Using Explained Variance Allocation to analyse Importance of Predictors

Henri Wallard

CEDRIC, Centre d' Etude et de Recherche en Informatique et Communications.
CNAM CEDRIC, 292 rue Saint Martin ,75141 Paris , France.
Email :henriwallard@hotmail.com

Abstract. Using applications of linear regression, Market Research practitioners want to determine a ranking of predictors or a quantification of their respective importance for a desired outcome. As predictors are often correlated, regression coefficients can be difficult to use directly because they can be instable across samples and have negative values that are counterintuitive. To overcome these difficulties other methods have been proposed in the industry using squared semi partial correlation coefficients, squared zero order correlation coefficients or methods such as Shapley Value decomposition or decomposition via orthogonalisation in the space of predictors.

The proximity between the results obtained by different Variance Decomposition methods has led some authors to conclude that they are a fully valid approach. This paper will highlight theoretical reasons why these methods present similarities, offer a simple alternative new way to decompose variance but will also show the flaws and risks of relaying on Variance Decomposition for quantification of importance of predictors and why a Game Theory approach like Shapley Value can lead to misinterpretations. It will also present additional methods developed to compute β coefficients using Variance Decomposition as an intermediate step and propose recommendations for driver analysis.
Keywords: Variance Decomposition, Regression, Importance.

1 Introduction

In the field of Market Research, practitioners want to help their clients identify how to act on some factors such as quality of service or design of a product to achieve a desirable outcome such as purchase intent or satisfaction and loyalty of the customers. This is done with the desire to identify what are the best drivers of improvement, and quantify their respective impact. This is achieved through statistical modeling and simulation like in other fields such as Psychology, Social Sciences or Economics. The classic reference method used to do this Ordinary Least Square regression (OLS), but while this approach is recommended in the business literature it has some limitations. Because of sample size and design of the questions in Market research surveys, there may occur instability of coefficients. And also as questions can relate to similar topics there may be multi-collinearity and negative coefficients that are counter-intuitive. Causal assumptions and modeling options can lead to a variety of

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

results when one wants to quantify or simulate the impact of a given action on a predictor on the desired response. In Market Research many techniques are used: Regressions, Path Models, Bayesian Belief Networks, Random Forest to name a few. This paper focusses on relative importance in the context of Linear Regression.

2 The concept of relative importance

Johnson [1] points out that the terminology of relative importance is confusing because of the many different definitions used, and introduced the term "relative weights" to define the proportion of explained variance by the linear model allocated to each individual predictor. We propose first to define relative importance in a general way.
Let us consider p random variables as predictors and y a response. We can define a Relative Importance Function as a function that associates to each predictor (defined by its index j in the set P={1,…p}) a value of importance:

$$RI : \mathrm{P} \rightarrow \square$$

In practical applications, functions of relative importance are defined using matrix calculus and polynomial functions applied to the correlation matrix, explained variance of models and bivariate or multivariate correlation coefficients. These computations are applied using the estimates of these values. As a consequence we should like Grömping [2] clearly refer to estimators of relative importance. For instance we will see later that some relative importance functions are based on a full decomposition of the explained variance. As the estimator of explained variance in Linear Regression is biased, it is impossible that all estimators of relative importance for each predictor are unbiased as their sum is actually biased. This is why in all that follows we will only discuss the properties of estimators .The topic of relative importance has been discussed in many publications since at least 1936 and a history of the use of relative importance has been presented by Johnson and Lebreton [3]. This article will focus on relative importance evaluation based on allocation of shares of variance for linear regression. Grömping [2] gives an overview of Variance Decomposition methods. Some approaches allocate shares of Variance that can be negative and this has attracted criticism. Others propose the usage of values of relative importance that are all positive but do not add up to the total of explained variance. Lastly some methods fully decompose the explained variance across predictors. We will designate in this paper the methods that assign relative importance values to each predictor so that the sum of these relative importance adds up to the estimated R^2 as Variance Decomposition as opposed to Variance Allocation when the sum of the relative importance estimates is different from the estimated R^2.

In terms of notation we will use the following in reference to the Linear Regression Model and focus this article on allocation of shares of variance of the response y into proportions due to the p predictors X's (and errors) :

$$y = X\beta + e = \sum \beta_i X_i + e \quad (1)$$

We will work in the case of p predictors that are linearly independent, and in the case of n observations n greater than p and the n x p matrix of predictors score is of full rank p. The variance explained by the p predictors (P is the set of the p predictors) is:

$$V(P) = \sum_{i=1}^{i=p} \beta_i^2 v_i + 2 \sum_{i<j} \beta_i \beta_j \sqrt{v_i v_j} \quad (2)$$

with vi variance of Xi and ρ_{ij} coefficient of corrélation between Xi et Xj.

We can assume a regression model without intercept and w.l.o.g that all X's are centered (i.e. have expectation 0). To simplify the notations in the rest of the article we will assume, unless specified, that y and the Xj's are centered and standardized.

We have identified 8 methods published and included in R packages. A detailed documentation on 6 of these methods is available on Pr. Grömping's website [4] dedicated to resources on the relaimpo R package. Another method proposed by both Genizi [9] and Johnson [1] called here Relative Weights (RW's) and finally Zuber and Strimmer [5] CAR scores (Correlation-Adjusted (marginal) correlation) are also available via R packages (relaimpo and yhat). We will first present 3 methods of allocation and then 5 methods of decomposition and then discuss some points of difference and convergence between these approaches.

.

3 Methods for Variance Allocation

3. 1 Allocation "first".

The measures are the squared correlations of the predictors with the response:

$$first(j) = \mathrm{cov}(y, Xj)$$

When the predictors are mutually de-correlated the sum of the measures "first" adds up to the overall R^2 of the model. When this is not the case, the sum of the first (j) over all p predictors is often higher that the overall R^2 of the model.cf. Grômping [4].

3.2 Allocation "last".

This measure attributes as Relative Importance for a predictor j the increase in R^2 when predictor j is included last in the model compared to the R^2 with only the other p-1 predictors. This measure is identical to the squared semi-partial

correlation $sr^2(j)$, which is sometimes presented as the amount by which the R^2 is reduced when this predictor is deleted from the regression equation. See for instance Tabachnick and Fidell [6].

3.3 Allocation "betasquared".

This relative importance measure consists in attributing as importance the square of the standardized regression coefficient. Like the measures 3.1 and 3.2 these are variance allocations as the sum of these measures for all p predictors do not in general add up to the R^2.

4 Methods of Variance Decomposition

4.1 Decomposition Hoffman-Pratt

This measure of relative importance noted pratt(j) attributes to a predictor j the product of the standardized multiple regression coefficient by the marginal correlation between the predictor j and the response When the predictors are standardized :

$$pratt(j) = \beta_j \rho_{yj}$$

From the properties of the OLS regression we can easily confirm that this measure is leads to a decomposition of the R^2.

$$R^2 = \text{cov}(y, \sum_j \beta_j Xj) = \sum_j \beta_j r_{yj} \quad (3)$$

4.2 Shapley Value or LMG or Average

This method has been assigned several names. See for instance Grömping [2] for an historical overview. We will call this measure here lmg(j) or SV(j). This measure is computed by averaging on all possible ordering of the p predictors the increase of the R^2 when the predictor j is added to the model based on the other predictors entered before j in the model. These values have been proposed by Lindeman, Merenda and Gold (1980), hence the name lmg. If we consider a game theory perspective where we assimilate the p predictors as players and define the game function of a coalition of k players as the R^2 achieved by the model based on these k predictors, it turns out that the application of Shapley Value to the game described above generates exactly the same values as lmg,, hence the possible notation SV(j).

Let us present below one notation and one of the ways to compute lmg. Let r be a permutation of P, this constitutes an ordering of the predictors. Each permutation r enables to define an order of entry of the predictors in the model.

Let Sj(r) be the set of predictors entered before j in the permutation r. We can compute:

$R^2(S_j(r))$ as the R² of the model including the predictors in Sj(r)

$R^2_{+j}(S_j(r))$ as the R² of the model including the predictors in Sj(r) U{j}

And define $\Delta_j(r) = R^2_{+j}(S_j(r)) - R^2(S_j(r))$

$\Delta_j(r)$ is the increase in R² when the predictor j is added to the predictors entered before j in the model with order resulting from the permutation r.

$$lmg(j) = \frac{1}{p!}\sum_r \Delta_j(r) \tag{4}$$

averaged on all 2^p permutations of the p predictors. This formula can be rewritten in different forms, combining the permutations that have the same sets Sj(r).

4.3 PMVD (Proportional Marginal Variance Decomposition).

This measure is also a variance decomposition and is computed similarly as for lmg but with weights attached to each single permutation:

$$pmvd(j) = \frac{1}{p!}p(r)\sum_r \Delta_j(r) \tag{5}$$

For more details about PMVD see Feldman [7] and also Grömping [2].

4.4 Relative Weights

Fabbris [8] has proposed a way to decompose the explained variance using the Singular Value Decomposition of the matrix X. Later Genizi [9] and Johnson [1] used this approach in a different way. This decomposition is a particular case of a more general approach consisting of using a set of mutually uncorrelated variable to decompose the explained variance. We will formalize the orthogonal decomposition in general and then present the Relative Weights computation.

Let $z_i, i = 1,...p$ as set of p orthogonal standardized predictors:

Let us note $\lambda_{ji} = \mathrm{cov}(z_j, X_i)$ and $\beta_i = \mathrm{cov}(y, z_i)$

We compute the Orthogonal Decomposition RW with the zi's as follows

$$RW(j) = \sum_{i=1}^{i=p} \lambda_{ji}^2 \beta_i^2 \tag{6}$$

The Relative Weights generate a full variance decomposition because:

$$\sum_{j=1}^{j=p} RW(j) = \sum_{j=1}^{j=p}\sum_{i=1}^{i=p} \lambda_{ji}^2 \beta_i^2 = \sum_{i=1}^{i=p} \beta_i^2 \sum_{j=1}^{j=p} \lambda_{ji}^2 \quad (7)$$

$\lambda_{lm} = \mathrm{cov}(z_l, x_m)$, and the z_i being a set of standardized orthogonal vectors and as the x_j are also standardized finally:

$$\sum_{i=1}^{i=p} RW(\mathrm{i}) = V(y) = 1 \quad (8)$$

So the Relative Weights computed with any set of z_i enables the computation of a full decomposition of the R^2 using the RW_j. The decomposition proposed by Genizi and Johnson consists in computing the Relative Weights using a specific set of orthogonal predictors that minimize the sum of the squares between each X_j and z_j.

So in terms of variables minimizing: $\Psi = E[(z - X)'(\mathrm{z} - X)]$

In the case of a specific dataset with n observations and p predictors this leads to consider a specific matrix Z of the zj is as follows:

Let X be the n x p matrix of standardized centered observations. Let $X = P\Delta Q'$ the singular value decomposition of X. The set of zi minimizing the abovementioned sum of squares is $Z = PQ'$.The orthogonal decomposition using this specific set of orthogonal vectors are the Relative Weights. We will use the notation RW(j) from now on for this specific decomposition and Vo(j) in case we use another set of zi's.

4.5 CAR scores

The CAR scores are the squared correlations between the response and the vectors Z as defined in 4.4.So: CAR(j)=λj²

This is a recent Variance Decomposition proposed by Zuber and Strimmer [10]. They use the term CAR standing for Correlation-Adjusted (marginal) coRelation.

We have limited the presentations of these methods to the strict minimum detail, but the documents in reference offer additional perspective on the Game Theory approach and axiomatic definitions of desirable properties in variance decomposition.

5 Results on Variance Decomposition

There are important difference between the usage of the Linear Model and the interpretation of Variance Decomposition values. In the case of lmg for instance, it is important not to use in a simplistic way the variance decomposition as if they were equivalent to the coefficients generated by Linear Regression Model. First because they are terms of variance that are actually homogeneous to squared values of the β's. If we consider an ideal case with mutually decorrelated predictors the lmg value would be distorted compared to the relative values generated by the Linear Model.
Also another way to write lmg in the case of two predictors is as follows:.

$$\text{lmg}(1) = \frac{V(y) + (\beta_1^2 * \text{v}_1 - \beta_2^2 * \text{v}_2) * (1 - \rho_{12}^2)}{2} \qquad \text{lmg}(2) = \frac{V(y) + (\beta_2^2 * \text{v}_2 - \beta_1^2 * \text{v}_1) * (1 - \rho_{12}^2)}{2}$$

If we consider a case with two predictors a sufficiently high correlation and a third predictor uncorrelated with the first two the reallocation of importance between the two lmg values can lead ultimately to different rankings between the importance measures if we apply the beta squared versus lmg.

So all in all lmg produces distortion and can potentially change the ranking between the importance of predictors versus the results derived from the Linear Model. This is why we need to be careful not to consider them as a full alternative to standard models. Conklin and Lipovetsky [11] have considered adjusting regression coefficients using Shapley Value as an intermediate step of calculation and computing coefficients in resolving a quadratic equation equalizing the Hoffman values and lmg for each predictor. Grömping and Landau [12] have criticized this approach.

Regarding similarities, Johnson and Lebreton [3] have observed the proximity between the results of relative weights and lmg and state :« *Despite being based on entirely different mathematical models, Johnson's epsilon and Budescu's dominance measures (Note : Budescu's dominance is one of the denomination of Shapley Value /lmg) provide nearly identical results when applied to the same data these two mathematically different approaches suggests that substantial progress has been made toward furnishing meaningful estimates of relative importance among correlated predictors. The convergence between these two mathematically different approaches suggests that substantial progress has been made toward furnishing meaningful estimates of relative importance among correlated predictors".*

We will first analyze and formulate results in the case of two predictors. Starting with two uncorrelated standardized variables E1 and E2, we will construct X1, X2, and y:

$$X_1 = \cos(\varphi)E_1 - \sin(\varphi)E_2 \qquad X_2 = \cos(\varphi)E_1 + \sin(\varphi)E_2$$
$$y = \cos(\psi)E_1 + \sin(\psi)E_2$$

From this we can actually compute:

$$\beta_1 = \frac{\sin(\varphi - \psi)}{\sin(2\varphi)} \qquad r_{y1} = \cos(\varphi + \psi)$$
$$r_{y2} = \cos(\varphi - \psi)$$
$$\beta_2 = \frac{\sin(\varphi + \psi)}{\sin(2\varphi)} \qquad \rho_{12} = \cos(2\varphi)$$

$$\text{last}(1) = \sin^2(\varphi - \psi); \text{first}(1) = \cos^2(\psi + \varphi)$$
$$\text{last}(2) = \sin^2(\psi + \varphi); \text{first}(2) = \cos^2(\psi - \varphi)$$

$$\text{SV}(1) = \frac{(1 - \sin(2\varphi)\sin(2\psi))}{2}$$
$$\text{SV}(2) = \frac{(1 + \sin(2\varphi)\sin(2\psi))}{2}$$

It is also possible to compute the result of orthogonal decomposition using any orthogonal base of the considered plane let us consider z1 and z2 such as:

$$z_1 = \cos(\omega)E_1 + \sin(\omega)E_2$$
$$z_2 = -\sin(\omega)E_1 + \cos(\omega)E_2$$

The results of an orthogonal decomposition process using the zi defined by the choice of a specific value of ω are Vo(1) and Vo(2) as computed below :

$$Vo(1) = \cos^2(\psi - \omega)\cos^2(\omega + \varphi) + \sin^2(\psi - \omega)\cos^2(\omega - \varphi)$$
$$Vo(2) = \cos^2(\psi - \omega)\sin^2(\omega + \varphi) + \sin^2(\psi - \omega)\sin^2(\omega - \varphi)$$

It is easy to demonstrate (w.l.o.g with $\varphi \leq \pi/2$),that the specific z_i considered earlier to implement the Relative Weights of the variance decomposition proposed by Johnson and Genizi, corresponds to the case when $\omega = -\pi/4$. Taking $\omega = -\pi/4$ in the computations of Vo(1) and Vo(2) and simplifying we get:

$$\text{RW}(1) = \frac{(1 - \sin(2\varphi)\sin(2\psi))}{2}; \text{RW}(2) = \frac{(1 + \sin(2\varphi)\sin(2\psi))}{2}$$

We recognize here the formula for $SV(1)$ and $SV(2)$. So we have demonstrated through a trigonometric approach that in the case of two predictors the relative weights and the lmg (or Shapley Values) are identical (cf. also Thomas and al. [14]). This result is just the application of simple trigonometric equivalences and should not in our view lead to conclude that

because the two methods converge this is in itself a justification of their validity. The demonstration proposed here enables easy visualization of the impact of the choice of orthogonalisation if we let ω vary. In the case of two predictors it also enables to demonstrate that the CAR scores may remain constant even when the correlation between predictors vary. We can also notice some other links between orthogonalisation procedures and lmg if we use some particular sets of orthogonal vectors in the space generated by the Xj's. As Relative Weights is a particular case of decomposition by orthogonalisation , these links help understand the proximity between lmg which is an averaging of last values over submodels and relative weights , which is a decomposition by orthogonalisation.

Case 1: Let us consider y* as the projection of y on the space of the predictors and let us choose one given predictor and an orthogonal set of zi's with the condition that:

$$zj = \frac{y^*}{\square y^* \square}$$

We have

$$\forall i \neq j, y.zi = 0 \quad and \quad y^*.zj = 1$$

Let us now use the RW calculation formula:

$$\mathrm{Vo(j)} = \sum_{i=1}^{i=p} \lambda^2 \mathrm{ji}^* \beta^2 \mathrm{i} \text{ as:}$$

$$\lambda ji = \mathrm{cov(zj}, Xi) \quad \beta i = \mathrm{cov(y, zi)} \quad \beta j \neq 0; \beta j = 1$$

we have :

$$\mathrm{Vo(j)} = \mathrm{cov}^2(\mathrm{y, xj}) = \mathrm{first(j)}$$

This means that for any j there is always at least one choice of orthogonal decomposition that will allocate first (j) to that predictor.

Case 2: This time we will consider an orthogonal set so that:

$$zj = \frac{uj}{\square uj \square}$$

u_j being the residual of the regression of X_j on the other variables.

$$if \quad i \neq j \quad \lambda ji = \text{cov}(\text{zj}, Xi) = 0, \quad and \quad \lambda^2 jj = \text{cov}^2(\text{z}\, j, Xj) = 1 - \text{R}_j^2$$

$$As \quad \beta^2 j = \text{cov}^2(\text{y}, \text{zj}), \quad \text{Vo(j)} = last(j)$$

These examples show that orthogonalisation methods do enable with specific sets of orthogonal vectors to allocate either first(j) or last(j) for one given predictor. As Johnson is a particular case of orthogonalisation and lmg (j) is an average of last (j) across submodels, it confirms why there can be a proximity between variance decomposition via orthogonalisation and lmg in the case of more than 2 predictors. Both lmg and Relative Weights are computer intensive methods. We introduce an alternative variance decomposition method that is much more computationally efficient and offers very similar results to lmg and relative weights. The method will allocate to each predictor j a share of variance that is a weighted average between first(j) and last (j) , hence the name weifila for weighted first last.

Here are the computation steps: let L and F be the sum of first and last for all predictors:

$$L = \sum_j last(j) \qquad F = \sum_j \text{first}(j)$$

We will consider two cases in the usual situation where $L \neq F$

$$If \quad L < R^2 < F \quad \text{W}(j) = last(j)\left(\frac{F - R^2}{F - L}\right) + first(j)\left(\frac{R^2 - L}{F - L}\right) \tag{9}$$

$$If \quad F < R^2 < L \quad \text{W}(j) = last(j)\left(\frac{R^2 - F}{L - F}\right) + first(j)\left(\frac{L - R^2}{L - F}\right) \tag{10}$$

The case where R^2 would be outside of the interval between F and L is not encountered in practice. By construction in both cases above: $\sum w(j) = R^2$

We will note also that in the case of two predictors the weifila values equate to the lmg and relative weights values as shown below:

$$F = first(1) + first(2) = 1 + \cos(2\psi)\cos(2\varphi)$$

$$L = last(1) + last(2) = 1 - \cos(2\psi)\cos(2\varphi)$$

$$F - L = 2\cos(2\varphi)\cos(2\psi) \qquad w(j) = \frac{first(\text{j}) + \text{last(j)}}{2}$$

So we recognize the formula above for lmg and this confirms that:

$$w(j) = lmg(j) = SV(j) = RW(j) \tag{11}$$

Weifila is a way to select an intermediate point w(j) inside the interval between first(j) and last(j) for each j. We have compared these 3 measures on two different datasets. The results are presented below:

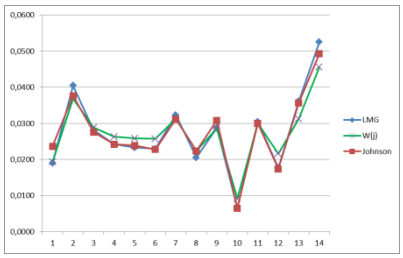


Fig 1: 1499 obsevations.14 predictors

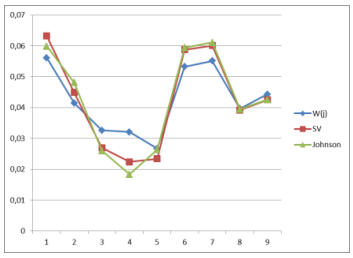


Fig 2: 499 observations 9 predictors

The weighted first last average "weifila" is much simpler to compute and delivers very similar results at least in the typical size of datasets and number of drivers used in practical applications for marketing and social research.

6 Conclusions

Among several methods to allocate variance among predictors, the proximity between lmg and Relative Weights has been noted (Johnson and Lebreton [3]), and seen as a justification of their validity. This proximity is actually a complete equality in the case of a model with two predictors and results from simple geometric properties. Also there exists variance decomposition via orthogonalisation that allocate exactly last(j) or first(j) for any of the predictors. So this proximity should not be seen in itself as a justification of validity.

The new method of variance decomposition proposed in this paper via a weighted average between first(j) and last(j) for each predictor provides very consistent results with lmg and Relative Weights but is simpler and less computer intensive. This method has been successfully tested with datasets typical of situations encountered in marketing research applications.

As underlined by Johnson [1] and Grömping [2], variance decomposition should not be seen as a substitute for linear regression models or path analytical models and models based on theory driven explanations can be more relevant than using directly variance decomposition. However when a model based on theory is not available variance decomposition can help identify important variables. Lastly the usage of modern machine learning techniques can also be considered.

