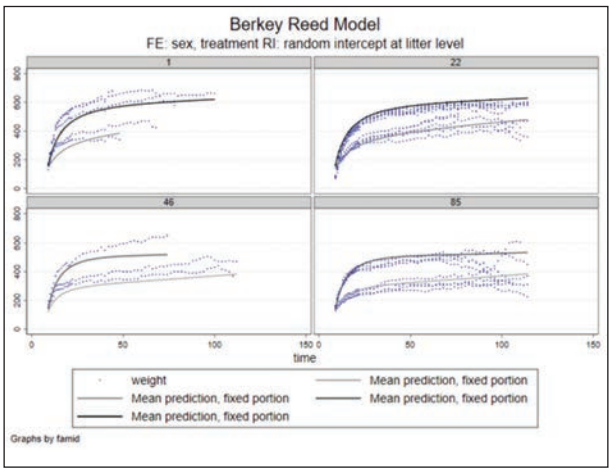

Figure 7. Plot of weights (in gram) and linear predictions with fixed and random part in four randomly selected litters, nonlinear multilevel mixed effects model using Berkey-Reed function to represent individual growth

clusion of a quadratic term to represent time should be the best options, since rarely growth models are built to model the weight trajectories during the whole lifespan of individuals, so a simpler and more efficient alternative is to be preferred.

A possible alternative to build flexible yet simple models are General Additive Models: it could be interesting to evaluate their performance in this context, since they may allow to handle data with such an irregular trend, and at the same time to maintain a simple and understandable interpretation for the regression parameters.

Some issues remain open, like the problem of how to handle the extreme trends that some animals showed in the last part of their life; they represent an interesting feature, that is usually associated with ageing and the onset of pathological conditions (for example mammary tumours increase the individual weight, other tumours decrease individual weight).

Conclusions

Continuous experimental longitudinal data, in particular those consisting in body weights, have some very peculiar characteristics similar to the human counterpart, the most relevant for their analysis are non-linearity and the fact that they can take unexpected, extreme turns upwards or downwards, mostly when rats are close to the end of their life, reflecting the presence of large neoplastic mammary lumps or a worsening of the health conditions due to ageing. These features should discourage the use of methods based on the comparison of measures of synthesis like the group means, because they could be heavily affected by the atypical recordings, giving an unrealistic picture of the situation and possibly preventing to detect subtler differences caused by experimental factors.

The use of multilevel mixed effects models is therefore to be encouraged, since they allow to analyse directly the recordings of each subject, without concerns about the differences in the duration of the follow-up. They are also a precious tool in case of clustered data, like in this rather peculiar experimental design, where no randomization was performed on a whole cohort of rats of second generation. The most straightforward specification of such models using a proper linear function isn't the best option because it requires a transformation of the variables, so the advantage of a simple functional form is counterbalanced by the difficult interpretation of the transformed variables. Even the introduction of polynomial terms, that allow to represent a curve trajectory remaining in the frame of a linear function, slightly reduces the ease of the interpretation, but it can still be acceptable in case it allowed a more faithful representation of the growth trajectories; as these analyses showed, nevertheless, it's not always the case. New tools in this field, that may have potential advantages, are the nonlinear growth models that were "borrowed" from human studies: a wide variety exists, so one can select the most appropriate every time, according to the characteristics of data. The model's parameters are not always easy to interpret, but a clear indication is provided about the direction, magnitude and statistical significance of the effect of each covariate.

As a more general recommendation, deepening the knowledge and understanding of the methods used for the statistical analysis of experimental results is an important strategy to enhance the quality of the research, in particular for toxicology (27). This work is an attempt in this direction: it aims to go beyond the statistical techniques that are routinely used, to explore the characteristics of the data and to try to understand the mechanisms that determined them. In this framework, some methodologies to answer the research questions in a more comprehensive way were proposed and applied to the carcinogenicity bioassay on a sweetened beverage (Coca Cola) performed by the Ramazzini Institute. All estimated models confirm that the unconstrained consumption of Coca-Cola in this experiment was associated with a significant increase in body weight (linear mixed-effect model, logarithmic function of time: $p= 0.021$; linear mixed-effect model, cubic function of time: $p= 0.016$; non-linear mixed-effect model, Berkey-Reed growth function: $p= 0.000$ for all parameters).

To conclude, it's worthy to remind once more the importance of properly choosing the methods and specifying the models, where properly means in a data-and-experience-driven way. A thorough knowledge of the data and of the dynamics that contribute to determine them is always a good starting point to

build plausible, representative and meaningful models. Another crucial point is the fact that model checking and verification of the respect of the assumptions that lie at the foundations of any method, should become a routine embedded in every analysis, while it still remains not so common (or, at least, not always explicitly reported) in the literature regarding carcinogenicity bioassays, in particular for non-cancer outcomes.

Ultimately, the use of more adequate statistical models helps refine and reduce the use of experimental animals, in line with the EU Directive 2010/63/EU (28).

