JOURNAL OF APPLIED QUANTITATIVE METHODS

Applied Quantitative Methods in Medicine

# SUBJECT-LEVEL TREND ANALYSIS IN CLINICAL TRIALS

**Alexandru-Ionut PETRISOR**
PhD, Assistant Professor, Sections Urbanism and Landscape, School of Urbanism,
"Ion Mincu" University of Architecture and Urbanism, Bucharest, Romania

**E-mail:** alexandru_petrisor@yahoo.com
**Web-page:** http://www.geocities.com/a_petrisor

**Liviu DRAGOMIRESCU**
PhD, Associate Professor, Department of Ecology,
Faculy of Biology, University of Bucharest, Romania

**E-mail:** liviu_dragomirescu@yahoo.com

**Cristian PANAITE**
PhD Candidate, Intern, Diabetes Nutrition and Metabolic Diseases,
N. Paulescu Institute, Bucharest, Romania Romania

**E-mail:** cristian.panaite@gmail.com

**Alexandru SCAFA-UDRISTE**
PhD Candidate, University Assistant, „Carol Davila" University of Medicine and Pharmacy,
Department of Internal Medicine and Cardiology,
Clinic Urgency Hospital, Bucharest, Romania

**E-mail:**

**Anamaria BURG**
PhD, Assistant Professor,
University of South Carolina, Lancaster, S.C., USA

**E-mail:** a_rusu@hotmail.com

**Abstract:** *In particular situations, clinical trials researchers could have a potential interest in assessing trends at the level of individual subjects. This paper establishes a common approach and applies it in two different situations, one from nutritional medicine and one from cardiovascular medicine. The approach consists of running as many regression models as the number of subjects, looking at the behavior of some parameter of interest in time. The regression parameters, particularly the slope of the regression line, offer the general sense of the trend and allow for testing its statistical significance. Extrapolation at the level of the entire sample is possible using some version of the binomial test. In both cases, significant results were obtained despite of small sample sizes.*

**Key words:** *simple linear regression; slope; coefficient of determination; binomial test*

## 1. Introduction

Clinical trials are used to evaluate new drugs or treatments, including new technologies, assess new screening programs, or ways to organize and deliver health services [1][1]. Epidemiologists distinguish three types of clinical trials: prophylactic - used to prevent diseases, therapeutic - used to treat diseases, and interventional - used to intervene before the disease is developed [2],[3]. Whilst most often clinical trials, similar to other epidemiologic studies, analyze the impact of intervention on the development of a certain disease, assessing the impact of potential risk factors, researchers might be occasionally interested to analyze trends at the level of the subject.

Due to the fact that there are only several important moments when the parameters of interest are checked (beginning of the study, midterm, endpoint and eventually some other intermediary moments), there are too few data to use a time series model. Nevertheless, the evolution of the variable of interest (weight in the first case, in the second) in time translates into a simple linear regression model.

For each regression model, several parameters of interest describe the trend. The sign of the regression slope, $\beta$, indicates either a decreasing or increasing trend. Whereas a test of significance for $\beta$ could pinpoint significant trends in some patients, the limited number of time milestones results into a general lack of significance. The very few significant trends cannot allow for further analyses. The same behavior characterizes the coefficient of determination, $R^2$.

To assess the overall trend, a sign test could be used to assess whether the trend is decreasing (a percentage significantly larger than 50% or some other value of the regression slopes are negative) or positive. If sufficient data are available, the values of the regression slopes could be used in conjunction with their significance test and trends can be classified as either significantly decreasing, not significant, or significantly increasing.

Two examples of applying the proposed approach are presented in this paper. In the first case, while comparing the efficiency of three weight loss programs, a question of interest is whether weight loss is consistent during the period when the treatment is administered. Weight is checked at some intervals, and for each patient the efficiency should translate into a continuous decrease of weight. In the second example, two echocardiographic parameters (the ejection fraction and the kinetic score) were analyzed in relationship in a clinical study of the Acute Myocardial Reperfusion Syndrome looking at risk factors, predictors and criteria assessing the success of interventions.