References

1. J. W.Johnson. A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression. Multivariate Behavioral Research,35(1) 1-19. 2000.
2. U.Grömping. Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. The American Statistician, Vol 61 , No2 p139 2007.
3. J.W Johnson and J.M. Lebreton. History and Use of Relative Importance Indices in Organizational Research. *Organizational Research Methods* **7**, 238 - 257. . 2004.
4. http://prof.beuth-hochschule.de/groemping/relaimpo/
5. V.Zuber, and K.Strimmer. Variable importance and model selection by decorrelation. Preprint. http://arxiv.org/abs/1007.5516 (2010).
6. B. Tabachnick L. Fidell , Using Multivariate Statistics , Fifth Edition , 5.6.1.1 Pearson 2006.
7. B Feldman . "Relative Importance and Value." Manuscript version 1.1, 2005 http://www.prismanalytics.com/docs/RelativeImportance.pdf
8. L. Fabbris. Measures of regressor importance in multiple regression: an additional suggestion. *Qual Quant* 1980, 4:787–792
9. A. Genizi. Decomposition of R^2 in Multiple Regression with correlated regressors. Statiscica Sinica 4 , 407-420. 1993.
10. V. Zuber, K Strimmer. High-Dimensional Regression and Variable Selection Using CAR Scores. Statistical Applications in Genetics and Molecular Biology **10**: 34. 2011.
11. S**.**Lipovetsky, M. Conklin. Analysis of Regression in Game Theory Approach. *Applied Stochastic Models in Business and Industry* **17**, 319-330.2001
12. U Grömping , S.Landau.Do not adjust coefficients in Shapley Value Regression *Applied Stochastic Models in Business and Industry* **17**, 319-330.2009
13. U Grömping. Variable importance in regression models *WIRE's Comput Stat 2015,7:137-152. Doi:10.1002/wics.1346*
14. R D Thomas , Bruno D Zumbo, Ernest Kwan, Linda Schweitzer. On Johnson's (2000) Relative Weights Method for Assessing Variable Importance A Reanalysis : *Multivariate Behavioral Research July 18 2014*

The Antioxidant Properties of the Stevia Rebaudiana Plant

Rivka Weiser Biton
Department of Chemistry, ORT Braude College, Carmiel, Israel
(E-mail:rivkaw@braude.ac.il)

Abstract.
The worldwide use of Stevia rebaudiane as substitute sweetener for sugar, increases every year. Stevia rebaudiane is not involved in the insulin mechanism and as so has no calories. This makes Stevia rebaudiane also a natural substitute for the more common synthetic substitutes for sugar. Various studies show that the plant Stevia rebaudiane contains substances with properties of antioxidant reagent, and as such can prevent antioxidation damage to DNA.[1]
In this work protocols for extraction are developed and antioxidating activity tested.
The aim was to find optimal conditions for the process, so the antioxidants in the Stevia plant were preserved. Methods used for testing: Redox-titrations and spectrophothometric methods.
In all results Vitamin C was used as reference for the antioxidation activity. The results show that the Stevia rebaudiane plant is active as an antioxidant reagent; and that the extent of the antioxidant activity depends on the solvent and on the conditions of the extraction process.

Keywords: Stevia rebaudiane, natural substitue for sugar, extraction process, antioxidating activity, methods for testing

1 Introduction

Artificial sweeteners are the most widely used food additives[2]. The consumption is relative safe and can reduce the use of sugar-sweetened beverages, sweets, cakes without the impact of weight gain. Many people have turned to high-intensity sweeteners – but there is still a question if those food additives are safe?
Saccharin had shown no side effects in 120 years of use. Cyclamate also had not shown any side efffects in over 70 years of use around the world- but is still banned in the USA. Sucralose seems safe for most people, except people who are allergic to it.
Aspartame is not safe because it is metabolized by the body and broken down to harmful products, one of them is Methanol.
An additional problem is, that the study with mice showed us, that the mice were not satisfied with the artificial sweetener, instead they were pulled towards sweeter food rich in calories.
There is some evidence that frequent consumption of high intensity sweeteners may have a counterintuitive effect and increase risk of excessive weight gain and metabolic syndrome.
This leads to the assumption that the sweeteners cannot fill in for the "crave for sweets"[3].
The Stevia plant is a natural sweetener which is composed of 12 different kinds of sweeteners from the glycoside family[4] and has no influence on the blood sugar level. The plant is well known for its high sweet diterpene content (4-20%) in dry mass, which includes Stevioside, Rebaudiosides A-F, and Dulcoside A. 5-10% of the dry Stevia plant is Stevioside[5] which is 200-300 times sweeter than Sucrose with a bitter aftertaste. 2-4% of the dry Stevia plant is Rebaudioside A which is 300-400 times sweeter than

Sucrose with a mild aftertaste. 1-2% of the dry Stevia plant is Rebaudioside C, approximiately 30 times sweeter then Sucrose with a bitter aftertaste. 0.5-1% of the dry Stevia plant is Dulcoside A, also approximiately 30 times sweeter then Sucrose but with a very bitter aftertaste. Other kinds of the glycocides in the plant are not used as sweeteners.

The Stevia leaves are known as sugar leaves in South and Central America[6]. The leaves are used in South America to sweeten the "Mata" drinks and in Japan as substitute for sugar since 1971, Canada approved the Stevia plant as food additive. In addition there are commercial products which use only specific kind from the glycoside family, such as mainly Rebaudiosides A for Reb-A; a non calorie sweetener.

Polyalcohols with Rebaudiosides are used for Truvia, sweetener developed by Cargill and Coca-Cola. The name of the sweetener for Pepsi-Cola is Pure Via.

Can the Stevia plant offer more then being in use as a non – calorie sweetener? It turns out that the Stevia leaves contain substances which are antioxidant[7,8]. There are no reports so far about the antioxidant activity as function of the consumption of Stevia.

Antioxidant activity was studied in common plants[9]. Lertsiri[10] published a paper about different plants used as food and tested the antioxidative properties of those plants. He showed that the infusion from the Stevia plant had similar antioxidant properties to an infusion of green tea.

In addition Stevia rebaudiana leaves exhibited preventive activity against DNA strand scission by •OH, generated in Fenton's reaction.

In this work we will show the antioxidant activity of the Stevia plant compared to Vitamin C (Ascorbic acid). Vitamin C is an electron donor, as such a potent water-soluble antioxidant in humans[11]. Antioxidant effects of Vitamin C have been demonstrated in many experiments in vitro. Oxidation of lipids, proteins and DNA results in specific oxidation products that can be measured in the laboratory.

Why this research?

The purpose of the research is to characterize and to indentify the antioxidant properties of the Stevia plant as a function of the extraction process of the plant[12]. The main advantage of this research is the combination of sweetener and antioxidant activity found in the Stevia Rebaudiana plant.

2 Materials

The antioxidant properties of Stevia were tested by a number of methods and compared to ascorbic acid. All of those methods measure amount of ascorbic acid.

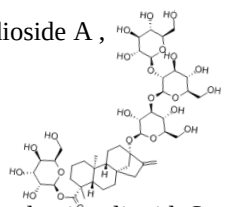
Chemicals:

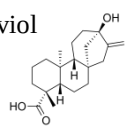
C_2H_5OH(aq) 70% , KI(aq) 0.01M , KIO_3 , HCl(aq) 5M , Starch 1% , Ascorbic acid ,

H_2SO_4(aq) 0.125M , Hydroxylamine hydrochloride ($NH_2OH{\cdot}HCl$) ,

H_2SO_3(aq) , N-(1-Naphthyl)ethylenediamine dihydrochloride (NEDA with 2HCl) , NaF ,

$K_3Fe(CN)_6$(aq) 1% , $FeCl_3$(aq) 1% , H_3PO_4(aq) 0.2%,

Rebaudioside A , 

Steviol 

Sucrazit classic – liquid, Sucrazit Stevia – powder

Plant materials:

Orange juice and peel, Lemon juice and peel,

Fresh leaves of Stevia, dried leaves of Stevia, dry leaves of Stevia from Nature Store

3 Methods

Determination of water percentage in fresh Stevia leaves:

Dry leaves are compared to fresh leaves- the main difference is the water content of the plant. Two drying methods were used: I.Weighting the plants before and after heating at 40°C for 2 weeks. II.Weighting the plants before and after drying at room temperature for 4 weeks.

Comparison between the drying methods showed consistency. Both drying methods resulted in the same procentage of water in the plants.

Extraction methods:

Cold extraction, hot extraction with various temperatures, organic extraction [13,14,15].

Active carbon was added to all extraction of the stevia plant in order to accomplish clear and colourless solutions. Every cycle of extraction started with approxiemately 4 gram of Stevia plant and was repeated many times in order to allow at least four repetitions of each measuring method.

Procedure for cold extraction:

a. weight the extract; b. add water as solvent; c. stir the solution for 6 hours at 25°C;

d. filter; e. transfer to volumetric flask.

Procedure 1 for hot extraction:

a. weight the extract; b. add water as solvent; c. stir the solution for 1 hours at 75°C;

d. filter; e. transfer to volumetric flask.

Procedure 2 for hot extraction:

a. weight the extract; b. add water as solvent; c. heat for 5 minutes at 95°C;

d. stir the solution for 55 minutes at 90°C; e. filter f. transfer to volumetric flask.

Procedure 3 for hot extraction:

a. weight the extract; b. add water as solvent; c. stir the solution for 5 minutes at 95°C;

d. filter; e. transfer to volumetric flask.

Procedure for organic extraction:

a. weight the extract; b. add Ethanol 70%-water 30% as solvent; c. stir the solution for 30 minutes at 70°C;

d. filter; e. transfer to volumetric flask; f. add Ethanol 70%-water 30% till fill line.

Determination of Ascorbic acid by titration[16] :

1. Iodate (IO_3^-) oxidizes iodide (I^-) to iodine (I_2) : $IO_3^- + 5I^- + 6H_3O^+ \rightarrow 3I_2 + 9H_2O$

2. Iodine oxidizes Ascorbic acid ($C_6H_8O_6$) : $C_6H_8O_6 + I_2 + 2H_2O \rightarrow C_6H_6O_6 + 2H_3O^+ + 2I^-$

3. Starch-iodine endpoint : I_2 forms blue complex with starch indicator

In this work the titration was done with three different concentrations of KIO_3 in order to obtain a valid titration volume.

Every titration was repeated five times to achieve accurate results.

Determination of Ascorbic acid by Spectrophotometer:

Two methods: I. Determination by iron and Phenanthroline[17] measured at 420nm and II. Determination with Iodate at 540nm gave no consistent results.

The formation of Prussian Blue is a classical qualitative test to detect Fe(II) using hexacyanoferrate(III)[18] as reagent. The procedure was successfully applied for the determination of Ascorbic acid by Mantei, Dorinas and Radu[19].

1. Oxidation of iron(II) : $Fe^{2+} + [Fe(CN)_6]^{3-} \rightarrow Fe^{3+} + [Fe(CN)_6]^{4-}$

2. Formation of hexacyanoferrate (II) ferric insoluble complex (Prussian Blue):

$$4Fe^{3+} + 3[Fe(CN)_6]^{4-} \rightarrow Fe_4[Fe(CN)_6]_3$$

With the addition of KCl 0.1M and HCl 0.01M this reaction is quantative in low concentrations.

3.A deep blue soluble compound is formed when Fe(III) is reduced to Fe(II) by Ascorbic acid:

$$2Fe^{3+} + C_6H_8O_6 \rightarrow 2Fe^{2+} + C_6H_6O_6 + 2H^+$$

The reduction of Prussian Blue reagent by Vitamin C is a pH dependent process. A steady-state signal for the spectrophotometer is reached after fiveteen minutes.

4. Prepararion of calibration curve with concentration of Ascorbic acid in the range of 0 – 20.000mg/liter.

5. Prussian Blue has an absorbance peak at 700nm.

Samples of Stevia leaves after organic extraction were not repeatable when measured by this spectroscopic method.

4 Results

List of abbreviations:

STTC – fresh leaves of Stevia – cold extraction
STTH – fresh leaves of Stevia – hot extraction
STTO – fresh leaves of Stevia – organic extraction
STYC – fresh and dried leaves of Stevia – cold extraction
STYH – fresh and dried leaves of Stevia – hot extraction
STYO – fresh and dried leaves of Stevia – organic extraction
STKC – dry leaves of Stevia from nature store – cold extraction
STKH – dry leaves of Stevia from nature store – hot extraction
STKO – dry leaves of Stevia from nature store – organic extraction

A - Rebaudioside A
B - Steviol
VC - Vitamin C
SU - Sucrazit classic – liquid
SUST - Sucrazit Stevia - powder

Table 1: Water percentage in fresh Stevia leaves:

mass of fresh leaves [gram]	mass of drying process leaves after drying process [gram]	water content %	dry leaves %
3.9961	0.4967	87.57	12.43
3.9993	0.4976	87.56	12.44
4.0010	0.4981	87.55	12.45

standard deviation:0.01

All results are at least average of four repititions. This measurement is important for comparisson between dry leaves and fresh leaves, dried at room temperature.

Amount of Ascorbic acid relative to 1mg Vitamin C determent by titration:

Table 2. All titration results

*** the titrant was diluted 100 fold to obtain valid titration volume

the titrant was diluted 20 fold to obtain valid titration volume

	sample	extraction temperature ° C	extraction time min	weight mg	weight after drying mg	volumetric flask ml	concentration of sample mg/ml	volume of sample ml	weight of sample for titration mg	volume of titrant *** / ## ml	volume needed for 1mg of sample ml	antioxidant activity relative to 1mg vitamin C
***1	STKH	75	60	4012.5		100	40.125	20	802.5	0.160	1.99E-04	2.36E-04
***2	STKH	90	60	4012.5		100	40.125	20	802.5	0.160	1.99E-04	2.36E-04
***3	STKH	95	5	4012.5		100	40.125	20	802.5	0.240	2.99E-04	3.54E-04
***4	STKC	25	420	1999		100	19.990	20	399.8	0.145	3.63E-04	4.29E-04
##5	STKO	70	30	3990		50	79.800	10	798.0	0.980	1.23E-03	1.45E-03
***6	STYH	90	60	4012		100	40.120	20	802.4	0.220	2.74E-04	3.24E-04
***7	STYH	95	5	4012.5		100	40.125	20	802.5	0.270	3.36E-04	3.98E-04
***8	STYC	25	420	4012.5		100	40.125	20	802.5	0.330	4.11E-04	4.86E-04
##9	STYO	70	30	400		50	8.000	10	80.0	0.430	5.38E-03	6.36E-03
***10	STTH	90	60	4012	499.1	100	4.991	20	99.8	0.032	3.21E-04	3.79E-04
***11	STTH	95	5	4034	501.8	100	5.018	20	100.4	0.039	3.89E-04	4.60E-04
***12	STTC	25	420	4034	501.8	100	5.018	20	100.4	0.052	5.18E-04	6.13E-04
##13	STTO	70	30	3720	462.8	50	9.255	10	92.6	0.700	7.56E-03	8.95E-03
***14	SU	25	420	8000		50	160.000	20	3200.0	0.135	4.22E-05	4.99E-05
***15	SUST	25	420	8000		100	80.000	20	1600.0	0.170	1.06E-04	1.26E-04
***16	A	25	420	100		100	1.000	20	20.0	0.170	8.50E-03	1.01E-02
***17	B	25	420	10		50	0.200	20	4.0	0.037	9.25E-03	1.09E-02
18	**VC**			**104.2**		**100**	**1.042**	**20**	**20.8**	**17.620**	**8.45E-01**	**1.00E+00**

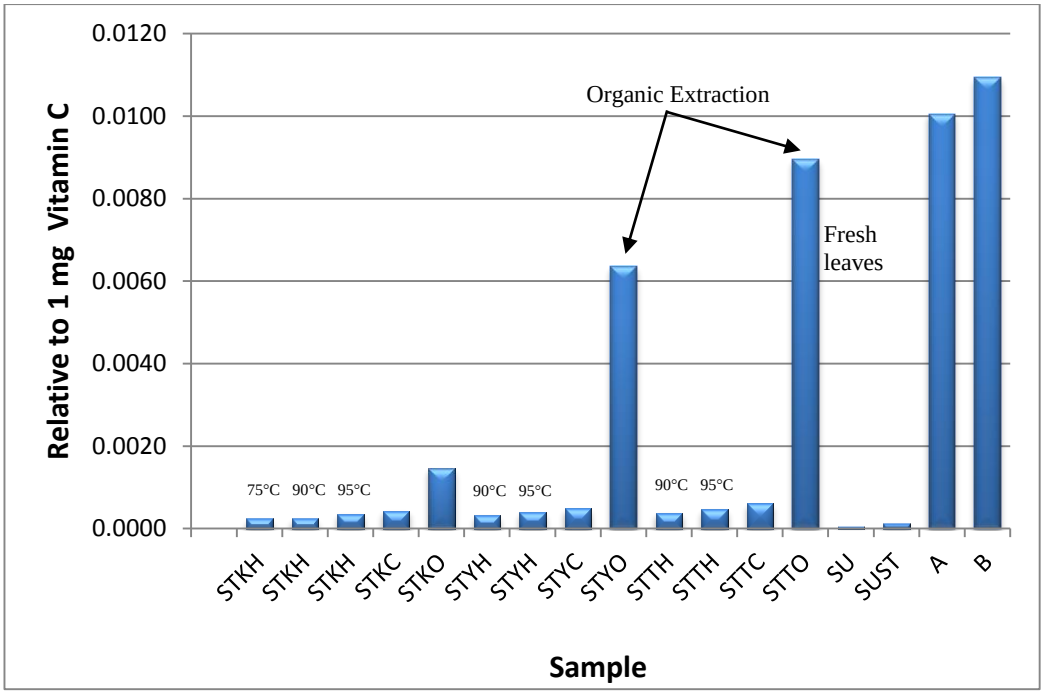


Fig.1. Antioxidant Activity of all Samples as a function of the Activity in 1 mg Vitamin C determent by Titration

The most antioxidant activity was found in Steviol, Rebaudioside A, and in fresh leaves after organic extraction (Fig.1.). Leaves which were dried in the laboratory also contained more antioxidants than dry leaves which were bought in a store.

For the fresh leaves the most efficient extraction by far is the organic extraction (Fig.2.).

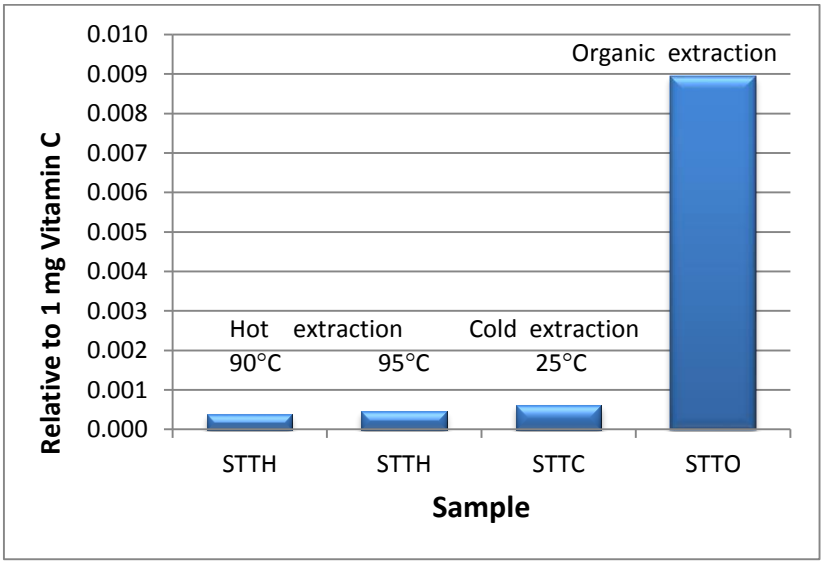


Fig.2. Antioxidant activity in freah Stevia leaves as function of extraction method

For Calibration curve: various concentration of Ascorbic acid $1 \cdot 10^{-4}$M (17.6mg/liter) solution.

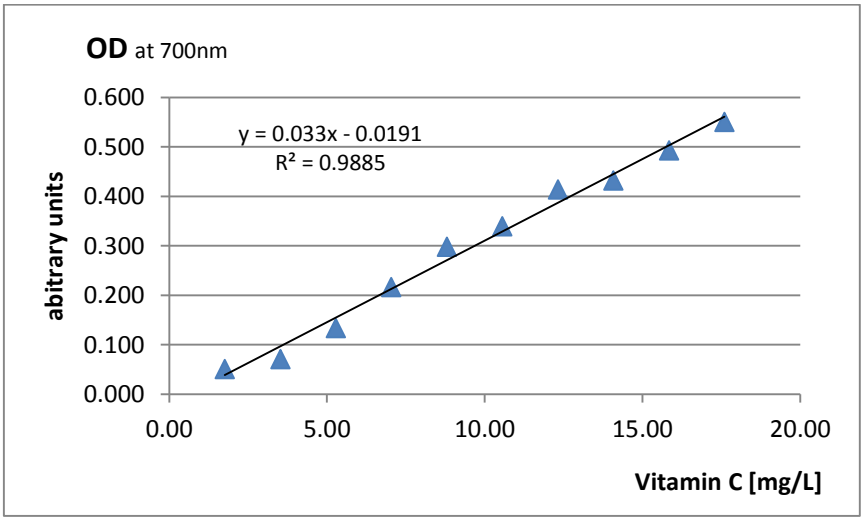


Fig.3. Calibration curve of Vitamin C with Prussian Blue Complex at 700nm

Table 3. Results of spectrophotometric method using the Prussian Blue complex and measured at 700nm

sample	extraction temperature ° C	extraction time min	weight mg	weight after drying mg	volumetric flask ml	concentration of sample mg/ml	volume of sample ml	volume with Prussian blue	weight of sample for spectro mg	OD	concentration of ascorbic acid by calibration curve mg/L	antioxidant activity relative to 1mg vitamin C
STKH	90	60	3920		100	39.200	5	10	19.6	0.217	7.155	3.650E-04
STKH	95	5	3900		100	39.000	5	10	19.5	0.298	9.609	4.928E-04
STKC	25	420	3920		100	39.200	5	10	19.6	0.309	9.942	5.073E-04
STYH	95	5	4000		100	40.000	5	10	20.0	0.372	11.852	5.926E-04
STYC	25	420	4120		100	41.200	5	10	20.6	0.416	13.185	6.400E-04
STTH	90	60	4140	515.0	10	51.502	5	10	25.8	0.478	15.064	5.850E-04
STTH	95	5	3720	462.8	10	46.277	5	10	23.1	0.426	13.488	5.829E-04
STTC	25	420	4360	542.4	10	54.238	5	10	27.1	0.593	18.548	6.840E-04

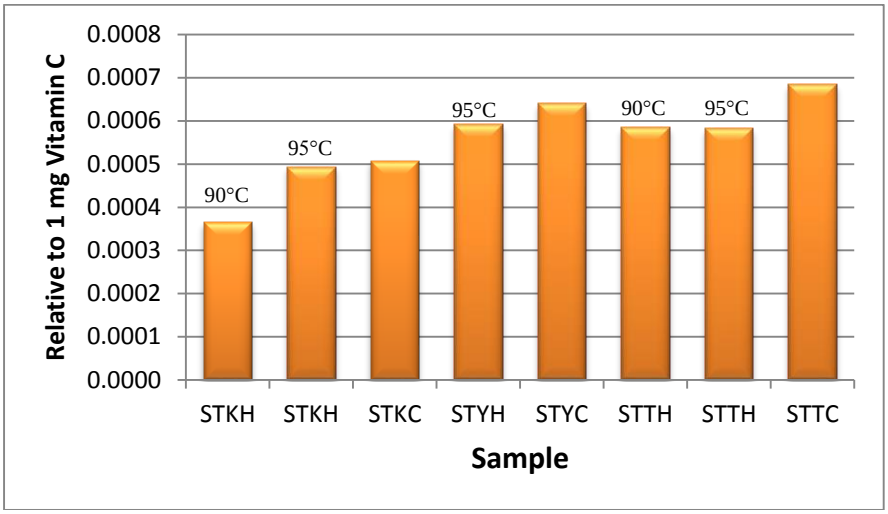


Fig.4. Antioxidant Activity of all Samples as a function of the Activity in 1 mg Vitamin C determent by Prussian Blue Complex at 700nm

The most antioxidant activity was found in fresh leaves after cold extraction (Fig.4.). Leaves which were dried in the laboratory also contained more antioxidants than dry leaves which were bought in a store. The concentration of leaves after organic extraction could not be measured by this spectrophotometric procedure.