Author's contribution:
Daria Sgargi: concept and design of study, data collection; Daria Sgargi, Simona Panzacchi, Daniele Mandrioli: data interpretation and analysis, drafting, revision, approval of final manuscript.; Rossella Miglio and Fiorella Belpoggi: supervision of data interpretation and analysis, critical revision of the entire text, approval of final manuscript. This work is adapted from the PhD Thesis of Daria Sgargi "The Analysis of Survival and Longitudinal Data from Life-span Carcinogenicity Bioassays on Sprague-Dawley Rats", Department of Statistical Science, University of Bologna.

References

1. Global Burden of Disease Cancer Collaboration. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015. A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2017; 3 (4): 524–548
2. International Agency for Research on Cancer. Preamble. In: IARC monographs on the Evaluation of Carcinogenic Risks to Humans. WHO- IARC, Lyon. Last update September 2015
3. Bucher J R. The National Toxicology Program Rodent Bioassay. Designs, Interpretations, and Scientific Contributions. Ann N Y Acad Sci 2002, 982: 198-207
4. Organization for Economic Cooperation and Development. Test No. 451: Carcinogenicity Studies. OECD Publishing, Paris, 2009
5. Organization for Economic Cooperation and Development. Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting Test Guidelines 451, 452 and 453. OECD Publishing, Paris; 2nd edition, 13 April 2012
6. Shockley K R, Kissling G E. Statistical Guidance for Reviewers of Toxicologic Pathology. Toxicol Pathol 2018, 46(6): 647-652
7. Hothorn L. Statistical evaluation of toxicological bioassays-a review. Toxicol Res, 2014; 3: 418-432
8. Belpoggi F, Tibaldi E, Soffritti M, Falcioni L, Bua L, Trabucco F. Results of long-term carcinogenicity bioassays on Coca-Cola administered to Sprague-Dawley rats. Ann NY Acad Sci 2006; 1076: 736-752
9. Organization for Economic Cooperation and Development. Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies. Paris, OECD, 2002
10. Laird N M, Ware J H. Random effects models for longitudinal data. Biometrics 1982; 38: 963–974
11. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York Springer, 2000
12. Fitzmaurice G, Molenberghs G. Advances in longitudinal data analysis: an historical perspective. In Davidian M, Verbeke G, Molenberghs G. Fitzmaurice G. Longitudinal data analysis: a Handbook of modern statistical methods, 3-30. London, Chapman & Hall/CRC Press, 2008
13. Pinheiro J, Bates D. Mixed-Effects Models in S and S-PLUS. New York, Springer, 2000
14. West B T, Welch K B, Galecki A T. Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition. London, Chapman & Hall/CRC Press, 2010
15. Singer J D, Willet J B. Applied longitudinal data analysis. Modeling change and event occurrence. Oxford, Oxford University Press, 2003
16. Lindstrom M, Bates D. Nonlinear mixed-effects models for repeated measures data. Biometrics 1990; 46, 3: 673-668
17. Demidenko E. Mixed Models: Theory and Applications with R, 2nd Edition. Hoboken, Wiley, 2013
18. Hauspie R C, Cameron N, Molinari L. Methods in Human Growth Research. New York, Cambridge University Press, 2004
19. Jenns R M, Bayley N. A mathematical method of studying the growth of a child. Hum Biol 1937, 9: 556-563
20. Count E. Growth pattern of the human physique. Hum. Biol. 1943, 15: 1–32
21. Berkey CS, Reed RB. A model for describing normal and abnormal growth in early childhood. Hum Biol. 1987 Dec; 59 (6): 973-87
22. Sengupta P. The laboratory rat: Relating its age with human's. Int J Prev Med 2013; 4: 624-30
23. Soffritti M, Belpoggi F, Degli Esposti D. Cancer Prevention: the lesson from the lab. In: Biasco G, Tanneberger S. Cancer Medicine at the Dawn of the 21st Century: the view from Bologna 49-64. Bologna, Bonomia University, 2006
24. Calle E E, Rodriguez C, Walker-Thurmond K, Thun M J. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. Adults. N Engl J Med, 2003: 1625-38
25. Rapp K, Schroeder J, Klenk J, Stoehr S, Ulmer H, Concin H, Diem G, Oberaigner W, Weiland SK. Obesity and incidence of cancer: a large cohort study of over 145000 adults in Austria. Br. J. Cancer 2005; 93: 1062–1106
26. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer – viewpoint of

the IARC Working Group. N Engl J Med, 2016; 375:794-798
27. Mandrioli D, Silbergeld E K. Evidence from Toxicology: The Most Essential Science for Prevention. Environ Health Perspect. 2016 Jan; 124 (1): 6–11
28. Eurpean Parliament and Council. Directive 2010/63/UE of Eurpean Parliament and Council, of September 22nd, 2010, on the protection of animals used for scientific purposes (Text with EEA relevance) OJ L 276, 20.10.2010, p. 33–79

Correspondence:
Fiorella Belpoggi,
Cesare Maltoni Cancer Research Center, Ramazzini Institute
Via Saliceto, 3, 40010 Bentivoglio, Bologna, Italy
Tel. +39 051 6640460
Fax +39 051 6640223
E-mail: belpoggif@ramazzini.it