## 2. Methods

### 2.1. Statistical tests

The steps taken in applying the proposed methodology are:

1. For each subject, run simple linear regressions according to the model $Y=\alpha+\beta\times TIME$, where Y is the dependent variable monitored in the study.
2. For each model, record the slope of the regression line, $\beta$, or the coefficient of determination, $R^2$. Also, test for their significance [6]:

$$t_0 = \frac{\beta}{\sqrt{\frac{\sum_{i=1}^{n}\left(Y_i-\left(\alpha+\beta\cdot X_i\right)\right)}{n-2}}\sqrt{\sum_{i=1}^{n}X_i^{\,2}-n\cdot\left(\sum_{i=1}^{n}X_i\Big/n\right)^2}} \approx t_{n-2} \quad \text{, or } t_0 = R\sqrt{\frac{n-2}{1-R^2}} \approx t_{n-2}.$$

3. Define an indicator variable to describe the trend as either:
   I=-1, if $\beta<0$; 1, if $\beta>0$; and no value otherwise, if $\beta$ was found significant in very few cases, or
   I=-1, if $\beta<0$ and $p\leq0.05$; 1, if $\beta>0$ and $p\leq0.05$; and 0 otherwise, if the number of subjects with a significant trend is large enough to allow for further statistical testing.
4. To analyze the overall trend, run the binomial test to compare the proportions of subjects with I=1 and, respectively, -1 (derived from Piegorsch and Bailer [7]):

$$z = \frac{p - 0.5}{0.5/\sqrt{n}} \approx N(0,1)$$

   *p* is the proportion of subjects with I=1 or -1. 0.5 is the proportion corresponding to the null hypothesis $H_0$: there is no overall trend (the proportions of subjects with I=1 and, respectively, -1 are equal, and each of them is 0.5).
5. The indicator variable can be used in the Analysis of Co-Variance or logistic regression, either as dependent or independent variable, depending upon the interest of the researcher.

**2.2. Software implementation**

The steps described above were implemented in SAS. In order to use SAS, data were stored in an array with the following columns: TIME, the time when each observation was recorded; CLASS, any classification variable (identifying the group to which subjects belong, if any), and S01, S02, ..., S0n, an identifier of each subject (could be automatically generated in Excel). Given this structure, the SAS code is provided below (comments are inserted between accolades {}).

```
data name_of_dataset;
input time class S01 S02 ... S0n;
{Paste actual data here}
;
proc reg;
model S01 = time / influence; {The "influence" option was used in the second example to see if
observations recorded at some particular time are more relevant for diagnosis; if there is no interest in
testing it, do not use the "/ influence" statement}
by class; {Should analyses be run only for a particular class, use "where class = class_level" instead; if
there are no classes, do not use any statement}
{Repeat the statements starting with "proc reg;" for S02, ..., S0n}
run;
```

From the SAS output, retrieve for each regression model the values of ß and corresponding p-values for the test of $H_0$: ß = 0 vs. $H_A$: ß ≠ 0). Store all these in Excel in three columns and define the following based on Excel functions, replacing SIGN and BETA with corresponding column names:

The sign of ß: SIGN= IF(BETA<0,"MINUS",IF(BETA=0,"ZERO","PLUS"))

The significance of ß: SIGNIFICANCE=IF(p<0.05,"S","NS")

To compute the proportion of subjects with ß > 0 (respectively ß < 0), use: X=COUNTIF(RANGE,"PLUS"), respectively X=COUNTIF(RANGE,"MINUS"), where X is the position of the cell where the result of the formula is computed and RANGE corresponds to FIRST CELL:LAST CELL of the column where the sign of ß is stored.

The sign test can also be computed in Excel using:

Y=(X/N-0.5)/(0.5/SQRT(N)), where N is the total number of observations.

## 3. Results and Discussion

### 3.1. Example #1: Weight loss

Data had been produced by the study "Efficiency of the intensive nutritional, pharmacologic and behavioral management of obesity – correlation of genetic, bio-morphological and psychological factors" (National University Research Council grant #163 of 2006). The aim was to compare the efficiency of three weight loss programs among 84 subjects: classical intervention (24 subjects), intensive intervention assisted by nutritionists (33 subjects), and intensive intervention assisted by psychologists (27 subjects). The later two categories were joined in a group labeled "intensive interventions" (60 subjects).

In this case, the variable of interest was actual weight, recorded in the beginning of the study (0 months), during the study, after 1 and respectively 4 months, and in the end of the study (12 months). Its values were not recorded uniformly, and the actual sample sizes were diminished (*Table 1*).