5 Discussions and Conclusions

The basic idea of this work was to use the Stevia plant as sweetener and at the same time as antioxidizing reagent. This was the reason, first to study the antioxidizing potential of the Stevia plant compared to a known antioxidizing reagent: Ascorbic acid (Vitamin C). Second to study how to use the Stevia plant while guarding its properties, sweetness and antioxidizing potential. Significant for the way of using the plant are: kind of leaves (fresh or dried) and kind of extraction.

At the lab all kind of Stevia leaves were tested – fresh and dried in a few ways; so were different kind of extraction tested as cold, hot extraction at different temperatures and at a different time of extraction. In addition organic extraction was tested. The purpose of all of the tests were to understand how those methods effect the antioxidizing activity of the plant.

The results of this research show that the cold extraction of all kind of Stevia leaves is preverable over the hot extraction as function of antioxidant activity; so is the five minute hot extraction preverable over a thirty minute hot extraction. In other words the extraction of Stevia leaves suits cold beverages, which can contain also sweetness and antioxidizing activity. In hot beverages and baking, because of the temperature, most of the antioxidizing activity is lost[9,10].

The results also show that the most effective extraction is the organic one. In addition the fresh leaves show more antioxidizing activity then the dry leaves (bought or dried in the lab) which was consistent in all kind of extractions (cold, hot and organic). Comparing the dry leaves of the Stevia plant, the one dried at the lab had more antioxidizing activity then the one bought.

Steviol and Rebaudioside A, which origin in the Stevia leaves were also tested. Steviol was found more antioxidizing active than Rebaudioside A. So the Sucrazit Stevia (SUST) was found more antioxidizing active then the Sucrazit (SU), both are widely used as food additives. Again the Stevia plant contributes to the antioxidizing activity.

In this work we compare Steviol and Rebaudioside A, which were pure compounds bought from Fluka Analytical, to the dry and fresh leaves of the Stevia plant. That is to say that 1 mg of pure material was compared to 1 mg crude compound, so it was expected that the pure material will show more antioxidizing activity.

References

1. F.M. Kengne, A.L. Woguia, P. Fogue Souopgui, B. Atogho-Tiedeu, G. Simo, B. Thaddée. Journal of Coastal Life Medicine, 2(12), 962-969, 2014.
2. Merck Index, 11th Edition, 8282.
3. S.E. Swithers. Artificial Sweeteners produce the counterintuitive Effect of inducing metabolic Derangements, Trends Endocrinology and Metabolism, 24(9),431-441, Sep. 2013.
4. Varki A et al. Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, 1999.
5. S. Kailasam. Quantifi Cation of Stevioside and Rebaudioside A in Stevia Rebaudiana Bertoni Leaves using Agilent 1260 Infinity LC, Agilent Technologies, 1-4 (2011).
6. U. Woelwer-Rieck, C. Lankes, A.Wawrzun, M. Wüst. Improved HPLC Method for the Evaluation of the major Steviol Glycosides in Leaves of Stevia rebaudiana, Springer Verlag, Eur Food Res Technol, 231, 581-588, 2010.
7. J.S. Valentine, C.S. Foote, A. Greenberg and J.F. Liebman Eds., Active Oxygen in Biochemistry, Search Series Vol. 3, Blackie Academic & Professional, London, 1995.
8. M. Valko, D. Leibfritz, J. Moncol, M. Cronin, M. Mazur, J. Telser. Free Radicals and Antioxidants in normal physiological Functions and Human Disease, The International J. of Biochemistry&Cell Biology, 39 (1), 44-84, 2007.s.
9. N.A. Khalaf, A.K. Shakya, A. Al-Othman, Z. El-Agbar, H. Farah. Antioxidant Activity of Some Common Plants, Turk J. Biology, 32, 51-55, 2008.
10. P. Deetae, P. Parichanon, P. Trakunleewatthana, C. Chanseetis, S. Lertsiri. Antioxidant and anti-Glycation Properties of Thai herbal Teas in Comparison with conventional Teas, Food Chemistry, 133, 953-959, 2012.
11. S.J. Padayatty, A. Katz, Y. Wang, P. Eck, O. Kwon, J.H. Lee, S. Chen, C. Corpe, A. Dutta, S.K. Dutta,M. Levine. J. of the American College of Nutrition, 22(1), 18-25, Feb. 2003.
12. I.A. Jahan, M. Mostafa, H. Hossain, I. Nimmi, A. Sattar, A. Alim and Syed M.I. Moeiz. Bangladesh Pharmaceutical Journal, 13(2), 67-75, July 2010.
13. S. Singh, V. Garg, D. Yadav, M. Nadeem Beg, N. Sharma. In vitro antioxidative and antibacterial Activities of various Parts of Stevia rebaudiana (Bertoni), International J. of Pharmacy and Pharmaceutical Sciences, 4(3), 468-473, 2012.
14. Sh. Shukla, A. Mehta, V.K. Bajpai, Sa. Shukla. In vitro Antioxidant Activity and total phenolic Content of ethanolic leaf Extract of Stevia rebaudiana Bert, Experimental and Toxicologic Pathology, 47, 2338-2343, 2009.
15. S. Shukla, A. Mehta, P. Mehta, V.K. Bajpai. Antioxidant Ability and total phenolic Content of aqueous leaf Extract of Stevia rebaudiana Bert, Experimental and Toxicologic Pathology, 64, 807-811, 2012.
16. Determination of Vitamin C Concentration by Titration, University of Canterbury, Christchurch, New Zealand.
17. J. Anwar, M.I. Farooqi, S.A. Nagra and A.M. Khan. A new Method for Spectrophotometric Determination of Ascorbic Acid, Jour. Chem. Soc. Pak., 12(1), 76-79, 1990.
18. L.N. Jattinagoudar, K.S. Byadagi, N.M. Shirhatti, S.T. Nandibewoor and S.A. Chimatadar. J. of Chemical and Pharmaceutical Research, 5(4), 290-300, 2013.
19. N. Matei, S. Dobrinas, G.L. Radu. Spectrophotometric Determination of Ascorbic Acid in Grapes with the Prussian Blue Reaction, Ovidius University Annals of Chemistry, 23(1), 174-179, 2012.

Active Learning vs. Traditional Learning in the View of Lecturers and Students

Rivka Weiser-Biton[1], Miri Shacham[2], David Pundak[3]

[1]Prof. Ephraim Katzir Department of Biotechnology Engineering, Teaching and Learning Center, ORT Braude College, P.O. Box 78, Karmiel 21982, Israel, Tel: 972-4-9901901, Fax: 972-4-9901839, E-mail: : rivkaw@braude.ac.il

[2] Teaching & General Studies Depts., Teaching and Learning Center, ORT Braude College, P.O. Box 78, Karmiel 21982, Israel, Tel: 972-4-9901724, Fax: 972-4-9901738, E-mail: mshacam@braude.ac.il

[3]Kinneret College; Web-Learning Unit, ORT Braude College, P.O. Box 78, Karmiel 21982, Israel, Tel: 972-4-9901843, Fax: 972-4-9901886, E-mail: dpundak@gmail.com

Abstract.
When using online teaching to train students for the modern world it becomes very important to use active learning including independent learning skills, critical thinking, and team/collaborative work. Lecturers in ORT Braude academic college are involved in the development of new teaching technologies and a transition from traditional to active teaching. The lecturers have developed new learning systems involving the activation of students during lectures, peer learning, illustration from activity, learning through self-guidance, work with simulations, problem-solving and a dynamic online course site (Maharshak and Pundak[1]).

research goals: (a) identify attitudes of "active learners" and lecturers towards active learning compared to those of traditional learners and lecturers. (b) examination of the characteristics of "active lecturers". (c) identification of students' attitudes towards active learning. (d) comparison of achievements of students studying according to traditional teaching methods and those studying according to active learning (Pundak, Rozner[2]).

Research tools were: (a) attitudes questionnaire and (b) in-depth interviews. The research's population included 4 lecturers and 152 engineering students in a Chemistry course. The research findings indicate significant differences between "active lecturers" and traditional ones. From the interviews with lecturers it appears that the attempt to encourage them to investigate new attitudes concerning active learning necessitates undermining perceptions concerning the lecturer's role. This situation generates resistance that must be treated. Additionally, it appears that a significant number of the college lecturers lack knowledge concerning group activation, peer learning or the management of illustrations.
Students studying according to the active learning approach reported more involvement in learning and better understanding of the studied subject matter. The students especially noted that they enjoyed meaningful dialogue and brainstorming with group members and that they were able to cope with complex questions in the group. Students' criticism of active learning related to the large investment demanded from them. Comparison of students' achievements in final exams indicated that a significantly higher grade was

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

achieved by students studying according to active learning than by those studying in the traditional manner.
Keywords: Active learning, Investigative learning, Traditional learning

1. Introduction

The need for active learning in science subjects has been discussed in numerous studies (Hake[3], Michael[4], Hall et al.[5]). Researchers have shown that passive learning does not lead to changes in the system of terms with which students arrive at the natural sciences course (Pappalardo and Gunn[6]).
Following past research, this research focused on students' attitudes towards active teaching in general chemistry courses for engineering students in four major aspects: teamwork, application of chemical knowledge in daily life, understanding versus memorization, and enhancing self-efficacy in coping with chemistry questions.
In chemistry studies in Ort Braude College, barriers were identified in understanding basic concepts and linking them to everyday life. Students particularly find it difficult to understand such topics as limiting factor, shifting between different units of concentration, the relationship between the structure and characteristics, and making a synthesis. They find it difficult to apply knowledge learned in class when analyzing chemical events. Consequently, it was decided to adopt the active learning approach in chemistry. For the last ten years, an active learning class has been implemented in the college and basic science courses are taught in this class .In this learning environment, students sit at round tables that allow them to study and work in groups according to the Scale-Up approach (Beichner et al.[7]). Lecturers who teach by the active learning approach have complex roles beyond teaching the material. The lecturer acts as a group facilitator, leads a discussion, helps to solve problems, summarizes and makes comparisons, demonstrates phenomena and more. Therefore lecturer's position is not fixed.. For example, during a lecture, the lecturer stands close to the lecturer's place, where teaching aids are and a central computer are placed; during a discussion the lecturer is near the boards surrounding the class; while working in groups the lecturer is close to students and guides them in coping with the academic tasks.
The development of an active learning chemistry course had several goals: developing tools for teaching in an online center, improving collaborative learning, preparing lessons that will lead to perceptional changes among students about the nature of chemistry. Developing challenging and interesting lessons where fundamental problems in the various topics are discussed by exposure students to experiments / films / problems which will entail thinking and class discussion, encouraging teamwork, creating a social collaborative learning atmosphere between students and lecturer. Engaging in dialogue and brainstorming within the group while coping with complex questions and issues
Key steps in the development of the course included: developing learning materials, identifying areas where students have perceptual problems, preparation of lessons that will exposure students to problem through the

various aids in the active learning center, preparing a short presentation to summarize each topic, constructing a website, integrating movies, animations, experiments, and demonstrations that raise a problem and stimulate thinking .
The lessons included brief lectures and diverse students' activities in the groups. Students' responses were obtained via checking network tasks, a system for collecting responses and submitting exercises. In many lessons the teacher will act as a group work facilitator, and there are topics that the students will learn in groups and then teach the class, taking conceptual tests.

2. Theoretical Background

Active learning can be generally defined as any instructional method which involves students in the learning process; it requires students to engage in meaningful learning activities and reflect upon what they are doing, as opposed to a traditional lecture where students passively receive information from the lecturer (Prince[8]). Previous research on benefits of active learning demonstrates that students learn and understand the content better and also develop improved communication, leadership, ethical decision making and critical thinking abilities (Hake[3], Michael[4]).

Introducing active learning to science and engineering classes
Paulson[9] introduced active learning in an organic chemistry lecture and laboratory classes. The methods he used in order to involve the students, included: group work in and out of class, reading assignments followed by easy single question quizzes, a finger signalling method which allows students to answer questions in class while keeping their answers private, reflection cards. To examine the effects of active learning, he compared 3 classes taught in traditional lecture format and 3 parallel classes taught in cooperative and active learning . His finding revealed significantly higher pass rates for all three classes. Paulson concluded that the change from traditional method to active learning should be done in a gradual manner- it takes time and a fair amount of practice to effectively employ active learning techniques He points out that it is important to explain to the students why lecturers are trying new techniques.
Hall et al.[5] focused on the integration of active learning in four courses in engineering, collectively titled *Unified Engineering* in MIT. The faculty instructors used various active learning techniques, such as concept tests, turn-to-your partner discussions and muddiest-point-in the lecture cards method. The instructors' response to the implementation of active learning was mostly positive, although some reported that certain techniques were hard to implement, they eventually caught on and accomplished their goals of actively engaging the students. The students' response was positive and active learning was viewed as an effective technique to improve their understanding of content. Students commented on positive social dynamics with their peers .
Larson and Ahonen[10] described a case study of using active learning in a Finnish engineering course. The common Finnish class reflects Finnish culture of quiet contemplation and shyness in public settings, thus, the dominant

teaching method is mainly focused on passive knowledge delivery and little or no engagement by the students. As a pre-emptive measure, to address possible student's fears and resistance to change, the modified lecture became a course objective and was evaluated by the students; also, in-class time was taken to explain why new active learning methods were implemented. The active learning methods ranged from long term projects beyond class and short in-class activities, and pre-planned questions intended to activate critical thinking. The results were evaluated using a self-evaluation pre and post-course survey,. Findings showed that students appreciated the active learning methods, recognizing their efficacy and motivational forces. Pappalardo and Gunn[6] studied a different case study of utilizing active learning strategies in first year university introductory chemistry courses. This study took place in the American university of Sharjah (AUS), which is based in the United Arab Emirates. While all courses were taught in English, students and faculty were from more than 35 different nationalities, exposed to many different views on teaching and learning. Active learning strategies included in-group activities, problem solving sessions, pre-class quizzes and group discussions.. The results showed that students found active learning techniques helpful for their encouragement, promoting better understanding, interactive manner, illuminating errors and clarifying concepts. Positive advantages viewed by instructors included: improved professor-students interaction and a more effective way to help students understand the material. On the other hand, the main drawback mentioned was time consuming aspects of preparation activities.
Even though there is clear evidence of the benefits of active learning over to traditional learning methods, most lecturers still adhere to the latter (Prince[8]). To better understand the gap between lecturing styles of different teachers, Pundak et al.[11] developed an attitudes questionnaire, on the basis of active learning instructors' experiences, which yielded six key domains characterizing their attitudes. This served as a diagnostic tool to identify how close faculty's current attitudes are to the ones common among active learning instructors, thus indicating the gap extant needed to be overcome. Findings revealed significant differences between attitudes of instructors practicing active learning and their colleagues who adhered to the traditional method in all six domains. Results showed that the most significant difference existed in the domain of ***activation of a big size class*** – active learning instructors believed it was possible to actively involve students in the learning process and create intimate working groups, whilst the others believed that trying to involve students would not lead to any progress and that it is not possible to achieve personal contact in groups, in a large class. The Five other domains represented, were ***'understanding versus quality'*** in which active learning instructors preferred to ensure that most of the students have reached an adequate level of understanding as opposed to prioritizing course curriculum demands; ***'function of instructor'*** in which active learning instructors recognized that their roles involved not only knowledge delivery, as in the traditional method, but also other roles such as recognizing students' difficulties, and also guiding, directing and encouraging students in various exercises; the possibility of more ***independent learning*** with use of

proper guidelines as opposed to expecting the students to only have knowledge of what was presented in class*;* students can ***develop new knowledge***, present new arguments and ideas; and the importance of ***involvement*** in class as opposed to only measuring success by passing the final exam. Another study by our research group (Pundak et al.[12]) compared e-teachers' attitudes with the previous results using the AIT questionnaire and showed similar results.
The following research is a part of a longitudinal research with several goals:

(a) To identify attitudes of "active learners" and lecturers towards active learning compared to those of traditional learners and lecturers.
(b) To examine the characteristics of "active lecturers"
(c) To examine students' attitudes towards active learning
(d) To compare academic achievements of students who study in traditional teaching methods and those studying according to active learning.

Research Questions

1. What are the attitudes towards active learning of students who are learning in the active approach and those learning in the traditional approach?
2. What are the differences in academic achievements between students who are learning chemistry in the active approach and those learning in the traditional approach
3. What are the perceptions towards active learning of lecturers who teach chemistry in the traditional approach and those who teach in the active approach?

3. Research Design and Methodology

Research Population

Students: 350 students participated in the research: 189 in active learning classes and 161 in traditional classes. The students studied mechanical engineering, electric and electronic engineering, biotechnology and participated in a general chemistry course.
70 students answered the attitudes questionnaire. 66 students answered all the questions and were included in the research (94% response rate): 35 that participated in active learning and 31 in traditional learning. 10 students participated in in-depth interviews.
Lecturers: 4 Chemistry lecturers participated in the research – 2 teaching in the traditional approach and 2 teaching according to the active learning approach.

Research Tools

Four research tools were used in this research:
For students: an attitudes questionnaire which was devised especially for this research, an achievements test at the end of the course and in-depth interviews at the end of the course.
For lecturers: in-depth interviews at the end of the course
Table one depicts the research stages and use of the different research tools:

Research Stage	Research Tools	Goal	Population	Ways of Analysis
1	Attitudes questionnaire (Weiser-Biton et al.[13])	Examine differences in attitudes of students in traditional and active learning	66	(a) Closed-ended questions: quantitative statistical. (b) Open-ended questions – content analysis by categories
2	Achievement tests at the end of the course	Examine differences in achievements of students in traditional and active learning	350	Quantitative-Statistical
3	In-depth interviews with students	Examine perceptions of active learning	10	Qualitative - Content analysis
4	In-depth interviews with lecturers	Examine perceptions of lecturers' role in traditional and active learning	4	Qualitative - Content analysis

Table 1: Research Stages

Attitudes' Questionnaire with Regard to Active Learning

The questionnaire was devised by the researchers, in 2008. The questionnaire contained 24 statements divided into four content aspects: team work and coping with problems, applying chemistry problems in daily life, showing comprehension in coping with chemistry problems versus memorizing, evaluating self-efficacy in learning new chemistry topics. Table 2 summarizes the desired attitudes in each of the four topics examined in this research. the questionnaire was validated by 7 experts in the field of science teaching. Each topic's reliability according to Cronbach α is presented in Table 2.

	Aspect	Preferred Attitudes	Non-preferred attitudes	Reliability according to Cronbach's α
Team	Active learning and team work in the course	Active learning and team work in the course contribute to understanding contents and motivation for self-learning	To succeed in the course the student does not have to show involvement, interest or special comprehension	0.704
Daily	Course contents' link to daily life	Course contents are linked to daily life outside class settings	Course contents required only for success in the course and have no link to daily life	0.945
Compreh	Comprehens	The course allows	It is important to	0.844

ension	ion versus memorizing in the course	for significant understanding of taught chemical contents	memorize contents. Only a few can really comprehend chemistry	
Self	Ways of learning in the course	Self-learning, practice in class and homework help learn and comprehend the material	It is better to get a ready-made solution or formula to learn the material and succeed in the course	0.899

Table 2: the four aspects of learning

4. Research Findings

Research findings relating to the first research question: The question was examined through an attitudes questionnaire answered by 66 students.

Attitudes of students in both groups: the traditional learning group (1) and the active learning group (2).

To examine differences in the attitudes of students learning in the traditional format and those participating in the active learning format, t tests were conducted with regard to the four aspects: teamwork (team), link to daily life (daily), comprehension (comprehension) and assessment of self-efficacy in learning new topics in chemistry (self). The findings are presented in Tables 3 and 4.

	Group	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
team	1	31	3.0538	.81239	.14591
	2	35	3.5381	.88846	.15018
daily	1	31	2.9419	.81354	.14612
	2	35	2.9143	.78409	.13253
comprehen-	1	31	3.1815	.77247	.13874
sion	2	35	3.4643	.91534	.15472
self	1	31	2.7677	.84237	.15129
	2	35	2.5714	.83475	.14110

Table 3: Differences in attitudes between traditional and active learning –four aspects

Dimension	**t-test**	**df**	**alpha**
team	-2.300	64	.025
daily	.140	64	.889
compr	-1.347	64	.183
self	.949	64	.346

Table 4: Significant differences between groups

Table 4 reveals that the differences between the groups were significant with regard to the issue of teamwork in learning.

A more detailed analysis of each item on the Attitudes questionnaire yielded significant differences in five statements as depicted in Table 5. Of the 24

statements, statistically significant differences were found in the level of agreement to five statements. The most significant difference between the groups pertained to the level of agreement with the statement "**Learning in groups helps me understand the material in class**" ($z=-5.03, p<0.01$). Approximately 83% of the students in the active learning group answered that they agreed or strongly agreed, compared to 23% of the control group students.

60% of the students in the active learning group compared to 17% of the students in the control group expressed the highest level of agreement with the statement that "Lessons that include **teamwork contribute more to understanding** than lectures without activity" ($z=-3.92$, $p<0.01$).

No.	Item	Group	N	Mean	Std. Dev.	t	P
2	Lessons that include teamwork contribute more to understanding than lectures without activity	traditional	30	3.60	.932	-4.262	.000*
		active	35	4.49	.742		
4	Practice in class helps me understand at least some of the developments or proof provided in lectures or textbooks	traditional	30	3.50	.861	-2.529	.014*
		active	35	4.06	.906		
7	Learning in groups helps me understand the material in class	traditional	30	3.00	.830	-6.245	.000*
		active	35	4.29	.825		
9	As a result of learning and practicing in class, I have high motivation to learn chemistry topics	traditional	30	3.13	.819	-3.081	.003*
		active	35	3.77	.843		
12	Practice in class enables me to learn in a way that allows me to understand the material better than in a traditional class	traditional	30	2.93	.828	-5.434	.000*
		active	35	4.17	.985		

*P<0.0005

Table 5: Statements with significant differences between students who learnt in an active format and those who learnt in a traditional format

Approximately 74% of the students in the active learning group compared to approximately 61% of the students in the control group agreed that **practice in class helped** them understand at least some of the developments or proof provided in lectures or textbooks. 16% of the students in the control group compared to approximately 6% of those the active learning group objected to this statement ($z=-2.25$, $p<0.05$).

Approximately 80% of the students in the active learning group compared to approximately 23% of the students in the control group expressed the highest

level of agreement with the statement "**Practice in class enables me to learn in a way that allows me to understand the material better than in a traditional class**". It should be noted that none of the students in the control group definitely agreed with this statement and most students in this group (approximately 53% expressed a neutral attitude towards it ($z=-4.83$, $p<0.01$).

Regarding the statement that only a few people with special skills can really understand chemistry, it was found that the active learning group tended to object to the statement more significantly than the control group (57% compared to 30% respectively) ($z=-2.13$, $p<0.05$).

Much like the difference with regard to the previous statement, approximately 83% of the students in the active learning group reported that they objected or definitely objected to the statement that course assignments did not contribute to understanding of the course materials. 40% of the students in the active learning group compared to approximately 67% of the students in the control group expressed significant objection to this statement ($z=-2.03$, $p<0.05$).

Apart from the statistically significant differences, other differences in the level of agreement were not far from significant. For instance, students in the active learning group tended to agree more with the statement that all they had to do in order to understand the main course ideas was to read the required materials, solve most of the problems assigned to them and pay attention to what is said in class ($z=-1.47$, $p=.143$). A high percentage of students in the active learning group agreed that as a result of learning and practicing in class, they have high motivation to learn chemistry topics ($z=-1.11$, $p=.192$). Students in the control group compared to those of the active learning group tended to agree more with the statement that the abilities they developed in the course were strictly theoretical and had nothing to do with reality ($z= -1.69$, $p=.092$). Similarly, students in the control group also agreed with the statement that dedicating a long time (half an hour or more) to solving one chemistry problem was a waste of time, and if they did not manage to find a solution fast, they had better approach someone who knew better who would provide a solution ($z=-1.37$, $p=.172$). It was also found that students in the active learning group tended to agree more with the statement that they are helped by mistakes they had made in their homework and exams as clues to what they ought to do so as to better understand the materials taught in class ($z=-1.58$, $p=.113$).

Differences between students in the active learning group and traditional learning pertaining to comments (positive and negative) written in the attitudes questionnaire

Analysis of the comments written in both group showed 32 comments in the active learning group, most of which (69%) were positive with regard to the different aspects, compared to only 17 comments, most of which (88%) were negative in the traditional learning group.

From the he traditional learning group's comments, it is apparent that the two isolated positive comments pertained to varied teaching strategies, mostly with regard to auxiliary materials such as booklets and exercises on the chemistry course website.

In contrast, most **positive comments** (approximately 64%) in the active learning group pertained to the activation offered by the approach, and the method of working in groups and teams. Approximately 23% of the comments sought to praise the alternative teaching methods. Approximately 9% of the positive comments pertained to the lecturers and their teaching capabilities. A single comment (approximately 5% of the positive comments in the active learning group) referred to the lecture-tutorial continuum, claiming the combination is most useful.

In contrast, comparison of the **negative comments** revealed that of those who commented about the lecture-tutorial continuum, about 67% were students from the traditional learning group and only 33% from the active learning group. In this category, half of the traditional learning group students' comments were that the practice is unnecessary, or that it does not constitute a complementary factor in understanding and assimilating the material taught in the lecture. Conversely, the negative comments pertaining to the lecture-tutorial continuum in the active learning group focused on the fact that the practice is too long and might lead to difficulties in concentration. Nevertheless, 3 comments in the active learning group pertained to difficulties in the teaching environment (focusing on a number of boards, faraway boards that do not enable students to see solutions accurately etc.), compared to 2 comments in the traditional learning group maintaining that the teaching method and teaching ways complicate learning, and that it is best to learn in the simple method. Three comments made by students from the traditional learning group with regard to the nature of lecturer-students' communication, whereas not even one negative comment regarding this issue was made by the active learning group.

Two comments (20% of the negative comments made by the active learning group) pertained to obstacles posed by group work, and by the fact that this type of learning might lead to unnecessary chats and diversions during the lesson.

Perceptions of active learning chemistry course

The perceptions of students in the active learning group were examined via open-ended questions in the questionnaire and in-depth interviews.

Analysis of the open ended questionnaires yielded the following findings: 83% would recommend participation this course to their friends; The reasons for this recommendation were: a different atmosphere, the lesson is interesting, the approach contributes to the social aspect of learning and allows for acquaintance among students in the course, learning in a group contributes to understanding and mutual enrichment, learning is active and more challenging than traditional passive learning.

A few students would not recommend the course to their friends as they claimed they found the traditional learning method more effective, and active learning required more efforts from them.

66% believed that active learning in Chemistry contributed to enhancing their motivation to learn. The students listed three reasons for that:

A. Learning in groups enhanced their self-confidence and contributed to their motivation to learn, both when they explained things to their peers and when they listened to explanations given by peers.

B. The lecturer personal attitude and support contribute to motivation even when the topic is difficult.

C. Active learning allows for practice during the lesson, which increases motivation to make efforts and succeed in solving problems during the lesson and at home.

Content analysis of the students' answers to the open-ended questions on the questionnaire and in-depth interviews revealed that students who learnt in the active approach format saw differences between the active learning approach and the traditional approach in four major categories: learning in groups, instant practice, lecturer-students' communication and varied ways of teaching.

The differences between traditional and active learning:

- Learning in groups: The group as a motivational factor – persuasion to learn "*Each student pushes the other*". Explanation at eye level – "*Sometimes it is clearer when a student explains to another student, and together, it is possible to cope with complex questions*.". Brainstorming –"*The group allows for presenting different perspectives, different ways of thinking*". Better understanding – "*The help of group members allows for better understanding of the material*."
- Instant practice- Instant practice helps understand - "Active practice in class is more significant than passive absorption in traditional learning and assimilation at home"
- Lecturer-students communication: Students can be guided by the lecturer – "*This does not happen in traditional learning. In active learning, there is a possibility of asking questions and the lecturer always answers. Always!*". Sense of the lecturer's caring and interest – "*In traditional learning courses the lecturer does not care whether you have understood*." "*In active learning there is more investment on the part of the lecturer*."
- Varied ways of teaching: In active learning there is more demonstration via practice and animations, presentations on the board and computerized illustrations.

The in-depth interviews referred to the advantages and disadvantages of active learning in small groups and raised an important point: the students noted that learning in a group allows for enrichment, collaboration and explanations at "eye level" by peers. However, it also has the disadvantage of noise, difficulties in concentration, difficulties in sharing with peers and also the phenomenon of students who rely on others and are inactive.

Differences in academic achievements.

The courses were taught according to both active and traditional teaching approaches for three years. The active approach focused on the students and provided constructive feedback after each lesson, encouraging learning in the group and discussing the group's products with the class. The traditional approach tended towards imparting the material through the lecturer, while providing summative feedback in the final exam. First, the students' achievements were examined in final exams and their final grades across the three years and the data is presented in Table 7.

Assessment	Group	N	Mean	Std. Dev.	t	P
Exam	tradition	161	66.73	17.551	-5.695	0.000*
	active	189	76.89	15.818		
Final	tradition	161	67.56	17.457	-4.749	0.000*
	active	189	76.13	16.285		

*P<0.0005

Table 7: Students' achievements in active and traditional learning – chemistry course

Table 6 reveals that there is a significant difference between achievements in the final test and the final grades of the students who learnt in the active format compared to students who learnt in the traditional format: the former had higher achievements than the latter.

Lecturers' perceptions towards active learning

The perceptions of the chemistry lecturers were examined in in-depth interviews. The "active lecturers" went through a process of changing attitudes as a result of the great efforts they had invested in developing new ways of teaching and activating the students throughout the course.

Both active and traditional lecturers claimed that the students experienced genuine learning and understanding problems, that there was a need to provide solutions for the problems in a different learning environment, and that there is great significance in focusing on conceptual understanding and not only on solution techniques which reflects learning in the traditional course.

During the active learning course the lecturers reported a number of problems that had arisen such as "*sitting in a group encourages students to speak to each other, not only to their neighbors*" and that "*noise is created and it is hard to get the students to work and quieter discussion*". Additionally, lecturers discussed problems in teaching the entire syllabus "*relatively low rate of teaching the material in comparison to traditional teaching*." In contrast, the lecturers reported that students had been highly active and contributed to the discussion and showed great involvement, "*in active learning lessons the students discuss the topic and then they start working*." "*In a traditional I am the one who*

determines the pace. Here, each student works at his pace, and I only set the framework."
The lecturers were required to develop tasks and activities which will engage students in learning the material in-depth. Demonstration films and animations were required to explain the principles of learning, practical experiences, analysis of situations and development of formulae, equations and formulating principles and rules in the different units, and simulations that illustrate the chemical processes, "Lectures were enriched by animations." Indeed all this required a good deal of work, creativity and planning on the part of the lecturers, "*teaching in the active learning class requires a great deal of preparation compared to traditional teaching, and more time has to be dedicated to communication with students*." Nonetheless, the investment begot results; the lecturers reported that students had indeed shown more profound understanding, interest, curiosity, and will to invest in the course. Relevance to daily life brought students closer to the field and sparked their interest, "*Students are alert and involved*."
Teaching in the active learning course influenced the "active" lecturers, who reported changes in their ways of teaching in other courses as well and in their "traditional" teaching, in organizing their lectures, carrying tasks and activating the students.

5. Discussion and Conclusions

This research reveals that active learning has numerous advantages over traditional learning**:**

1. Learning and practice in a team contribute more to **profound conceptual understanding** and to the ability to apply the material to complex chemistry problem than passive lectures s. These results are congruent with the findings of the interviews and the questionnaires**.**
2. Active learning in chemistry contributed to **improving the students' motivation** to learn chemistry because learning in a group allows for enrichment and collaboration and adds to their self-confidence and image.
3. **Lecturer-students communication –** the lecturer's support and explanations at "eye level" as well as varied ways of teaching encourage learning and motivation.
4. Active learning allows for instant practice and application in the lesson, which allows for **profound understanding and experiencing success** which will accompany the students to success in solving problems at home as well.
5. The highlight of this research is in the **students' achievements**. The achievements in the final exam and the final grades of students who learned in the active format were significantly higher than those of those who learnt in the traditional approach.

It should be noted that the active approach also has disadvantages such as noise resulting from the group discussions, leading to difficulties concentrating, mostly for students with attention disorders. Furthermore, there are students

who find it hard to study with peers, and those who are inactive and rely on others.

Nevertheless, the advantages outweigh the disadvantages, and hence, it appears that the research findings support the recommendation to train lecturers in chemistry and other science subjects to embrace the active and collaborative learning approach.

References

1. Maharshak, A. Pundak, D. (2004). Active Physics Learning – Combining the Marketing Concept with Information Technology. Journal of Educational Technology Systems, Baywood Publishing Company, Inc., NY. Vol. 32, 4, pp 399-418.
2. Pundak, D., Rozner, S. (2001). Teaching in Real Time - Just in Time Teaching. Studies in Technology and Science, 34, 17 - 20.
3. Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64-74.
4. Michael, J. (2006). Where's the evidence that active learning works?. Advances in physiology education, 30(4), 159-167.
5. Hall, S. R., Waitz, I., Brodeur, D. R., Soderholm, D. H., & Nasr, R. (2002). Adoption of active learning in a lecture-based engineering class. In Frontiers in Education, 2002. FIE 2002. 32nd Annual (Vol. 1, pp. T2A-9). IEEE.
6. Pappalardo, L., & Gunn, C. (2013). Utilizing active learning strategies to enhance learning in first year University Chemistry courses. Learning and Teaching in Higher Education: Gulf Perspectives, 10(1).
7. Beichner, R.J.; Saul, J.M.; Allain, R.J.; Deardorff, D.L.; Abbott, D.S. (2000). Introduction to SCALE-UP: Student-Centered Activities for Large Enrollment University Physics. Retrieve from http://eric.ed.gov/?id=ED459062
8. Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, *93*(3), 223-231.
9. Paulson, D. R. (1999). Active Learning and Cooperative Learning in the Organic Chemistry Lecture Class. *Journal of Chemical Education*, *76*, 1136.
10. Larson, D., & Ahonen, A. M. (2004). Active learning in a Finnish engineering university course. European journal of engineering education, 29(4), 521-531.
11. Pundak, D., Herscovitz, O., Shacham, M., & Weiser-Biton, R. (2009). Instructors' Attitudes toward Active Learning.*Interdisciplinary Journal of E-Learning and Learning Objects*,*5*(1), 215-232.
12. Pundak, D., Herscovitz, O., Shacham, M., & Weiser-Biton, R. (2010). Attitudes of Face-to-Face and e-Learning Instructors toward 'Active Learning'. *European Journal of Open and Distance Learning*.
13. Rivka Weiser-Biton, Miri Shacham, David Pundak, Active learning vs. traditional learning in the view of lecturers and students, The 10th Interdisciplinary Research Conference ORT Braude College.

Loss of conservation of graph centralities in reverse-engineered transcriptional regulatory networks

Holger Weishaupt[1], Patrik Johansson[1], Christopher Engström[2], Sven Nelander[1], Sergei Silvestrov[2], and Fredrik J Swartling[1]

[1] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Dag Hammarskjöldsväg 20, 751 85 Uppsala, Sweden (E-mails: holger.weishaupt@igp.uu.se, patrik.johansson@igp.uu.se, sven.nelander@igp.uu.se, fredrik.swartling@igp.uu.se)

[2] Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Box 883, 721 23 Västerås, Sweden (E-mails: christopher.engstrom@mdh.se, sergei.silvestrov@mdh.se)

Abstract. Graph centralities are often used to prioritize disease genes in transcriptional regulatory networks. Studies on small networks of experimentally validated interactions emphasize the general validity of this approach and extensions of such findings have recently also been proposed for networks inferred from expression data. However, due to the noise inherent to expression data, it is largely unknown how well centralities are preserved in such networks. Specifically, while previous studies have evaluated the performance of inference methods on synthetic expression, it has yet to be established how the choice of method can affect individual centralities in the network. Here we compare two centralities between reference networks and networks inferred from corresponding simulated expression data using a number of related methods. The results indicate that there exists only a modest conservation of centrality measures for the used inference methods. In conclusion, caution should be exercised when inspecting centralities in reverse-engineered networks and further work will be required to establish the use of such networks for prioritizing genes.
Keywords: Transcriptional network inference, simulated expression, centrality.

1 Introduction

With the increasing amount of 'omics' data made available to researchers during the last decades, biological network analysis has rapidly grown in its importance as one of the predominant methods of studying the underlying interactions and relationships between biological entities (Zhu *et al.*[1]). Current network based studies of topological properties of the related interactomes place particular interest on methods for clustering, pathway analysis, motif identification and graph based centrality measures on vertex or edge level (Ma and Lin[2], Zhang *et al.*[3], Aittokallio and Schwikowski[4]). Among these methods, vertex centrality can be considered the foremost technique in assigning importance to nodes in a variety of related networks (Zhang *et al.*[3], Koschützki and

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

Schreiber[5]). Importantly, it has previously been established that with respect to candidate gene identification, biological network analysis using centrality enrichment of nodes might in certain situations prove advantageous compared to a simple meta-analysis of genomic datasets (Langfelder *et al.*[6]).

Initial investigations of centralities in networks of *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Caenorhabditis elegans* have suggested that developmentally and functionally essential proteins, i.e. proteins whose disruption leads to embryonal lethality, might be associated with high degree, closeness or betweenness centralities (Jeong *et al.*[7], Joy *et al.*[8], Hahn and Kern[9]). More recently, graph centralities have also been investigated for the prediction of disease or cancer genes in human gene and protein networks (Huang *et al.*[10], Ortutay and Vihinen[11], Sidanni *et al.*[12], Izudheen and Mathew[13]). However, such studies have mainly been performed on networks built from databases of validated biological interactions. Since such data usually contains only limited numbers of generic interactions and biological entities, any related screen might miss important genes or might not be suitable to investigate the interactom specific for a given disease.

Alternatively, researchers often fall back on centrality based prioritization of genes from transcriptional regulatory networks (TRNs) reverse-engineered from expression data (Basso *et al.*[14], Jörnsten *et al.*[15], Emmert-Streib *et al.*[16], Cordero *et al.*[17], Knaack *et al.*[18]), which has become easily accessible for cells and tissues of normal and diseased conditions and can cover all known genes in the genome. Numerous methods for reverse-engineering of regulatory networks from expression data have been developed during the last decade (Margolin *et al.*[19], Faith *et al.*[20], Meyer *et al.*[21], Langfelder *et al.*[22], Altay and Emmert-Streib[23], de Matos Simoes and Emmert-Streib[24], Huynh-Thu *et al.*[25], Yip *et al.*[26], Zhang *et al.*[27]). A problem with this approach however is the noise inherent to most expression datasets and the uncertainty in any network reconstructed from such data with available inference methods (Margolin and Califano[28], Chalancon *et al.*[29]). Accordingly, a lot of effort has been dedicated to the evaluation of different aspects of inference accuracies and consistencies between such methods (Altay and Emmert-Streib[30], Schaffter *et al.*[31], de Matos Simoes *et al.*[32], Marbach *et al.*[33], Marbach *et al.*[34]).

However, it is still largely unexplored, how well the centralities in such inferred networks agree with the centralities of genes in the true underlying biological network. Considering that different methods are likely to make different systematic errors in inferring gene interactions (Schaffter *et al.*[31], Marbach *et al.*[34]), it is reasonable to also expect different effects on the conservation of centrality values between inferred and true biological networks.

Here we obtain reference biological networks from a database of interactions in *Escherichia coli* and from the *Pathway-Commons* database (Cerami *et al.*[35]), generate simulated gene expression for these networks using a model of stochastic differential equations, and utilize this data to reconstruct networks using a number of different inference methods. These benchmark datasets are then employed to compare the agreement of degree and betweenness centrality

measures between the reference and inferred networks for the different inference methods.

2 Generation of benchmark datasets

To determine how well centralities are conserved between true networks and networks inferred from gene expression data, we are here relying on benchmark datasets consisting of biological reference networks and synthetic expression data simulated from such reference networks according to Schaffter *et al.*[31]. Two types of reference networks are used here as described below

Ecoli250: A set of 100 networks with 250 nodes each, extracted using the GeneNetWeaver (GNW) software (Schaffter *et al.*[31]) from the included *ecoli-regulonDB-6-7* dataset. The entire ecoli database contained 1565 nodes connected by 3758 edges. For each network ten random regulators (nodes with high out-degree centrality) were selected and additional nodes added from the neighborhood using a greedy search (Schaffter *et al.*[31]).

PC200: A set of 100 networks with 200 nodes each, extracted from the *Pathway-Commons* database. From the 4667832 edges (connecting 19006 nodes) with diverse interaction types in the *Pathway-Commons* database, only the subset of 105178 edges (connecting 13390 genes) of transcriptional regulatory nature, i.e. with an *'controls-expression-of'* interaction type, was chosen as a reference. Each benchmark network was then generated by selecting a random seed node and adding nodes from the neighborhood using a greedy search based on degree centralities.

Based on the selection of seed nodes, the *Ecoli250* networks have a more nodular structure than the *PC200* networks (compare Figure 1), while both sets of networks exhibit a roughly scale free topology, as exemplified by a large number of nodes with low degree values and a small number of high degree values (compare Figure 2), which is one of the characteristic properties of biological networks (Albert[36]).

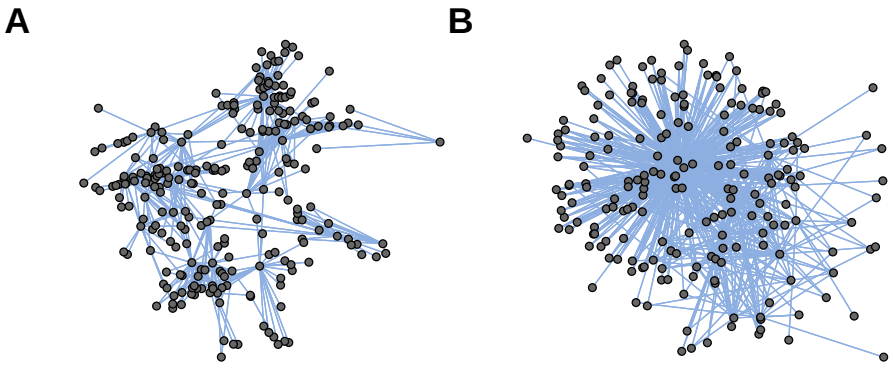


Fig. 1. Example of *Ecoli250* and *PC200* network structures. The network structure of one example of an *Ecoli250* network **(A)** and one *PC200* network **(B)**, without displaying the direction of links.

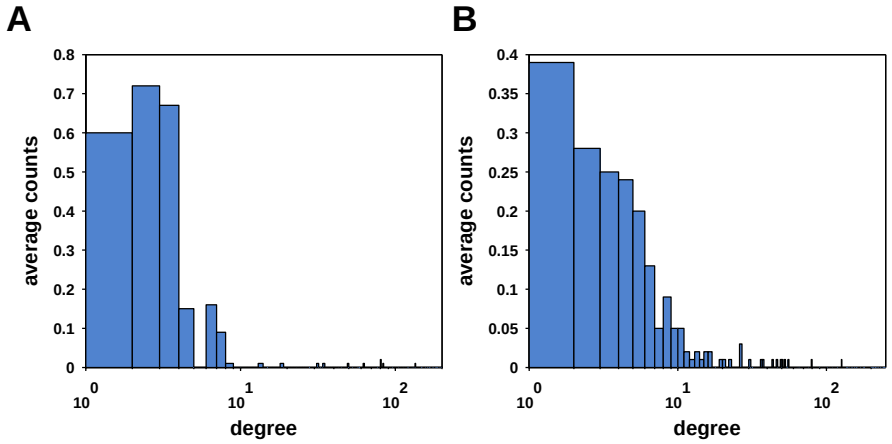


Fig. 2. Distribution of degree values in *Ecoli250* and *PC200* networks. The histograms show the number of nodes with certain degree values averaged over all *Ecoli250* networks **(A)** and all *PC200* networks **(B)**.