**Table 1.** Subject level linear regression coefficients of weight variation in time: Bucharest, 2008

| **Classical intervention** | | | **Intensive intervention assisted by nutritionists** | | | **Intensive intervention assisted by psychologists** | | |
|---|---|---|---|---|---|---|---|---|
| β | t(β) | p(t) | β | t(β) | p(t) | β | t(β) | p(t) |
| -0.52 | 5.74 | 0.1099 | -0.54 | 0.85 | 0.5532 | -0.75 | 1.14 | 0.4581 |
| 3.64 | -0.87 | 0.5456 | -1.09 | 1.48 | 0.3781 | 3.18 | -0.83 | 0.5584 |
| -2.26 | 1.62 | 0.3514 | -0.66 | 4.41 | 0.1419 | -2.91 | 0.79 | 0.5739 |
| -7.01 | 3.56 | 0.1743 | -0.89 | 6.41 | 0.0985$^{M}$ | -1.12 | 18.04 | 0.0352$^{*}$ |
| -12.00 | . | . | -1.91 | 1.61 | 0.3537 | 5.64 | -0.71 | 0.6088 |
| -2.94 | 2.24 | 0.2674 | -0.92 | 2.61 | 0.2329 | -2.12 | 3.74 | 0.1664 |
| -3.39 | 1.31 | 0.4162 | -0.79 | 5.9 | 0.1068 | -0.27 | 9.63 | 0.0658$^{M}$ |
| -40.00 | . | . | -0.79 | 1.02 | 0.4923 | -1.36 | 59.47 | 0.0107$^{*}$ |
| -1.18 | 2.12 | 0.2801 | -2.40 | . | . | -0.55 | 1.53 | 0.3686 |
| 3.00 | . | . | -0.54 | 6.31 | 0.1000 | -0.00 | . | . |
| -0.00 | . | . | -8.00 | . | . | -0.07 | 0.03 | 0.9807 |
| -0.00 | . | . | -0.77 | 1.45 | 0.3842 | -8.00 | . | . |
| -1.63 | 49.65 | 0.0128$^{*}$ | -1.88 | 1.47 | 0.3798 | -3.95 | 1.58 | 0.3600 |
| 1.05 | -0.12 | 0.9268 | -2.21 | 4.84 | 0.1298 | -0.81 | 1.87 | 0.3126 |
| -0.65 | . | . | -2.69 | 1.67 | 0.3440 | -0.57 | 3.08 | 0.1999 |
| -0.13 | 0.07 | 0.9534 | -0.54 | 5.94 | 0.1062 | -0.97 | 3.05 | 0.2017 |
| -2.64 | 20.4 | 0.0312$^{*}$ | -1.73 | 5.67 | 0.1112 | -3.47 | 3.99 | 0.1564 |
| -0.00 | . | . | -1.15 | 1.97 | 0.2997 | -0.69 | 0.59 | 0.6613 |
| -2.78 | 2.89 | 0.2123 | 2.68 | -0.3 | 0.8160 | -6.67 | 2.89 | 0.2123 |
| -2.86 | . | . | -0.50 | 0.83 | 0.5573 | -1.31 | 1.22 | 0.4380 |
| 120.00 | . | . | -2.11 | 2.19 | 0.2727 | -0.57 | 0.72 | 0.6047 |
| -3.47 | 1.85 | 0.3154 | -1.09 | 1.71 | 0.3376 | -3.72 | 4.1 | 0.1523 |
| -3.56 | 0.30 | 0.8143 | -5.69 | 4.13 | 0.1512 | -1.11 | 36.52 | 0.0174$^{*}$ |
| 13.33 | . | . | -0.55 | 3.45 | 0.1796 | -1.00 | 14.64 | 0.0434$^{*}$ |
| | | | -0.69 | 3.72 | 0.1674 | -1.78 | . | . |
| | | | -1.79 | 4.04 | 0.1544 | -1.24 | 3.61 | 0.1720 |
| | | | -0.67 | 0.99 | 0.5030 | -1.10 | 0.96 | 0.5122 |
| | | | -0.47 | 3.09 | 0.1993 | | | |
| | | | -0.40 | 0.33 | 0.7987 | | | |
| | | | -0.78 | 27.67 | 0.0230$^{*}$ | | | |
| | | | -0.72 | 4.26 | 0.1468 | | | |
| | | | -0.80 | 0.78 | 0.5769 | | | |
| | | | -0.76 | 60.62 | 0.0105$^{*}$ | | | |

**Notes:** p-values use a modified Michelin scale, adding marginal significance to the uncertainty region ($0.05 \leq p \leq 0.1$): $^{*}$ significant, $^{**}$ highly significant, $^{M}$ marginally significant

The overall trend was very highly significantly decreasing ($p<0.001$), with slight variations among subjects assigned to the classical intervention ($p=0.025$), intensive intervention assisted by nutritionists ($p<0.001$), and intensive intervention assisted by psychologists ($p<0.001$). Comparisons among groups suggested that intensive intervention methods are more efficient with respect to weight loss than the classical ones ($p=0.006$ when joining interventions assisted by nutritionists and psychologists, 0.027 otherwise), but no significant differences were detected between interventions assisted by nutritionists and psychologists ($p=0.388$).