For each reference network, expression data was then simulated using GNW, which makes use of stochastic differential equations (SDEs) to model transcriptional and translational processes (Schaffter *et al.*[31]). Specifically, the noise-free rate of concentration changes $\frac{dY_t}{dt}$ for mRNA and proteins are simulated using an ordinary differential equation (ODE) of the form

$$\frac{dY_t}{dt} = G(Y_t) - D(Y_t),$$

where $G(Y_t)$ describes the contribution through production at time t and $D(Y_t)$ describes the amount of degradation of the product at time t. The processes under biological noise are then simulated using the SDE of the form

$$\frac{dY_t}{dt} = G(Y_t) - D(Y_t) + a(\sqrt{G(Y_t)}\eta_G - \sqrt{D(Y_t)}\eta_D),$$

where a is a constant and η_G and η_D are two independent white-noise signals with zero mean. In addition to molecular noise GNW also adds a model of instrumental noise introduced in typical microarray experiments (Schaffter *et al.*[31]). Variable expression is then further generated through multifactorial variations in the initial activation of genes in the model (multifactorial data), through a 50% decrease in a single gene's initial activation (knockdown data) and through the complete inactivation of a single gene (knockout data).

In this study we used a white-noise term in the SDEs with standard deviation of 0.05 and generated 250 and 200 expression samples each of multifactorial, knockdown and knockout data for each *Ecoli250* and *PC200* network, respectively. These individual three datasets were then combined to obtain the final simulated expression dataset for each network.

3 Inference of transcriptional networks

From the simulated expression data, TRNs were inferred using ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks; Margolin *et al.*[19]), CLR (Context Likelihood of Relatedness; Faith *et al.*[20]) and MRNET (Minimum Redundancy NETwork; Meyer *et al.*[21]), all three of which are based on mutual information and are implemented in the R/Bioconductor package MINET (Meyer *et al.*[37]), as well as the correlation based WGCNA (Weighted Gene Co-expression Network Analysis; Langfelder *et al.*[22]) method. Importantly, since we were interested in centralities rather than the exact prediction of individual links, all links in the inferred and reference networks were considered to be undirected once the networks had been inferred.

4 Estimating the accuracy of inferred networks

Once the inferred networks were obtained, network accuracies were estimated using the area under the Receiver-Operator-Characteristic (ROC) curve (auROC) and the area under the Precision-Recall (PR) curve (auPR) for comparability with previously published results (Schaffter *et al.*[31]). ROC curves and PR curves were obtained by first sorting the network links predicted by the inference methods based on descending absolute strength. Subsequently, false-/true-positive rates and precision/recall values were sampled at certain intervals by including increasing numbers of highly scored links (compare Figure 3). The sampled curves where then linearly interpolated including also the additional start and end points (0,0) and (1,1) for ROC curves, as well as (0,1) and (1,$\mathrm{Pr}_{\max}$) for the PR curves (compare Figure 3), where $\mathrm{Pr}_{\max} = \frac{l}{(N^2-N)/2}$ is the maximal achievable precision with l denoting the number of (undirected) links and N denoting the number of nodes in the reference network.

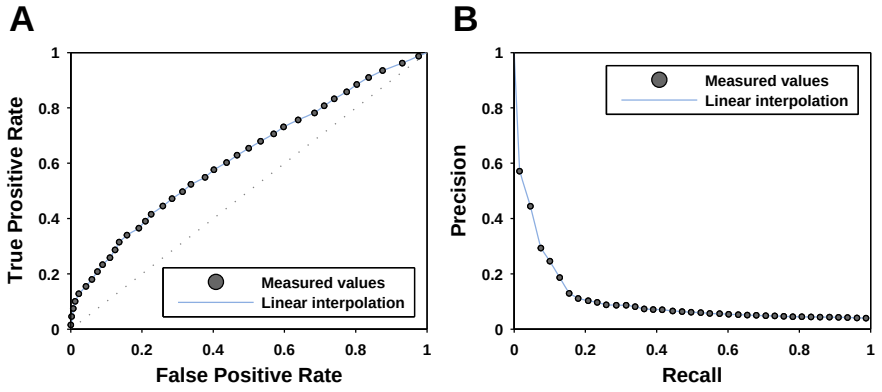


Fig. 3. Sampling of ROC and PR curves. ROC curve **(A)** and PR curve **(B)** for one WGCNA inferred *Ecoli250* network including sampled true-positive and false-positive rates or recall and precision values, respectively, and the interpolations over the sampled data points.

However, one concern with this approach was the difference in total number of predicted links observed for the individual methods, which led to substantial differences in the obtainable recall or true-positive rates (compare Figure 4).

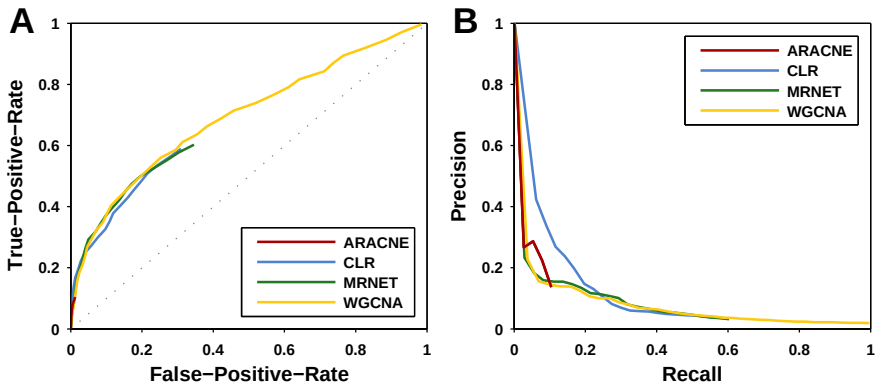


Fig. 4. Differences in maximum recall and true-positive-rates between network inference methods. True-positive and false-positive rates **(A)** as well as recall and precision values **(B)** of the four used inference methods calculated on one of the *Ecoli250* networks interpolated up to the maximum true-positive rate or recall value, respectively, obtained for the individual method.

To remove potential bias based on different amounts of missing data between the methods, we decided to (1) decrease the default threshold below which links in vertex-triplets are removed in ARACNE by 0.05 (Margolin *et al.*[19], Meyer *et al.*[37]) and (2) sample all networks only up to the smallest maximum recall value obtained among all inferred networks (compare Figure 5).

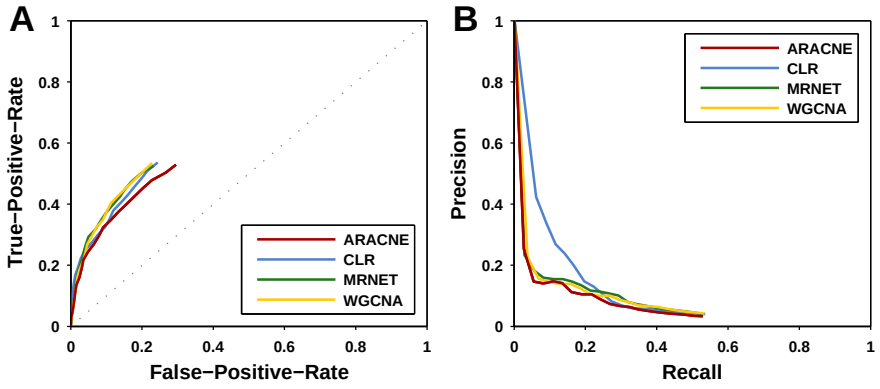


Fig. 5. ROC and PR curves after filtering links based on maximum true-positive rates and recall values. True-positive and false-positive rates **(A)** as well as recall and precision values **(B)** for the four used inference methods calculated on the network in Figure 4 and interpolated up to the smallest maximum true-positive rate or recall value, respectively, among all inference methods and *Ecoli250* networks.

The method specific auROC and auPR values where then calculated for all networks in the *Ecoli250* and *PC200* sets using the *trapz* function in MATLAB[38]. Both approaches discussed above were employed, i.e. either using all available interactions obtained by the inference methods with default parameters as illustrated in Figure 4 or using the interactions filtered based on the maximum recall value as outlined in Figure 5. The results are depicted in Figure 6A-B and Figure 6C-D, respectively.

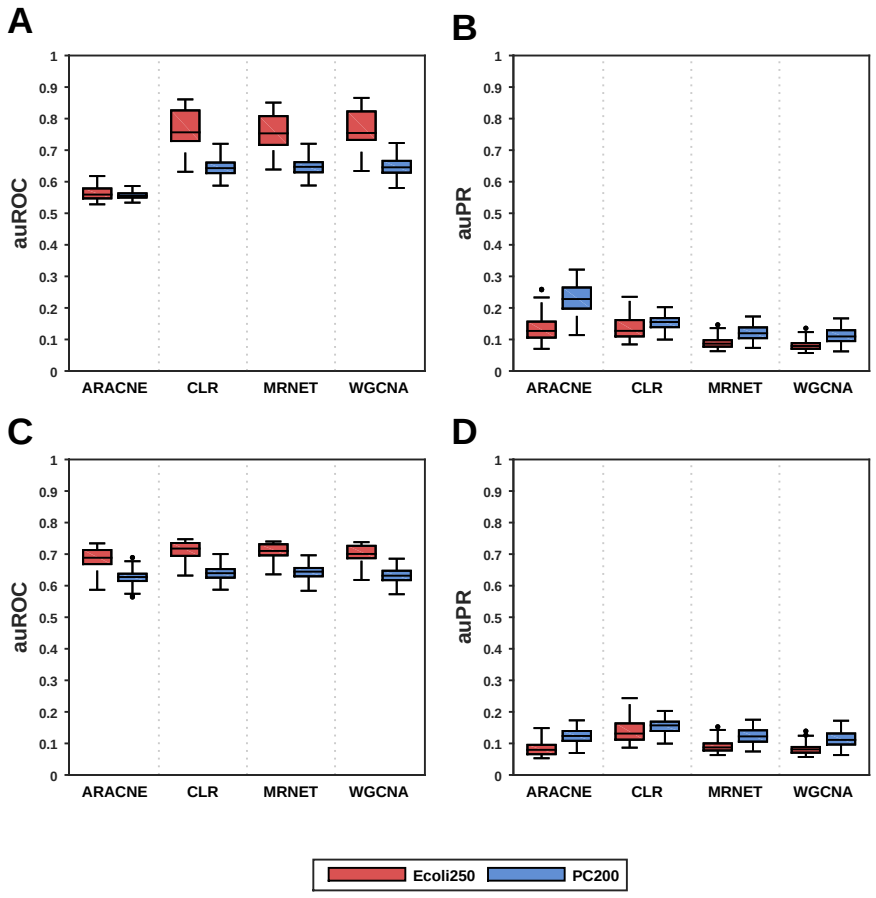


Fig. 6. Box-and-whisker plots of auROC and auPR values. The auROC **(A,C)** and auPR **(B,D)** values of the networks inferred by four different methods for the *Ecoli250* (red) and *PC200* (blue) are depicted as box-and-whisker-plots, with the median values represented by horizontal lines inside each box, lower and upper box borders indicating 25th and 75th percentiles, respectively, regions between whiskers including all non-outlier values, and dots representing outliers. The upper panel **(A-B)** depicts results for the first approach described above, i.e. including all available interactions obtained by the inference methods, while the lower panel **(C-D)** displays the results on filtered interactions.

Networks derived by ARACNE had generally lower auROC values than networks inferred by the other three methods (compare Figure 6A,C). Apart from the *PC200* comparison to WGCNA inferred networks with filtered interactions (Figure 6C), these differences in mean auROC values were significant at the $\alpha = 0.05$ level according to a two-sided Welch T-test (Welch[39]). Of note however, the difference of mean auROC values between ARACNE and the other methods was substantially diminished when using filtered interactions (Figure 6C), suggesting that part of the differences in Figure 6A can be explained due to different numbers of predicted links between methods. On the same note, while we observed comparable ranges of *Ecoli250* auPR values and a significantly higher mean *PC200* auPR for ARACNE as compared to CLR with default interactions (Figure 6B), auPR values for ARACNE in both network sets were substantially reduced when using filtered interactions, with CLR exhibiting significantly higher mean auPR values compared to all other methods (Figure 6D). The results in Figure 6A appear consistent with the original results presented by Schaffter et al. (2011), who documented comparable auROC differences between the ARACNE and CLR methods also coupled to low auPR values, when tested on expression data simulated with knock-out and knock-down perturbations (Schaffter *et al.*[31]).

In addition, we observed that the *PC200* networks exhibited significantly lower auROC values and significantly higher auPR values than the respective *Ecoli250* networks. Considering the different structures of the networks in the *Ecoli250* and *PC200* sets, this difference points towards a potential impact of the overall network structure on the inference accuracy of the used methods.

5 Conservation of centralities in inferred networks

In order to investigate the agreement of centrality measures between true and reverse-engineered networks for the four inference methods, we focused here on degree centrality and betweenness centrality (Freeman[40]).

Given the symmetric adjacency matrix $A = \{a_{n\times n}\}$, where n is the number of genes and $a_{ij} = a_{ji} = 1$ if there exists a biological interaction between gene i and j and $a_{ij} = a_{ji} = 0$ otherwise, then degree centrality is defined as

$$C_D(v_i) = \sum_{j=1}^{n} a_{ij},$$

and betweenness centrality can according to Brandes[41] be defined as

$$C_B(v_i) = \frac{1}{2} \sum_{k\neq i\neq l} \frac{\sigma_{kl}(v_i)}{\sigma_{kl}},$$

with v_i denoting the i'th gene, σ_{kl} denoting the number of shortest paths between two nodes v_k and v_l and $\sigma_{kl}(v_i)$ denoting the number of shortest paths between v_k and v_l that traverse through node v_i. Importantly, in order to allow the application of betweenness centrality on disconnected networks, if there is no path between nodes v_k and v_l in the network then one sets

$$d(v_k, v_l) \equiv \infty, \quad \frac{\sigma_{kl}(v_i)}{\sigma_{kl}} \equiv 0.$$

Degree and betweenness measures were then computed using the Matlab-BGL package (Gleich[42]) in the reference network and three different variants of the network obtained by each of the four individual network inference methods. The three different variants were established by including the same number of top scored links as present in the reference network (N_1), or using only a number of top scored links equal to 10% ($N_{0.1}$), or 50% ($N_{0.5}$) of the total number of links present in the reference network. This approach was chosen, since there might be differences in centrality conservations depending on the number and scores of links included in the generation of the reconstructed networks.

In order to determine the agreement of centralities between the reference and inferred networks, we were interested in comparing the ranks of nodes under the individual centralities between the two networks using a rank correlation metric. Importantly, since centralities are here understood as a method of prioritization of nodes, it is obvious that more importance should be placed on high ranked nodes, which dictates the use of a weighted rank correlation metric. During the recent years many different weighted variants of Spearman's ρ and Kendall's τ have been discussed in the literature, compare Dancelli *et al.*[43], Tarsitano[44], Pozzi *et al.*[45]. Since the chosen centralities in combination with the inferred network structures, which will often miss a subset of nodes thus receiving a zero centrality, implies a high number of ties in the rankings, Kendall's τ and specifically the τ_B variant, designed particularly for rankings with ties, appears here to be the preferable method. Accordingly, we use here the weighted Kendall's τ_B measure proposed by Pozzi *et al.*[45], which is computed for the two variables y^i and y^j over a running window Δt as

$$\tau_{ij}^{w} = \sum_{k=1}^{\Delta t-1} \sum_{l=k+1}^{\Delta t} w_{kl} \operatorname{sgn}(y_k^i - y_l^i) \operatorname{sgn}(y_k^j - y_l^j),$$

where we define the exponential smoothing weights as

$$w_{kl} = \frac{e^{\frac{2\Delta t-k-l}{\theta}}}{\sum_{v=1}^{\Delta t-1} \sum_{w=v+1}^{\Delta t} e^{\frac{2\Delta t-v-w}{\theta}}},$$

with $l > k$ and the constant θ indicating the specific time of the weights, such that

$$\sum_{k=1}^{\Delta t-1} \sum_{l=k+1}^{\Delta t} w_{kl} = 1$$

Importantly, the weights here are just the weights defined by Pozzi *et al.*[45] in reverse order, i.e. decreasing on both subscripts in order to place more importance on the nodes with high ranks in the reference networks. Δt was chosen equal to 250 and 200 for the *Ecoli250* and *PC200* datasets, respectively, and $\theta = \frac{\Delta t}{3}$ according to the recommendations of the authors (Pozzi *et al.*[45]).

The results of the weighted rank correlation analysis of centralities between reference and inferred networks are shown in Figure 7.

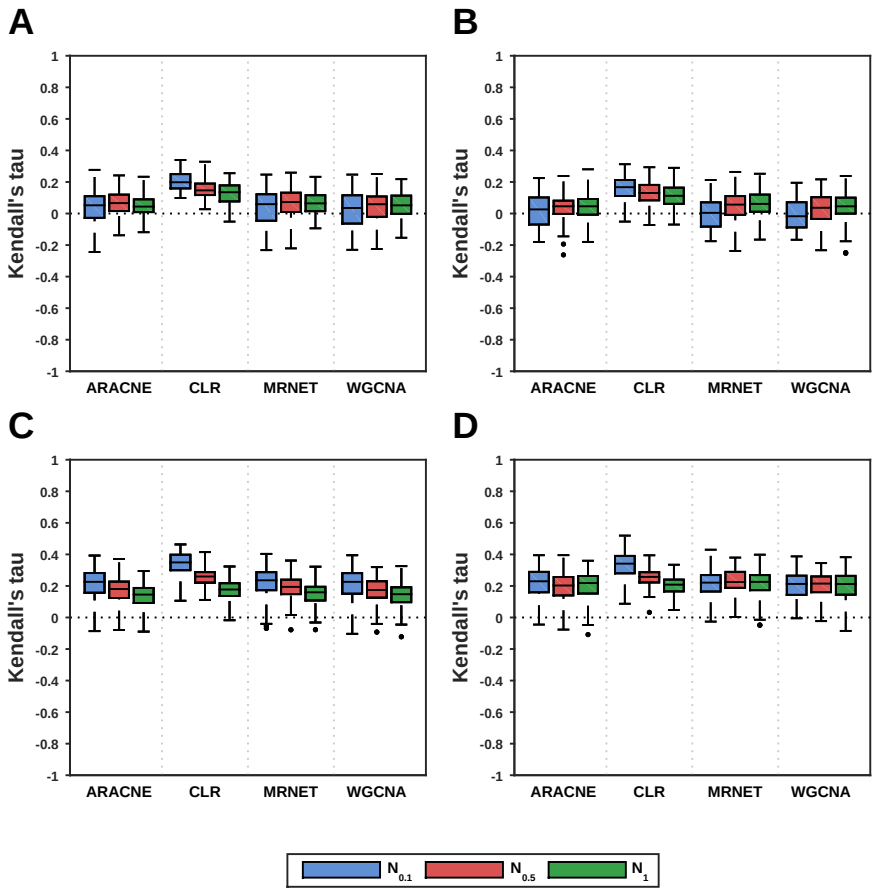


Fig. 7. Kendall's τ of centralities in reference and inferred networks. The Kendall's τ values for degree **(A,C)** and betweenness centralities **(B,D)** between reference and inferred networks for *Ecoli250* **(A-B)** and *PC200* **(C-D)** are shown as box-and-whisker-plots for the four inference methods. Centralities have been compared for three variations of the inferred networks; $N_{0.1}$ (blue), $N_{0.5}$ (red), N_1 (green).

As can be seen from Figure 7 none of the used inference methods produced a network with a high correlation of degree or betweenness centralities between reference and inferred networks, considering the three network variants $N_{0.1}$, $N_{0.5}$ and N_1. The highest mean correlations for both centrality types in *Ecoli250* and *PC200* are seen with the CLR method across all three network variants except the case of betweenness in the *PC200*-N_1 variant. Interestingly, in all comparisons of degree centrality for the *PC200* networks as well as the other three CLR results, the $N_{0.1}$ networks achieved a better mean correlation than the $N_{0.5}$ networks with a further reduction observed in the N_1 networks.

In order to explore the observed discrepancies in centralities between reference and inferred networks further, we inspected the expected and predicted links in the three network variants (compare example in Figure 8).

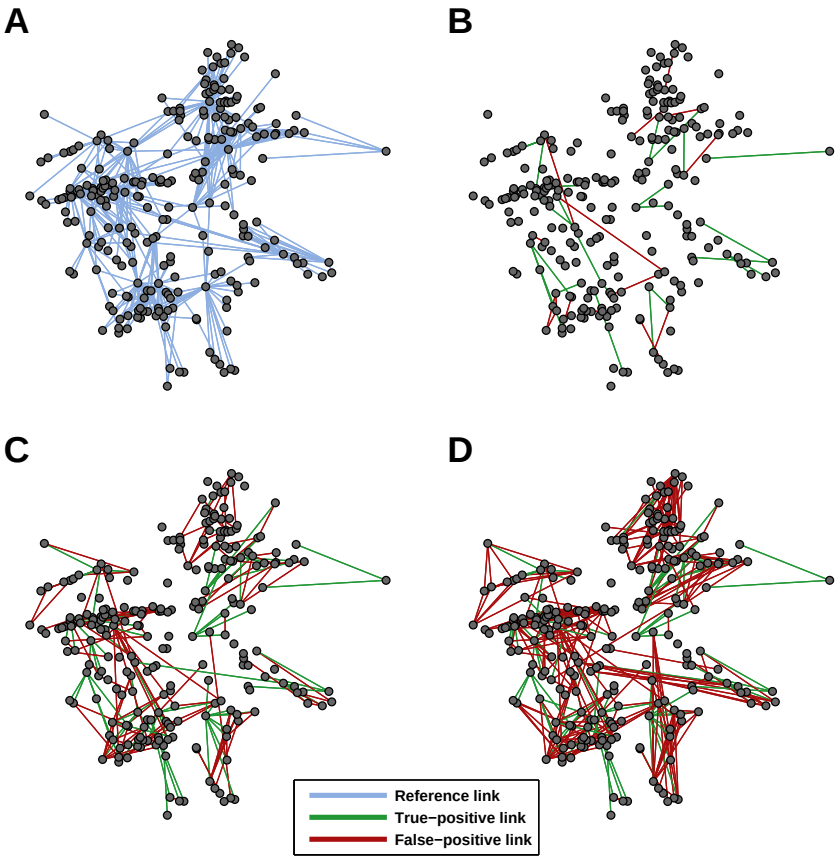


Fig. 8. Inspection of CLR inferred links. The reference *Ecoli250* network from Figure 1 **(A)** is shown together with the CLR inferred $N_{0.1}$ **(B)**, $N_{0.5}$ **(C)** and N_1 **(D)** networks, i.e. using a number of highly scored predicted links equal to 10%, 50% and 100% of the total number of links present in the reference network. Reference links in the benchmark network are shown in blue, true-positive predicted links are shown in green and false-positive predicted links are shown in red.

From the inspection of predicted links, as exemplified by the networks in Figure 8, it is obvious that when choosing only a small number of highly scored links from the network inference results (compare Figure 8B), one can among those links obtain a large proportion of true-positives, but the number of links does not suffice to reconstruct enough of the overall topology of the reference network to obtain a high degree of agreement between centralities. When increasing the number of links in the inferred networks, however, the topology

appears to become dominated by false-positive links (compare Figure 8C-D), which might lead to a more random distribution of centralities in the inferred networks.

6 Conclusion and future prospectives

The results discussed in this study show that the inference methods are indeed able to accurately predict a certain proportion of biological links from simulated expression data. However, the number of correctly predicted links does not appear to suffice in order to also reconstruct the underlying network topology to a degree necessary to obtain a high conservation of graph centralities between reference and inferred networks. In addition, it has previously been reported that network inference methods might show a lower prediction accuracy for links connecting to a high degree centrality node as compared to links connecting to low centrality nodes (Marbach *et al.*[33]). Together, these two observations might account for some of the revealed dissimilarities of centralities between reference and inferred networks.

It is possible that at least part of the observed disagreements of centralities can be accounted to difficulties of inferring networks from the chosen synthetic expression data and that the tested methods might actually perform better on true microarray expression data. It would hence be important to repeat the outlined experiments even using other models of expression data, compare for instance (Van den Bulcke *et al*[46], Langfelder *et al.*[22]). Additional and more comprehensive studies will also be required to evaluate the performance of other recently developed network methods with respect to the conservation of centrality methods in inferred networks. Specifically, considering the fact that methods make different systematic errors in inferring gene interactions (Schaffter *et al.*[31], Marbach *et al.*[34]) and that accurately predicted links also differ between networks (data not shown), it would be interesting to investigate how a combination of highly scored links from multiple network methods performs in reconstructing the relevant network topology (Marbach *et al.*[33]).

In summary, more work will be needed to establish the usefulness of networks inferred from expression data using current methodology for the purpose of centrality based gene prioritization.

7 Acknowledgments

This work was supported by grants from the Worldwide Cancer Research (formerly known as AICR) and the Swedish Childhood Cancer Foundation.

References

1. Zhu, X., Gerstein, M., and Snyder, M. "Getting connected: analysis and principles of biological networks." Genes & development 21.9 (2007): 1010-1024.
2. Ma, X, and Gao, L. "Biological network analysis: insights into structure and functions." Briefings in functional genomics 11.6 (2012): 434-442.

3. Zhang, M., Deng, J., Fang, C. V., Zhang, X., and Lu, L. J. "Molecular network analysis and applications." (2011). In Alterovitz, G., and Ramoni, M., eds. Knowledge-based bioinformatics: from analysis to interpretation (pp. 253). Wiley.
4. Aittokallio, T., and Schwikowski, B. "Graph-based methods for analysing networks in cell biology." Briefings in bioinformatics 7.3 (2006): 243-255.
5. Koschützki, D., and Schreiber, F.. "Centrality analysis methods for biological networks and their application to gene regulatory networks." Gene regulation and systems biology 2 (2008): 193.
6. Langfelder, P., Mischel, P. S., and Horvath, S. (2013). When Is Hub Gene Selection Better than Standard Meta-Analysis? Plos One, 8(4).
7. Jeong, H., Mason, S. P., Barabsi, A. L., and Oltvai, Z. N. "Lethality and centrality in protein networks." Nature 411.6833 (2001): 41-42.
8. Joy, M. P., Brock, A., Ingber, D. E., and Huang, S. "High-betweenness proteins in the yeast protein interaction network." BioMed Research International 2005.2 (2005): 96-103.
9. Hahn, M. W., and Kern, A. D. "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks." Molecular Biology and Evolution 22.4 (2005): 803-806.
10. Huang, Y. J., Hang, D., Lu, L. J., Tong, L., Gerstein, M. B., and Montelione, G. T. "Targeting the human cancer pathway protein interaction network by structural genomics." Molecular & Cellular Proteomics 7.10 (2008): 2048-2060.
11. Ortutay, C., and Vihinen, M. "Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies." Nucleic acids research 37.2 (2009): 622-628.
12. Siddani, B. R., Pochineni, L. P., and Palanisamy, M. "Candidate Gene Identification for Systemic Lupus Erythematosus Using Network Centrality Measures and Gene Ontology." PloS one 8.12 (2013): e81766.
13. Izudheen, S., and Mathew, S. "Cancer gene identification using graph centrality." CURRENT SCIENCE 105.8 (2013): 1143.
14. Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. "Reverse engineering of regulatory networks in human B cells." Nature genetics 37.4 (2005): 382-390.
15. Jörnsten, R., Abenius, T., Kling, T., Schmidt, L., Johansson, E., Nordling, T. E., ... and Nelander, S. "Network modeling of the transcriptional effects of copy number aberrations in glioblastoma." Molecular systems biology 7.1 (2011): 486.
16. Emmert-Streib, F., de Matos Simoes, R., Mullan, P., Haibe-Kains, B., and Dehmer, M. "The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks." Frontiers in genetics 5 (2014).
17. Cordero, D., Sol, X., Crous-Bou, M., Sanz-Pamplona, R., Par, L., Guin, E., ... and Moreno, V. "Large differences in global transcriptional regulatory programs of normal and tumor colon cells." BMC cancer 14.1 (2014): 708.
18. Knaack, Sara A., Alireza Fotuhi Siahpirani, and Sushmita Roy. "A pan-cancer modular regulatory network analysis to identify common and cancer-specific network components." Cancer informatics 13.Suppl 5 (2014): 69.
19. Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC bioinformatics, 7(Suppl 1), S7.
20. Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., . . . and Gardner, T. S. (2007). Large-scale mapping and validation of Escherichia

coli transcriptional regulation from a compendium of expression profiles. PLoS biology, 5(1), e8.
21. Meyer, P. E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. EURASIP journal on bioinformatics and systems biology, 2007.
22. Langfelder, P., and Horvath, S.. "WGCNA: an R package for weighted correlation network analysis." BMC bioinformatics 9.1 (2008): 559.
23. Altay, G., and Emmert-Streib, F. "Inferring the conservative causal core of gene regulatory networks." BMC Systems Biology 4.1 (2010): 132.
24. de Matos Simoes, R., and Emmert-Streib, F. "Bagging statistical network inference from large-scale gene expression data." PLoS One 7.3 (2012): e33624.
25. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. "Inferring regulatory networks from expression data using tree-based methods." PloS one 5.9 (2010): e12776.
26. Yip, K. Y., Alexander, R. P., Yan, K. K., and Gerstein, M. "Improved reconstruction of in silico gene regulatory networks by integrating knockout and perturbation data." PloS one 5.1 (2010): e8121-e8121.
27. Zhang, X., Zhao, X. M., He, K., Lu, L., Cao, Y., ... and Chen, L. "Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information." Bioinformatics 28.1 (2012): 98-104.
28. Margolin, A. A., and Califano, A. "Theory and limitations of genetic network inference from microarray data." Annals of the New York Academy of Sciences 1115.1 (2007): 51-72.
29. Chalancon, G., Ravarani, C. N., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., and Babu, M. M. "Interplay between gene expression noise and regulatory network architecture." Trends in genetics 28.5 (2012): 221-232.
30. Altay, G., and Emmert-Streib, F. "Revealing differences in gene network inference algorithms on the network level by ensemble methods." Bioinformatics 26.14 (2010): 1738-1744.
31. Schaffter, T., Marbach, D., and Floreano, D. "GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods." Bioinformatics 27.16 (2011): 2263-2270.
32. de Matos Simoes, R., Dehmer, M., and Emmert-Streib, F. "B-cell lymphoma gene regulatory networks: biological consistency among inference methods." Frontiers in genetics 4 (2013).
33. Marbach, Daniel, et al. "Revealing strengths and weaknesses of methods for gene network inference." Proceedings of the National Academy of Sciences 107.14 (2010): 6286-6291.
34. Marbach, D., Costello, J. C., Kffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., ... and DREAM5 Consortium. (2012). Wisdom of crowds for robust gene network inference. Nature methods, 9(8), 796-804.
35. Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, ., Anwar, N., ... and Sander, C. "Pathway Commons, a web resource for biological pathway data." Nucleic acids research 39.suppl 1 (2011): D685-D690.
36. Albert, R. "Scale-free networks in cell biology." Journal of cell science 118.21 (2005): 4947-4957.
37. Meyer, P. E., Lafitte, F., and Bontempi, G. (2008). minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information. BMC bioinformatics, 9(1), 461.
38. MATLAB, version 8.5 (R2015a), Natick, Massachusetts: The MathWorks Inc.
39. Welch, B. L. "The generalization of student's' problem when several different population variances are involved." Biometrika (1947): 28-35.

40. Freeman, L. C. "A set of measures of centrality based on betweenness." Sociometry (1977): 35-41.
41. Brandes, Ulrik. "A faster algorithm for betweenness centrality*." Journal of Mathematical Sociology 25.2 (2001): 163-177.
42. Gleich, D. F. Chapter 7 on MatlabBGL in Models and Algorithms for PageRank Sensitivity, Stanford University (2009).
43. Dancelli, L., Manisera, M., and Vezzoli, M. "On two classes of Weighted Rank Correlation measures deriving from the Spearmans ." Statistical Models for Data Analysis. Springer International Publishing, 2013. 107-114.
44. Tarsitano, A. "Comparing the effectiveness of rank correlation statistics." P: Dip. di Economia e Statistica, Univ. della Calabria (2009).
45. Pozzi, F., Di Matteo, T., and Aste, T. "Exponential smoothing weighted correlations." The European Physical Journal B 85.6 (2012): 1-21.
46. Van den Bulcke, T., Van Leemput, K., Naudts, B., van Remortel, P., Ma, H., Verschoren, A., ... and Marchal, K. "SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms." BMC bioinformatics 7.1 (2006): 43.

An application of the first exit time theory in some European populations

Konstantinos N. Zafeiris and Christos H. Skiadas[2]

[1] Laboratory of Anthropology. Department of History and Ethnology. Democritus University of Thrace.
(E-mail: kzafiris@he.duth.gr)
[2] Technical University of Crete, University Campus, 73100 Chania, Crete, Greece
(E-mail: skiadas@cmsim.net)

Abstract: The scope of this paper is to study the mortality and health trends in some European populations with the appliance of the first exit time theory. Five countries of Europe were chosen for that reason as they have recently experienced significant economic problems: Greece, Italy, Spain, Portugal and Ireland. Based on a life table analysis, several mortality and health indicators were estimated. Overall, mortality declined in all the populations studied and health status improved. However, in 2012, significant differences still existed among them.
Keywords: Life table, Health State Indicators, first exit time theory.

1 Introduction

According to the first exit time theory (Jansen and Skiadas [16], Skiadas and Skiadas [29]), the health state and the vitality of an organism can be estimated from the available population-death data sets. It is based on the idea that the health of an individual follows a stochastic process and death comes when health falls below a limit, a barrier as it is called in the first exit time theory. The health state is the opposite of mortality and the health state function is the opposite of the mortality function.

Five European populations were chosen in order for the first exit time theory to be applied: Greece (EL), Italy (IT), Spain (ES), Portugal (PT) and Ireland (IE). But the economic crisis they faced after the American one started in 2007 (FCIC [11]) is unquestionable as can be judged from the EUROSTAT data (Tables 1 and 2).

It has to be noted that, with the exception of Ireland, the economic levels of the countries studied were below the European Union (EU) and Euro area (EA) all the time. Economic growth was observed in all of them from 2000 until 2008 (2007 for IE), especially in Ireland, Greece and Spain. After 2008 (2007 for IE) a significant decrease in GDP was observed. Greece faced the biggest problems:

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

GDP per capita decreased by 19% from 2008 until 2012. Ireland comes next, (-15,9%), though it kept its position above EU and EA levels, followed by Spain, Portugal and Italy (Table 1).

Table 1. Gross domestic Product (GDP). Current prices, Euro per capita.

Geographic area	2000	Maximum GDP per capita before crisis			2012	% change with 2008 or 2007 (IE)
		Year achieved	GDP per capita	% change with 2000		
EU	19600	2008	25900	32,1	26500	2,3
EA	22100	2008	29000	31,2	29300	1,0
IE	28300	2007	44700	58,0	37600	-15,9
EL	13100	2008	21600	64,9	17500	-19,0
ES	16100	2008	24300	50,9	22600	-7,0
IT	21800	2008	27600	26,6	27000	-2,2
PT	12500	2008	16900	35,2	16100	-4,7

EU: European Union; EA: Euro-area, excluding Lithuania. Source: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_gdp_c&lang=en (last update 8/12/14)

Table 2. Unemployment rates in the years contained in Table 1. Annual averages. Both genders. Not seasonally adjusted data.

Geographic area	2000	Unemployment rates			2012	% change with 2008 or 2007 (IE)
		Year achieved	rates	% change with 2000		
EU	8,9	2008	7,0	-21,3	10,5	50,0
EA	8,7	2008	7,6	-12,6	11,3	48,7
IE	4,2	2007	4,7	11,9	14,7	129,7
EL	11,2	2008	7,8	-30,4	24,5	214,1
ES	11,9	2008	11,3	-5,0	24,8	119,5
IT	10,0	2008	6,7	-33,0	10,7	59,7
PT	5,1	2008	8,7	70,6	15,8	81,6

EU: European Union; EA: Euro-area, excluding Lithuania. Source: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_a&lang=en (last update 7/1/15).

The economic crisis was addressed politically (see for example the European Stability Mechanism, http://www.esm.europa.eu/) and several austerity and reform programs were carried out either with the involvement of the Troika

(Greece, Ireland and Portugal; IMF [13], [14], [15]) or not (Spain [1] and Italy, [30]). These developments left their marks clearly on the socio-economic environment of these countries. Unemployment rates increased 214% in Greece, 130% in Ireland, 120% in Spain and by a lesser degree in the other countries (between 2008 and 2012). In 2012 they were about 25% in Greece and Spain and lower – though still higher than the EU and EA average – in the other countries. Also, the people at risk of poverty increased in all of them between 2009 and 2012 etc. (ELSTAT [10]; see also Thomson et al. [31]).

The main question addressed in this paper is the examination of mortality and health trends of the populations of these countries before and during the ongoing economic crisis.

2 Data and methods

Following the first exit time theory the so called Ski-1995 6 parameters model was used in which the death probability density function g(x) is given by:

$$g(x) = k(x)^{-\frac{3}{2}} e^{-\frac{H_x^2}{2x}} \quad (1)$$

where H_x is the health state in age x

$$H_x = a_1 + ax^4 - b\sqrt{x} + lx^2 - cx^3 \quad (2)$$

and *k*, a_1, *a*, *b*, *l* and *c* are parameters which should be estimated. Skiadas and Skiadas ([24], [25], [26], [27], [28]) and especially Skiadas [26] discuss in detail the health state theory and Skiadas [26] has also produced an Excel sheet for the calculations. In its original form, death and population data are introduced in the sheet and after the normalization of the g(x) distribution in order to sum 1 and for different countries to be comparable, all the parameters are automatically calculated with a non linear regression model according to Skiadas [23]. Then the Excel solver was used for the final estimation of the parameters, in a way that the sum of squared errors of the fit to be minimized. The fit was excellent for most of the ages of the human life span - though some minor deviations were observed because of the mortality excess for ages 15 to 30. Thus, the original program was modified in order to further contain two seriatim Gompertz functions. Afterwards, Excel solver was used once again and in that way the coefficient of determination of the fitted mortality curve was perfect (R^2=1 and in few cases R^2=0.999).

In fact during the analysis several models of curve fitting were applied in the data (Gompertz, Weibull and some simplified exit time models) but none of them approached the perfect fit obtained with the ski-1995 6 parameters model. For example, it was found that in the older ages the Gompertz model had an R^2 that was quite lower than the model used here because it is not as flexible as the

Ski-1995 one and for some of the indicators it gave false results. In our opinion the Ski-1995 6 parameters model has three important properties. The first one is that gives a quite clear picture of the health status of the populations and it permits the calculation of several health indicators that will be discussed later. The second one is that it can be used in parallel with the health state analysis in order for several life table indicators to be calculated. The third property is that it can be seen as a sensitive and reliable data fitting method in life table analysis which produces perfect or almost perfect fits. In this paper, because of the maximum coefficient of determination that was achieved only the results from the fitted data will be discussed.

The death distribution curve is a bimodal one. In the past, the first mode corresponded to very high infant mortality and the second one to old age mortality (see for example Lexis [19]). In the modern era however, mortality is concentrated at older ages in most countries (Canudas-Romo [5]) and this triggered the development of various hypotheses in order to model and explain old-age mortality. For example, the mortality shifting hypothesis suggests that the old age mortality pattern retains its shape as it shifts towards higher ages through time (Bongaarts and Feeney [2] [3], Bongaarts [4]). In the rectangularization hypothesis, the human survival curve becomes more rectangular as mortality levels decrease (see Wilmoth and Horiuchi [32]; Robine [22]; see also Cheung at al. [8]). Closely linked with that notion, is the concept of compression of mortality which is based on the simple premise that mortality is being compressed when a given proportion of deaths takes place in a shorter interval as before (Kannisto [18]). Several indicators have been proposed and used in the literature in order to evaluate these suggestions. For example, Kannisto [18] has proposed 4 methods for measuring the compression of mortality, among them the well known C-family indicators, C10, C25 and C50, which are defined as the narrowest age intervals in which 10, 25 and 50% of all deaths occur. Additionally, Wilmoth and Horiuchi [32], used the inter-quartile range of life table ages at death in order to study the rectangularization of survival curves, discussing at the same time several other indicators (see also Robine [22]; Paccaud [21]; Egidi and Spizzichino [9]).

These indicators were used in parallel with several measures of longevity: the mean length of life, which corresponds to life expectancy at birth, the median age at death and the modal age at death (see for example Canudas-Romo [6]; Horiuchi et al. [12]), along with the life expectancy at a certain but advanced age like that of 65 (e65) plus 65 years (e65+65) etc. It must be noted that life expectancy at birth and the median age at death as measures of longevity are affected by the infant, child and reproductive-adult mortality and also that e65 tends to underestimate old age mortality shifts. On the contrary the modal age at death is considered to be a better indicator of mortality shifts and longevity, as it is determined only by old-age mortality, taking also into consideration that the current extension of human longevity in low mortality countries results from the improvements in old age survival (see Horiuchi et al. [12]). In this paper, the life

expectancy at birth and age 65 (expressed as e65+65) were calculated from the fitted data from the life table. The modal (M) and the median age at death were calculated from the death density function distribution (g(x)) along with the g(x) at modal age with the appliance of the formula (1) with decimal precision.

Additionally 4 fixed age points were estimated from the same distribution: the age at the 10^{th} and 90^{th} percentiles and the inter-percentile range, and the analogous 25^{th} and 75^{th} quartiles and inter-quartile range.

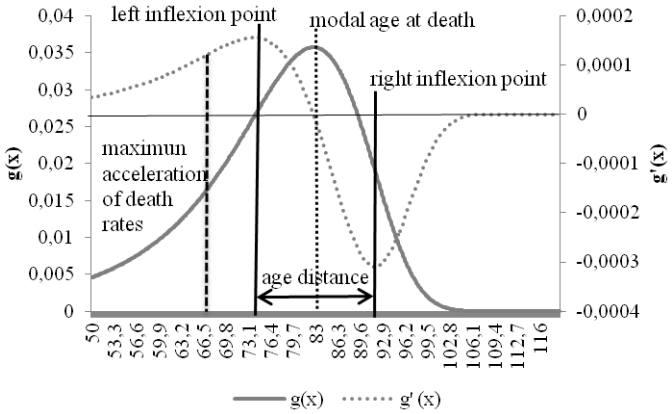


Fig.1. The left and right inflexion points, modal age at death, and death rate speed (g'(x)) Spain, males, 2000. x axis: age

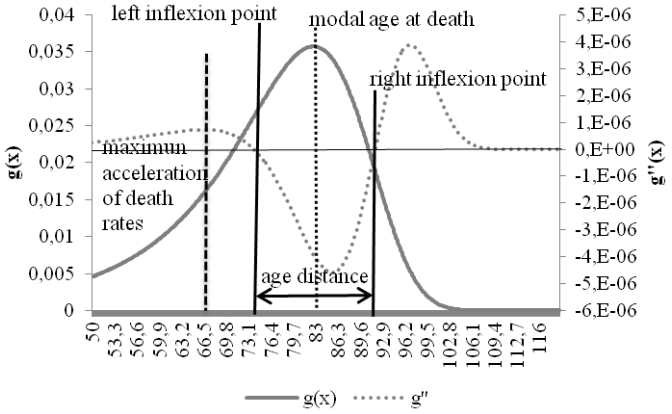


Fig.2. The left and right inflexion points, modal age at death, and acceleration of death rates (g''(x)). Spain, males, 2000. x axis: age

Also, three new indicators were proposed (Figures 1 and 2). The left inflexion point (LIP) of the death density distribution corresponds to the age at which the maximum speed (g') of death rates is observed; i.e. it corresponds to the age at which the human health is rapidly burdened. After that age the death rate speed is gradually reduced and tends to be 0 at the modal age of death while the death density function reaches its peak. The left inflexion point can be easily estimated from the g(x) distribution of ages of one or more decimal precision with the appliance of formula (1) as the age at which the maximum difference of two subsequent g(x) values is attained. Before the left inflexion point the

maximum acceleration of death rates is observed. The acceleration is calculated as the difference of two subsequent death rate speed values. The acceleration is positive before the age at maximum speed, but after its maximum it declines in order to be 0 in the LIP. In other words death rates at the older ages accelerate up to a time point and in that way the loss of vitality accelerates too. That result in the elevation of the death rate speed, whose maximum is attained at the left inflexion point. After the LIP acceleration becomes negative as the relative curve has an inverted "bath-tub" form (Figure 2) and death rate speed decreases and reaches a minimum which corresponds to the right inflexion point (RIP). At the same time g(x) values decrease. After the RIP acceleration becomes positive and the death rate speed increases up to some age and then remains stable. The age distance between the left and the right inflexion points (RIP-LIP) will be used as an estimation of the width of mortality curve at older ages.

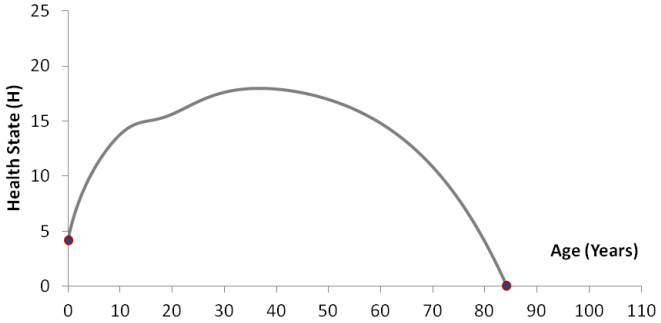


Figure 3. The health state function of the population up to the zero health age. Spain. Males, 2000.