The results were checked for consistency with findings using a traditional approach, employing the Analysis of Variance (ANOVA) to test whether there are significant differences between the three groups. In this case, we defined the weight loss as the difference between the initial ($t=0$) and final ($t=12$) weight. The global F test was $F=10.88$ with $p<0.001$, suggesting the existence of significant differences between groups. No differences were found between interventions assisted by nutritionists and psychologists, but both of them differed significantly from the classical approach, being more efficient.

### 3.2. Example #2: Clinical meaning of echocardiographic parameters

Data were generated within the Acute Myocardial Reperfusion Syndrome study "BNP – prognostic value of BNP correlated with echocardiographic indices of systolic and diastolic function in patients with ST elevation acute myocardial infarction with indication of reperfusion" (National University Research Council grant #22 of 2006) looking at risk factors, predictors and criteria assessing the success of interventions.

In this case, the variables of interest were the ejection fraction, defined as the difference between end-diastolic and end-systolic volumes divided by end-diastolic volume [4], and the kinetic score, computed according to the guidelines of the American Society of Echocardiography as the ratio between the sum of scores assigned to each segment of the left ventricle (1 = normal; 2 = hypokinetic; 3 = akinetic; 4 = dyskinetic) and their number, *i.e.* 16 [5]. Both values were recorded in the beginning of the study (0 days), during the study, after 1, 7 and respectively 30 days, and in the end of the study (365 days). Even though 88 subjects were included in the study, values of the ejection fraction and kinetic score were not always recorded, and the actual sample size was diminished.

The sign test did not detect any trend with respect to the ejection fraction ($p=0.198$), but detected a significantly decreasing trend of the kinetic score ($p=0.003$).

However, in this particular study further research questions that could be answered using the proposed methodology. Among them, of particular interest was the predictive value for screening purposes of recording a value of either or both the ejection fraction and/or the kinetic score at one of the particular moments used (0, 1, 7, 30 or 365 days). In order to answer this question, we assigned ranks from 1 to 5 to the pair (ejection fraction, time) and (kinetic score, time) with a maximum impact on the regression line, corresponding to its position on the time scale: 1, for t=0; 2, for t=1; 3, for t=7; 4, for t=30; and 5, for t=365.

The magnitude of the impact was assessed based on the jackknife residual [6]:

$$r_{(-i)} = \frac{y_i - (\alpha + \beta \cdot x_i)}{\sqrt{\sigma_{(-i)}^2 \left(1 - \frac{1}{n} + \frac{(x_i - \mu_x)^2}{\sum_{i=1}^{n} x_i^2 - n\mu_x^2}\right)}} \approx t_{n-3}$$

In situations where the jackknife residuals could not be computed, ranks were determined based on the value of the actual residuals [6]:

$$e_i = Y_i - (\alpha + \beta \cdot X_i)$$

The maximum impact corresponded to the maximum absolute value of either the jackknife residual or actual residual, within each set of five values computed for each patient.

In order to test whether some moment has a higher predictive value, we tested the statistical significance of the difference between the proportion of its corresponding rank and 0.2 (proportion under the null hypothesis), using a modified version of the test proposed by Pagano and Gavreau [8]:

$$z = \frac{|p - 0.2|}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Results suggest that for both variables the first moment (t=0) appears to be the most important, as its corresponding p-values are the lowest (*Table 2*). Nevertheless, for the ejection fraction none of the moment was significant at 0.05. Due to the reduced sample sizes, we accounted for additional marginal significance if p values were in the uncertainty area ($0.05 \leq p \leq 0.1$). In the same order of importance, the second moment is t=365 for the ejection fraction and t=30 for the kinetic score.

The usage of two variables in the second study allowed for checking the validity of the proposed approach. Since the ejection fraction and the kinetic score were very highly significantly correlated ($R^2$=-0.74 overall, -0.62 at t=0, -0.71 at t=1, -0.75 at t=7, -0.74 at t=30 and -0.84 at t=365, with $p<0.001$ in all cases), we assessed the correlation at the subject level looking at the correlation of the ranks described above. Spearman's coefficient of correlation was 0.189 with p=0.086, falling into the uncertainty region. This could be due to reducing the overall sample as the computation of ranks was not always possible since five values were not always recorded for each variable for individual subjects.