The health indicators are based on formula (2) by which the health state function distribution (Hx) of the population is calculated. This distribution increases up to an age which corresponds to the maximum health and then decreases at the older ages to reach the zero health point (Figure 3). It is seen that in the increasing phase of human health a disturbance is observed after the age of 15 and later on until about the 30s, which is connected with the excess of mortality at those ages which was discussed before. Several measures can be taken on this curve, in order for different populations to be compared. The Total Health State (THS) of the population can be estimated as the sum of H(x) values before the zero health age point and the Hmax will correspond to the maximum value of health that was attained during the life course. The age at zero health will also be discussed along with the expected healthy age, which represents the age close to the maximum health state. Also the Total Life Healthy Life Expectancy at birth (HLEB (Total)) which represents the expected life free of all disabilities will be discussed (for the calculations see Skiadas & Skiadas [24], [26]).

Finally the deterioration function will be used, which represents a measure of the curvature of the health state function and can be used to find how fast the health state function changes (see Skiadas and Skiadas [26], [27] for the calculations). As seen in Figure 4, the deterioration function is high in the first years of life and reaches its minimum when the adult development has been

completed. After its minimum point the "vitality" starts to disintegrate and it increases to its maximum level, to the time point in which the deterioration of the physical repair mechanism becomes maximum. After the age of maximum, the deterioration decreases to asymptotically low levels, what is called mortality leveling-off [26].

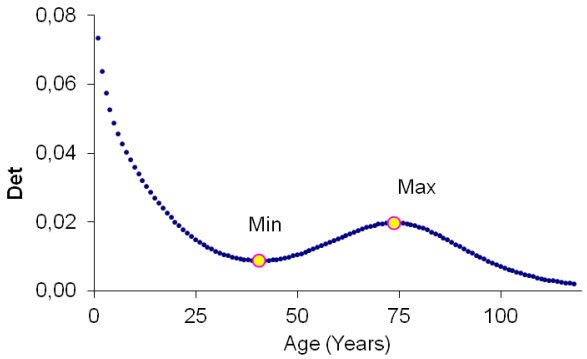


Figure 4. The deterioration function. Spain. Males, 2000.

Data come from the human mortality database (http://www.mortality.org/) for Portugal and Spain for the years 2000-2012 and for Italy and Ireland for the year 2000-2009. Also the Eurostat database was used (http://ec.europa.eu/eurostat/data/database) for Greece (2001-2012), Italy (2012) and Ireland (2010-2012)

3 Results

3.1 Mortality indicators

The analysis of mortality indicators and their temporal trends indicate that despite the general decrease of mortality observed in the 5 countries, significant differences still existed among them in 2012 (Figure 5). In the male population, life expectancy at birth (LEB) trends and differentials indicate that Portugal, the poorest country of the Euro area (see Table 1), had the highest mortality levels all the time. However, LEB increased linearly and quite rapidly there and the country finally converged with Greece, as LEB increased by more than 4 years between 2000 and 2012. Ireland ranked in the second position of higher mortality countries at the beginning of the century but because of a sharp and linear increase trend of LEB in the first years followed by a stepwise one later, it attained the 3rd lowest mortality in 2012, gaining by that time more than 4.5 years in LEB. During the last years of the study, after the spread of the economic crisis in the Euro area, any improvements in LEB levels were halted in Ireland. The same happened in Greece after 2010 where LEB decreased slightly in 2012.

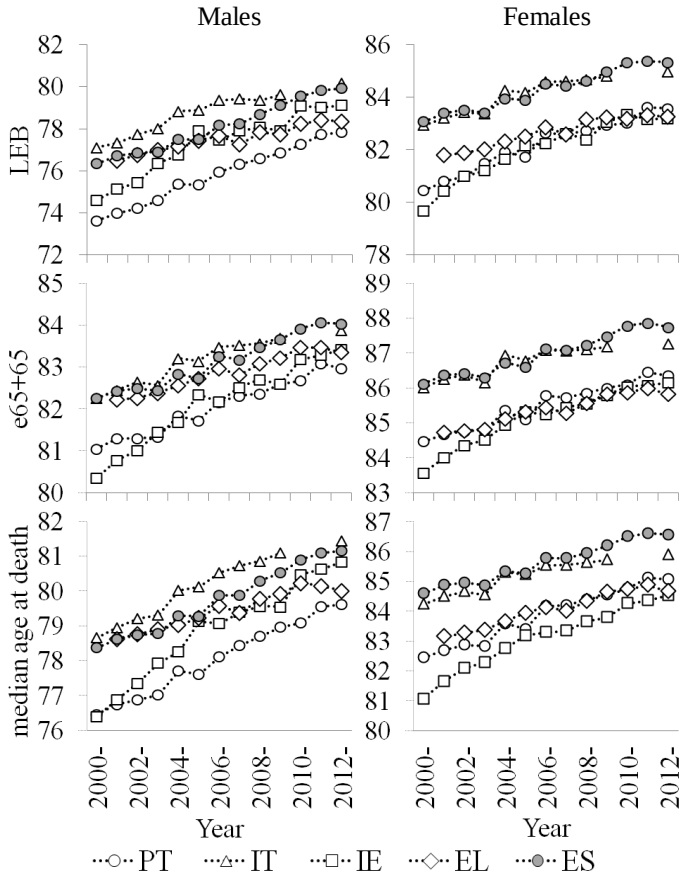


Fig. 5. Life expectancy at birth, e65+65 and median age at death.

The gains between 2001 and 2012 were less than two years there, being the smaller ones among the countries studied. Actually, Greece and Spain had many similarities as countries of lower mortality until 2006, when Ireland also converged with them. After 2006, a divergence trend is observed. In Spain the LEB improvements were more rapid and the country converged with Italy, which had the lowest mortality of all the others studied. The total gains in LEB between 2000 and 2012 were about 3 years in Italy and 3.6 in Spain. In the female population two groups of countries are observed. Spain and Italy were of rather lower mortality than the others all the time (gains in LEB between 2000 and 2012 about 2 and 3 years respectively). Greece - originally in the middle zone of mortality ranking - because of its moderate mortality transition (LEB increase 2001-2012: about 1.5 years) finally converged with Portugal and Ireland where - despite their worse original positions - LEB improvements were more rapid (more than 3 and 3.5 years respectively). In all the countries during

the last years of the study the previous mortality trends were halted and mortality changes became either minimal or negative.

If life expectancy at 65+65 is taken into consideration as a life span indicator, some differences are observed in comparison to the temporal patterns described by the LEB. The most important one is that in males a tighter grouping of Greece, Italy and Spain, as countries of higher longevity is observed for the first years of the study. After 2006, Italy and Spain started to differentiate and they gradually grouped apart from Greece. Portugal and Ireland and especially the latter, followed different courses and tended to converge with Greece, even though Portugal kept a more distant position in 2010 and 2012. Similarly, in females the most diverse country on the eve of the 21st century was Ireland, and Greece and Portugal had many similarities. Soon enough these countries converged. Spain overtook Italy in 2012, with which it previously clustered tightly. During the last years of the study, a general trend of halting of the improvements or even a minor decrease was observed in all countries and in both genders. A broadly similar picture emerges from the temporal trends of the median age at death in females, while in males the observed pattern is more similar to that of the life expectancy at birth.

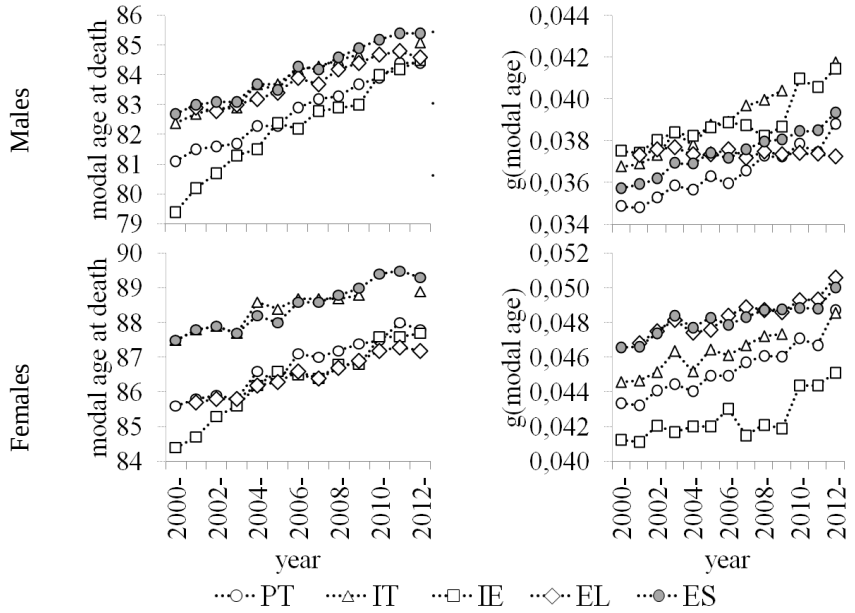


Fig. 6. Modal age at death (M) and death density at modal age at death.

A marginal decline of longevity is also obvious in 2012, as the modal age at death (M, Figure 6) decreased slightly in Greece (for both genders), Italy and Portugal (females). Besides that and with some minor exceptions, males' M was progressively sliding towards older ages since the beginning of the 20th century.

Portugal, which initially was placed in an intermediate position, experienced an almost linear improvement, followed by Ireland which was placed originally quite apart, both converging with Greece in 2012. During that course Ireland had the greater gains in the modal age at death (+5.1 years), followed by Portugal (+3.3), Italy and Spain (+2.7 years) and Greece (+1.7 years between 2001 and 2012). As could be expected then the lower mortality countries at the beginning of the century had smaller overall gains in M than those of the higher mortality.

The only exception is Greece where the observed improvements were more moderate. An additional difference is found there concerning the pattern of old age mortality, as the temporal fluctuations observed in the density of deaths at modal age [g(modal age)] were minor; i.e. the maximum of the death probability density function distribution did not change much over time (0.037-0.038) while at the same time the age at which this maximum is achieved was shifting towards older ages. In all the other populations, even if for some years the Greek pattern may be followed, the general trend is of the elevation of the peak of the death probability density function distribution in parallel with the increase of modal age at death.

In females, Italy and Spain, where modal age at death follows a stepwise but finally increasing trend, are clearly distinguishable. Similar increasing trends are found in the other countries. The gains they had in 2012 are lower than those of males: +3.3 years in Ireland, +2.2 in Portugal, about 1.5 in Italy and Greece and +1.8 in Spain. During the last years of the study any improvements became minimal or even reversed. Despite their grouping as the countries with the highest longevity, Italy and Spain had significant differences in the height of the peak of the death probability density distribution. Spain resembles Greece in both levels and temporal trends. In other words, females from Greece attain similar values in the peak of the death density distribution but, at the same time, at younger ages than Spain as M reveals. Italy attains the similar modal age at death with Spain in the majority of the years studied but the relevant death distribution peak is always lower.

Modal age at death trends in Portugal, in their turn, follow up to a time point those of Greece, and afterwards M becomes higher but in any case lower than in Spain and Italy. A similar stepwise increase is observed in the g(modal age), but its levels are always lower than the other countries discussed so far. Finally, Ireland, where originally M was low, converged quite randomly with Greece and Portugal, and after 2008 overtook Greece. The g(modal age) there even if it is in general improving order remains quite lower than all the other countries.

Conclusively, despite the general improvement of the modal age at death up to a time point which is observed in all countries, significant differences are found among them not only concerning the age at which the death distribution retains its maximum but also the mortality levels at that age.

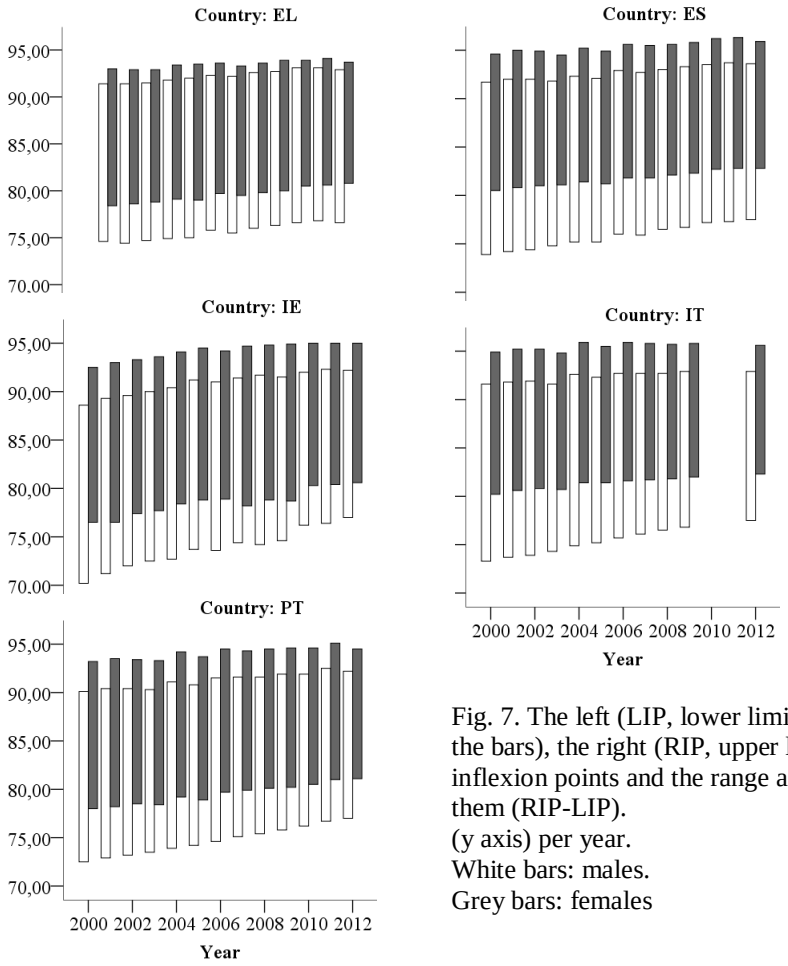


Fig. 7. The left (LIP, lower limit of the bars), the right (RIP, upper limit) inflexion points and the range among them (RIP-LIP).
(y axis) per year.
White bars: males.
Grey bars: females

The left and right inflexion points are quite informative for the shape of the mortality curve in the older ages (Figure 7). Because in females mortality transition has moved ahead considerably more than in males, the left and right inflexion points are located at higher age and their age distance is shorter, which means that the shape of the death density distribution at the older ages is narrower in them than in males. Additionally, both the LIP and RIP, even with some fluctuations, tend to slide over time towards higher ages in all populations in both genders. This happens at a slower rate concerning the right inflexion point; i.e. the age at which the minimum death rate speed is observed is increasing very slowly. In both genders the most important increase in the RIP is found in Ireland, which had the highest mortality in 2000, followed by Portugal, Spain and to a lesser degree Greece and Italy. However, despite the

differential overall improvements, in 2012 still significant differences existed among the countries. In males, Italy is clearly differentiated from the group formed by Spain and Greece, where the age at right inflexion point is lower. Beneath this group Portugal and Ireland are located. In females a greater variability is observed, though Italy and Spain are clustered together and apart from the other countries. Similarly, in the originally higher mortality countries the left inflexion point is moving with time towards older age faster than the others and the males have greater gains than females during that course. As a result males tended to converge in 2012, while in females the same grouping as the right inflexion point is observed. Within the two genders, as a result of the differential trends in the age location of the two points the relevant distance between them tends to decrease over time. The only exception is found in the males of Greece where the age distance fluctuated over time between the 16.3 and 17 years, even though a minor decrease was observed during the last years of the study. It seems then that the death density function distribution becomes narrower over time and in females this is even more pronounced.

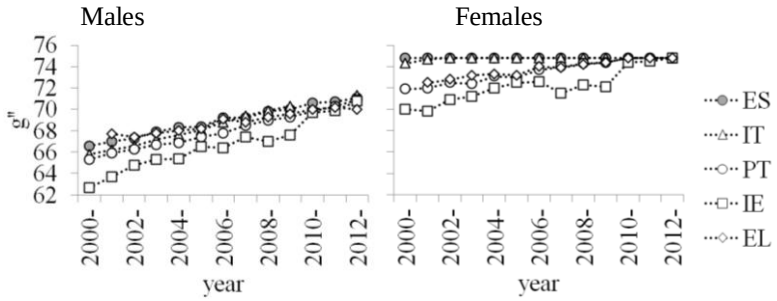


Fig 8. Age at the first maximum acceleration of death rate (g'').

These pattern shifts are accompanied by changes in the acceleration of the death rates (Figure 8), which have their age maximum (g'') before the left inflexion point and in fact signal the acceleration of the health burden processes and the maximization of the death rate speed some years later. All the populations by the end of the study, converged to similar values at the age where the g'' maximum is achieved. In females, in Spain and Portugal that age remained quite stable between 2000 and 2012 and the other countries gradually converged as longevity was increasing. A hypothesis for further research can be set up here: as mortality transition moves forwards the age at maximum g'' will be stabilized? Needless to say of course, additional evidence is needed in order for this hypothesis to be further examined and evaluated.

If the quartiles and percentiles of the whole mortality curve will be taken into consideration a similar picture described above concerning the right and left inflexion points is obtained and only some important findings will be discussed here (Figure. 9). In both genders the age at 10^{th} percentile (p10) and at the 25^{th} quartile (q25) are increasing faster than the relevant ages at the 90^{th} percentile

(p90) and 75th quartile (q25) this results in a limited shortening of inter-quartile and inter-percentile range, i.e. a compression of mortality is observed if especially the inter-quartile range is taken into consideration. An exception is found in males of Greece, where the age at 10th percentile does not change much through time but the 90th quartile increases somewhat and as a result the inter-percentile range increases between 2001 and 2012. A second exception is found in males of Ireland with the almost parallel increase of p10 and p90 results a limited decrease of the inter-percentile range. The ages at all quartiles and percentiles are always higher in females and the relevant range shorter.

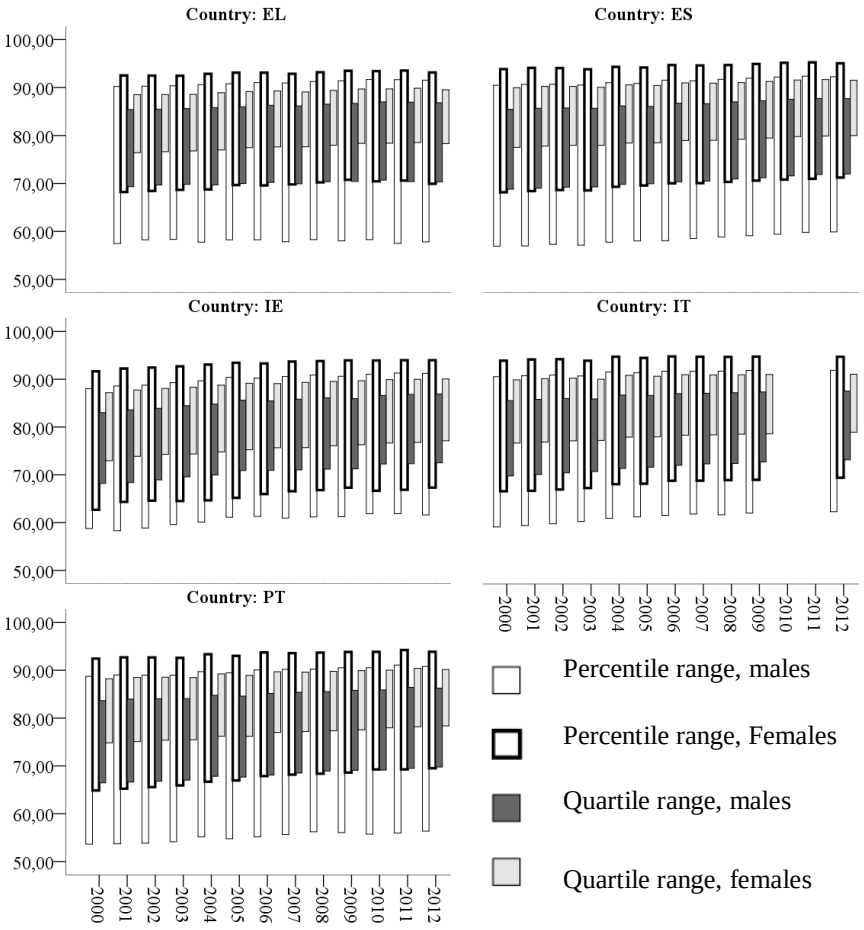


Fig 9. The inter-percentile and inter quartile range.

.

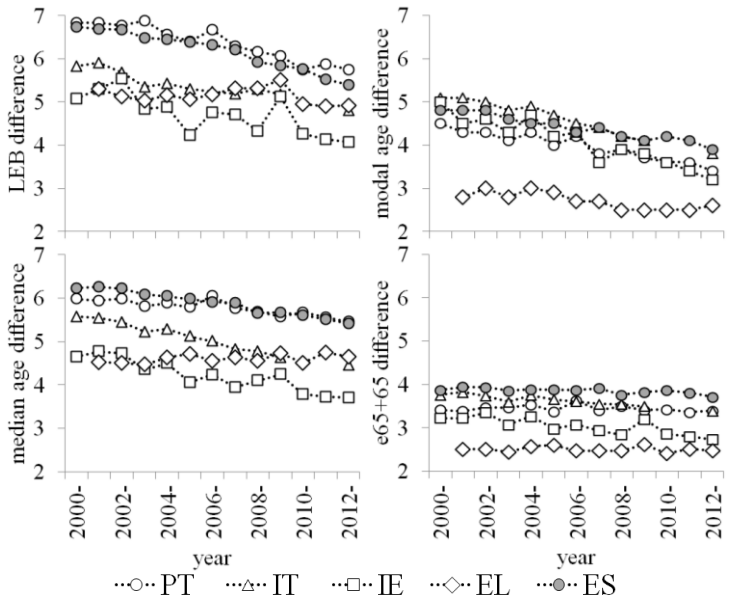


Fig. 10. Differences (females-males) between the two genders in life expectancy at birth and e65+65 and median and modal age at death.

Between the two genders the gap in LEB, modal and median age at death mainly tends to decrease over time (Figure 10). The greater differences in LEB among the two genders are found in the Iberian Peninsula. Italy and Greece follow a stepwise but decreasing course and for most of the time are placed above Ireland. However, Greece has the smallest differences concerning the modal age at death and the e65+65 indicator, with minor variations through time, as is the case with the median age at death. In fact the gender gap on the e65+65 indicator, with the exception of Ireland and despite some temporal variations, does not change much through time.