**Table 2.** Tests of the predictive value of the moment when the ejection fraction and kinetic score are recorded: Bucharest, 2008

| **Ejection fraction (n=81)** | | | | **Kinetic score (n=86)** | | | |
|---|---|---|---|---|---|---|---|
| **Rank** | **Frequency** | **z** | **p** | **Rank** | **Frequency** | **z** | **p** |
| 1 | 0.28 | 1.63 | 0.0516$^{M}$ | 1 | 0.32 | 2.33 | 0.0099$^{**}$ |
| 2 | 0.20 | 0.05 | 0.4801 | 2 | 0.15 | 1.31 | 0.0951$^{M}$ |
| 3 | 0.17 | 0.63 | 0.2643 | 3 | 0.11 | 2.55 | 0.0054$^{*}$ |
| 4 | 0.21 | 0.21 | 0.4168 | 4 | 0.12 | 2.09 | 0.0183$^{*}$ |
| 5 | 0.14 | 1.62 | 0.0526$^{M}$ | 5 | 0.30 | 1.90 | 0.0287$^{*}$ |

**Notes:** p-values use a modified Michelin scale, adding marginal significance to the uncertainty region ($0.05 \leq p \leq 0.1$): $^{*}$ significant, $^{**}$ highly significant, $^{M}$ marginally significant

## 4. Conclusion

Both examples indicate that the proposed approach was able to produce significant results despite of the reduced sample sizes. Comparisons with classical approaches, answering partially the same research question, suggest that the proposed methodology is valid and could be used to answer the particular question of assessing subject-level trends is

clinical trials. The only disadvantage is that it employs a large number of analyses, problem resolved by employing appropriate software.

## References

1. Bernard, Y. **L'échocardiogramme de stress**, Revue Médicale Suisse, no. 2258, June 2, 1999, online source: http://revue.medhyg.ch/print.php3?sid=19884, accessed on March 7, 2009
2. Gordis, L. **Epidemiology**, Philadelphia, PA: W. B. Saunders Company, 1996, pp. 90
3. Habash-Bseiso, D. E., Rokey, R., Berger, C. J., Weier, A. W. and Chyou, P.-H. **Accuracy of Noninvasive Ejection Fraction Measurement in a Large Community-Based Clinic**, Clinical Medicine and Research 3, no. 2, 2005, pp. 7-82
4. Hussey, J. R. **BIOS 757: Intermediate Biometrics. Course notes/Spring 1998**, Columbia, SC: School of Public Health, University of South Carolina, 1998, pp. 9,15, 29, 30
5. Lilienfeld, A. M. and Lilienfeld. D. E. **Foundations of Epidemiology,** 2nd ed., New York, NY: Oxford University Press, 1980, pp. 257-258
6. Mausner, J. S. and Kramer, S. **Mausner & Bahn Epidemiology – An Introductory Text**, Philadelphia, PA: W. B. Saunders Company, 1985, pp.195
7. Pagano, M. and Gauvreau, K. **Principles of Biostatistics**, 1st ed., Belmont, CA: Duxbury Press, 1993, pp. 296
8. Piegorsch, Walter W. and John A. Bailer **Statistics for Environmental Biology and Toxicology**, 1st ed., London, UK: Chapman and Hall, 1997, pp. 12

[1] Codification of references:

| | |
|---|---|
| [1] | Gordis, L. **Epidemiology**, Philadelphia, PA: W. B. Saunders Company, 1996, pp. 90 |
| [2] | Lilienfeld, A. M. and Lilienfeld. D. E. **Foundations of Epidemiology,** 2nd ed., New York, NY: Oxford University Press, 1980, pp. 257-258 |
| [3] | Mausner, J. S. and Kramer, S. **Mausner & Bahn Epidemiology – An Introductory Text**, Philadelphia, PA: W. B. Saunders Company, 1985, pp.195 |
| [4] | Habash-Bseiso, D. E., Rokey, R., Berger, C. J., Weier, A. W. and Chyou, P.-H. **Accuracy of Noninvasive Ejection Fraction Measurement in a Large Community-Based Clinic**, Clinical Medicine and Research 3, no. 2, 2005, pp. 7-82 |
| [5] | Bernard, Y. **L'échocardiogramme de stress**, Revue Médicale Suisse, no. 2258, June 2, 1999, online source: http://revue.medhyg.ch/print.php3?sid=19884, accessed on March 7, 2009 |
| [6] | Hussey, J. R. **BIOS 757: Intermediate Biometrics. Course notes/Spring 1998**, Columbia, SC: School of Public Health, University of South Carolina, 1998, pp. 9,15, 29, 30 |
| [7] | Piegorsch, Walter W. and John A. Bailer **Statistics for Environmental Biology and Toxicology**, 1st ed., London, UK: Chapman and Hall, 1997, pp. 12 |
| [8] | Pagano, M. and Gauvreau, K. **Principles of Biostatistics**, 1st ed., Belmont, CA: Duxbury Press, 1993, pp. 296 |