3.2 Health indicators

The age at minimum deterioration, i.e. of the onset of "vitality" disintegration in adulthood is located somewhere around 40 years of life in the male population, though some variations are observed among the countries (Figure 11). Clearer trends are observed at the age of maximum deterioration, which in general and despite some variations is in increasing order in all of them. Ireland, Portugal and Greece had the greater increase (4.8, 1.7 and 1.6 years respectively) during that course. It seems then that the health of the older people was gradually improved in all the countries studied; however it must be noted that Ireland was originally placed quite apart from the other countries but gradually converged partially with them in 2012. The age distance between the ages of minimum and maximum deterioration tends to enlarge slightly by time and the more important

improvements are observed in Ireland. In 2012 this distance is between 33 and 34 years. In females, the temporal trend pattern is quite similar to that of males. However, the ages of minimum and maximum deterioration are lower than in males. In contrast, females have consistently better health than males, as revealed by the temporal trends of the Total Health State Indicator (Figure 12).

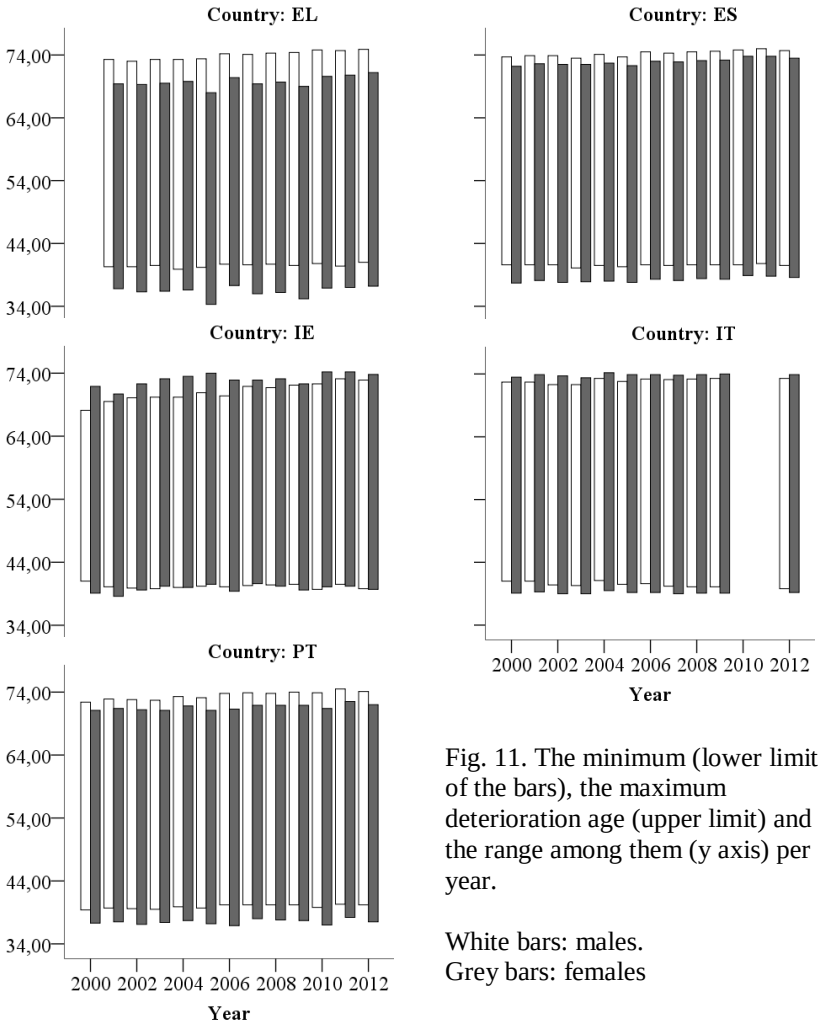


Fig. 11. The minimum (lower limit of the bars), the maximum deterioration age (upper limit) and the range among them (y axis) per year.

White bars: males.
Grey bars: females

Generally speaking, the total health of all populations and both genders becomes better as mortality declines over time. In males, the first years of the 21st century Greece, Spain and Italy, due to their lower mortality, were closely grouped together as the countries of quite higher total health than that of Ireland and Portugal.

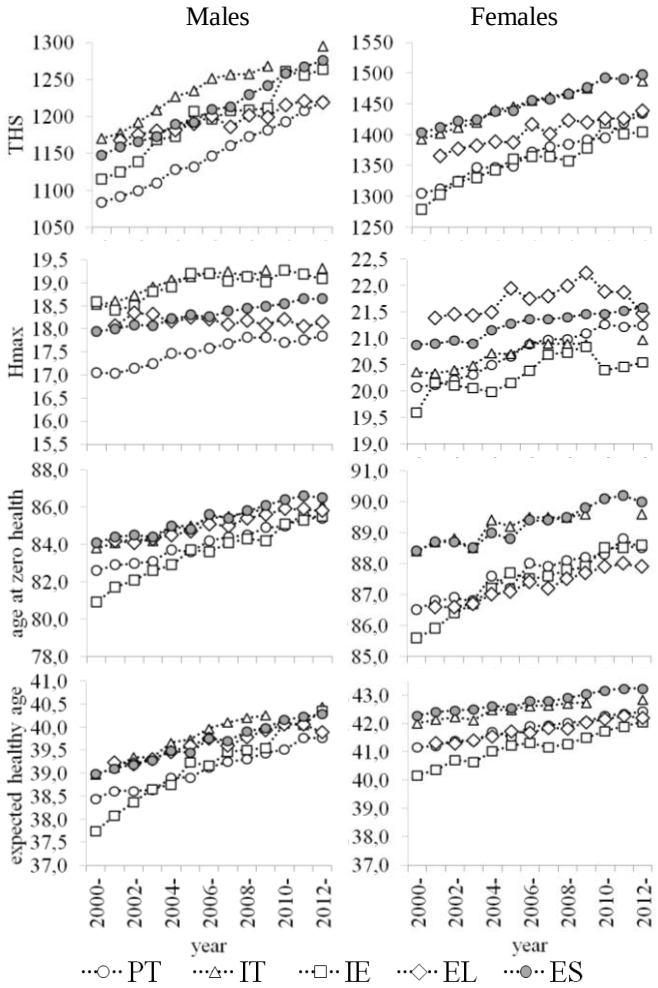


Fig. 12. Health states indicators in the countries studied.

Afterwards because of the differential temporal trends under held by all of them, two groups were formed. Ireland – which had the greatest total health gains even if in the last years of the study any improvements were halted - with Spain and Italy were clearly distinguishable from Portugal and Greece, the group of lower health populations. Meanwhile, Portugal had greater gains than Greece, where after 2010 there were no improvements in the health status at all. In females, Spain and Italy had the best health all the time. Greece was originally placed in an intermediate position and the health improvements observed there were not as rapid as in Portugal and Ireland. In Greece after 2007 and in Spain after 2010 the THS remained quite stable. Portugal converged with Greece in 2012. In Ireland THS decreased in the last two years of the study. It has to be noted that the analysis of health status trends in national populations conceals

the peripheral differences that may exist within the sub-populations of a country, as found in a relevant analysis for Greece (Zafeiris and Skiadas, [33]), a fact that must be taken into consideration for further research.

In any case the form of the health state distribution, besides its total 'length" (THS), also has other differences among the populations studied (Figure 12). First of all, despite their original differences in the THS, the males from Italy and Ireland achieved the highest health status (H_{max}) of all the others. Greece was very close to them in 2003 but later on no significant improvements were observed and in 2012 it had the 4th worst position among the five countries studied. In Spain H_{max} increased linearly attaining in 2012 an intermediate position. Males from Portugal had the worst maximum health status all the time. Besides the overall improvements observed over time there, after 2008 these were halted and H_{max} fluctuated close but beneath the Greek levels. H_{max} temporal trends were quite different in the female population of the countries studied. Following a rather variable but generally increasing course, the best maximum health was found in Greece until 2011. Afterwards H_{max} decreased and Spain, the second best country with maximum health, overtook Greece. Italy had the 3rd highest maximum health the first years of the study followed by Portugal. However, the latter attained 3rd position after 2008 and tended to converge with Spain and Greece. Finally women from Ireland had, with few exceptions, the worst maximum health of all the others and by the end of the study they had diverged a lot. Between the two genders, females all the time have higher values of H_{max}.

The second important difference found in the health status distributions is of the expected healthy age of the populations studied (Figure 12). Unlike the H_{max} distribution, originally males of Italy, Greece and Spain constitute a group of populations with the highest expected healthy age, followed by Portugal and Ireland. After 2008 the expected healthy age stabilizes in all populations around 40 years. Ireland had the most noticeable gains of all the other countries, as from the worst and most distant position it had in 2000 it fully converged with Spain and Italy in 2012. The other countries had more moderate gains. In females, Italy and Spain are clearly differentiated as the countries with the best expected healthy age, were an almost linear but very slow improvement is observed, which was halted during the last years of the study to values around the 42-43 years. Greece and Portugal had an intermediate position and after 2009 expected healthy age was very close to 42.3 years in both of them. Ireland had the worst expected healthy age all the time but because of the observed improvements it converged with Greece and Portugal in 2012. Finally females have higher expected healthy age than males.

A similar grouping of the countries to the one based on the expected healthy age is found for the age at zero health, which, originally comes later in life in the male population of Greece, Italy and Spain, albeit these countries had smaller improvements than the others (Figure 12). Any improvements were either halted

or reversed slightly in the last years of the study there. Once again the greatest increase was observed in Ireland. By the end all the countries converge somewhere between 85 and 86 years. In females, the grouping of Spain and Italy remains, though Greece is positioned far away, following a stepwise increase of age at zero health. Ireland soon converged with Portugal and Greece. After 2006 Greece tends to be placed more distantly and in the worst positions. As in males, in all countries age at zero health was either halted or reversed slightly in the last years of the study.

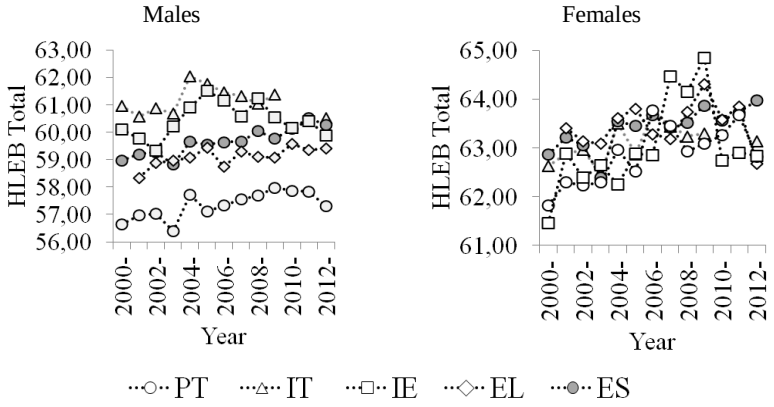


Fig. 13. Healthy life expectancy (Total).

Finally, the temporal rends of the healthy life expectancy of males are not as clear as those described above, as great fluctuations are observed over time (Figure 13). In Italy HLEB (Total), with some minor exceptions, was very close to 61 years until 2004 when it reached a maximum of 62.5 years. Afterwards HLEB (total) shortens. In Spain, it increases slightly to 60 years in 2008 and then fluctuates around there. In Greece, which is the country with the second worst HLEB (Total) after Portugal, it was somewhat shorter. Ireland, despite having for the first years of the study the worst LEB, at the same time had the second longest healthy life. After 2008 it converged with Italy and Spain. Females of Greece while originally having the healthiest length of life gradually converged with the other countries and in 2012 HLEB (Total) shortened significantly. In Italy female healthy life increased slightly until 2004 and then fluctuated above 63 years. During the last years of the study it decreased. In Ireland and Portugal a more complicated scheme persists through time, but in 2012 they converged with Italy and Greece to lower levels than in the past.

Conclusions

Based on the mortality indicators used in this paper an obvious decline of mortality is observed in all the countries studied. As the mean and median and e65+65 length of life revealed, the lower mortality countries had experienced a greater increase of longevity than the others. The death density distribution had

significant differences among the two genders and the countries. Modal age at death was sliding towards older ages up to a time point but at the same time it was quite variable among the countries, a fact that refers also to the mortality levels attained at that age. Also over time, as the left and right inflexion points revealed, the death distribution became narrower in both genders and this pattern was accompanied by changes in the acceleration of death rates. Both the left and right inflexion points and the analysis of inter-quartile and inter-percentile range revealed that a trend towards a compression of mortality was observed through time. However, sometime after the eve of the economic crisis, any improvements were halted if not slightly reversed, as for example found in the modal age of the female population of Greece, Italy and Spain. In fact, Greece gained a special place in the analysis of both mortality and health, because of its special patterns and its more moderate transition.

During the last years of the study the observed variability in mortality levels among the countries studied was significant, despite the overall improvements. Health indicators confirm this picture. Important differences were originally found in the maximum and minimum deterioration ages of human health and over time the populations tended to converge. The Total health was increasing at the same time, but after a time point in some countries the improvements were halted. The variability was also significant in the H_{max} and remained so by the end of the study. The expected healthy age increased and gradually stabilized, as happened with the age at zero health. Finally, the Healthy life expectancy followed a rather differentiated course as it exhibited several fluctuations irrespective of the temporal trends of life expectancy at birth. The differences found among the countries studied and the temporal trends of mortality and health observed may be attributed to the socio-economic differences which existed among and within them especially in the era of economic crisis and the analogous inequalities in health or to other factors like lifestyle, consumption of alcohol, smoking, environmental reasons etc. (for the socioeconomic and health inequalities see for example Mackenbach et al. [20], Karanikolos et al. [17]). A further study is connected with the isolation of these factors and the evaluation of their impact on the mortality and health differentials observed and also with the exploration of regional differences which existed within the countries studied.

References

1. BBC. Spain budget imposes further austerity measures. BBC. News business http://www.bbc.com/news/business-19733995
2. J. Bongaarts and G. Feeney. How long do we live? *Population and Development Review*, 28, 1, 13–29, 2002.
3. J. Bongaarts and G. Feeney. Estimating mean lifetime. *Proceedings of the National Academy of Sciences*, 100, 23, 13127–13133, 2003.
4. J. Bongaarts. Long-range trends in adult mortality: Models and projection methods. *Demography*, 42, 1, 23–49, 2005.
5. V. Canudas-Romo. The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19, 30, 1179-1204 2008.

6. V. Canudas-Romo. Three measures of longevity: time trends and record values. *Demography*, 47, 2, 299-312. 2010.
7. S. L. K. Cheung and J.-M. Robine. Increase in common longevity and the compression of mortality: the case of Japan. *Population Studies* 61, 1, 85–97. 2007.
8. S. L. K. Cheung, J. M. Robine, F. Paccaud and A, Marazzi. Dissecting the compression of mortality in Switzerland, 1876-2005. *Demographic Research*, 21, 19, 569-598, 2009.
9. V. Egidi and D. Spizzichino. Life expectancy: complex measures and the health related quality of life. *Italian Journal of Public Health*, 6, 5, 91-95, 2008.
10. ELSTAT. Hellenic Statistical Authority. Living conditions in Greece. 2 January 2015.
11. FCIC. Financial Crisis Inquiry Commission. Final report of the National commission on the causes of the financial and economic crisis in the United States. 2001.
12. S. Horiuchi, N. Ouellette, S. L. K. Cheung and J. M. Robine. Modal age at death: lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, 11, 37-69, 2013.
13. IMF. International Monetary fund. Letter of Intent, Memorandum of Economic and Financial Policies, and Technical Memorandum of Understanding; Greece: http://www.imf.org/external/np/loi/2012/grc/030912.pdf
14. IMF. International Monetary fund. Letter of Intent, Memorandum of Economic and Financial Policies, and Technical Memorandum of Understanding; Ireland: http://www.imf.org/external/np/loi/2012/irl/021012.pdf
15. IMF. International Monetary fund. Letter of Intent, Memorandum of Economic and Financial Policies, and Technical Memorandum of Understanding; Portugal: http://www.imf.org/external/np/loi/2014/prt/032814.pdf
16. J. Janssen and C. H. Skiadas. Dynamic modelling of life-table data, *Applied Stochastic Models and Data Analysis*, 11, 1, 35-49, 1995.
17. M. Karanikolos, P. Mladosvsky, J. Cylus, S. Thomson. S. Basu, D. StucklerJ. P. Meckenbach, M. McKee. Financial crisis, austerity and health in Europe. *The Lancet*, 381, 9874, 1323-1331, 2013.
18. V. Kannisto. Measuring the compression of Mortality. *Demographic Research*, 3, 6, 2000.
19. W. Lexis. Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de démographie internationale* 2, 447–460, 1878.
20. J. P. Mackenbach, I. Stirbu, A-J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinslau and A. E. Kunst. Socioeconomic Inequalities in Health in 22 European countries. *The New England Journal of Medicine*, 358, 23, 2468-2480, 2008.
21. F. Paccaud, C.S. Pinto , A. Marazzi and J. Mili. Age at death and rectangularization of the survival curve: trends in Switzerland 1969-1994. *Journal of Epidemiology and Community Health*, 52, 7, 412-5, 1998.
22. J-M. Robine. Redefining the stages of epidemiological transition by a study of dispersion of life spans: the case of France. *Population: an English Selection*, 13,1, 173-194, 2001.
23. C. H. Skiadas, Innovation Diffusion Models Expressing Asymmetry and/or Positively or Negatively Influencing Forces. *Technological Forecasting and Social Change* 30, 313-330, 1986.

24. C. H. Skiadas and C. Skiadas. Estimating the Healthy Life Expectancy from the Health State Function of a Population in Connection to the Life Expectancy at Birth. In *The Health State function of a population.* (Skiadas, C. H. and Skiadas, C., eds). 1st ed. Athens, 97-109, 2012.
25. C. H. Skiadas and Skiadas, C. *The Health State function of a Population. Supplement. Demographic and Human Development indicators. 35 countries studied*. ISAST, Athens, 2013.
26. C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*. 2nd edition. Athens, 2013.
27. C. H. Skiadas and C. Skiadas. *Supplement: The Health State Function of a Population.* ISAST, Athens, 2013.
28. C. H. Skiadas and C. Skiadas. *Demographic and Health indicators for 193 countries of the World Health Organization and the United Nations. Second supplement of the book The Health State Function of a Population.* Athens, 2014.
29. C. H. Skiadas and C. Skiadas. The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods*, 34, 1585-1600, 2014.
30. The Telegraph. What are Italy's fresh austerity measures. http://www.telegraph.co.uk/finance/financialcrisis/8883485/What-are-Italys-fresh-austerity-measures.html
31. S. Tomson, J. Figueras, T. Evetovits, M. Jowet, P. Mladovsky, Maresso, A. Cylus, J. Karanikolos, M. Kluge, H. Economic crisis, health systems and health in Europe : impact and implications for policy. WHO, Policy summary, 12
32. R. J. Wilmoth and S. Horiuchi. Rectangularization revisited: variability of age at death with in human population. *Demography*, 36,4, 475-495, 1999.
33. K. N. Zafeiris and C. H. Skiadas. Regional dimensions of mortality and health status in Greece. RELIK. Reprodukce lidského kapitálu -vzájemné vazby a souvislosti. 24. – 25. Listopadu, 564-575, 2014.

Random Sums of Dependent Random Variables: Strong Limit Theorems and Applications

Nadiia Zinchenko

Department of Informatics and Applied Mathematics, Nizhyn State Mukola Gogol University, Kropyv'yanskogo 2, 16600, Nizhyn, Ukraine
(e-mail: znm@univ.kiev.ua)

Abstract. We present sufficient conditions, which provide the strong approximation of the random sums of dependent random variables by a Wiener process, and apply them to study the asymptotic behavior of random sums. The cases of weakly dependent and associated summands are studied as well as random variables satisfying φ-mixing conditions. Various modifications of the LIL and Erdös-Rényi-Csörgő-Révész law are proved and used for investigation the rate of growth and fluctuations of random sums. Certain applications in the risk theory are discussed.
Keywords: Strong approximation, Invariance principle, Strong limit theorem, Random sums, Law of the iterated logarithm .

1 Introduction

This paper is a further development of the author's previous works ([19], [20]), where the strong approximation of the random sums $D(t) = \sum_{i=1}^{N(t)} X_i$ was studied, when $\{X_i, i \geq 1\}$ are *independent* random variables (r.v.) and $N(t)$ is a counting renewal process.

As usual, we say that a random process $\{D(t), t \geq 0\}$ admits *strong approximation* (other terms are *almost sure approximation*) by the random process $\{\eta(t), t \geq 0\}$ if $D(t)$ (or stochastically equivalent $D^*(t)$) can be constructed on the rich enough probability space together with $\eta(t)$ in such a way that a.s.

$$|D(t) - \eta(t)| = o(r(t)) \vee O(r(t)) \text{ as } t \to \infty, \tag{1}$$

where approximating error (error term) $r(.)$ is a non-random function. The class of limit theorems, which provide necessary and/or sufficient conditions for (1), are called *strong invariance principle*(**SIP**).

Zinchenko[18],[19] obtained some general results concerning sufficient conditions for strong approximation of random sums $D(t)$ by a Wiener or α-stable Lévy process under various conditions on the counting process N(t) and *independent* random summands $\{X_i, i \geq 1\}$. Corresponding proofs are based on the rather general theorems about the strong approximation of the superposition of càd-làg processes (not obligatory connected with partial sums). Such approach also occurs to be fruitful in the case of dependent summands. In the present paper we focuse on the random sums of two types of *dependent* variables (weakly dependent and associated r.v.), which have finite moments of order greater than 2, thus we consider only Wiener process as an approximating process in (1). Similarly to the case of i.i.d.r.v. SIP-type results itself can

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

serve as a source of a number of limit theorems. Really, using (1) with "small enough" error term, one can easily transfer the results about the asymptotic behavior of the Wiener process on the asymptotic behavior of random sums. In such a manner we establish various modifications of the LIL and Erdös-Rényi-Csörgő-Révész-type SLLN for dependent r.v., which describe the rate of growth and fluctuations of corresponding random sums.

2 Auxiliary results. SIP for superposition of the random processes

Next theorem (Zinchenko[19]) provide strong approximation of the superposition of the random processes $X(M(t))$, when càd-làg random processes $X(t)$ and $M(t)$ themselves admit a.s. approximation by a Wiener processes (see also Csörgő and L. Horváth[6]).

So, let $X(t)$ and $M(t)$ be independent separable real measurable càd-làg processes, $X(0) = 0$, $M(0) = 0$, $M(t)$ does not decrease with probability 1.

Theorem 1. *Suppose that there are independent standard Wiener processes $W_1(t)$ and $W_2(t)$, constants $m \in R^1$, $\lambda > 0$, $\tau_* > 0$, $\sigma > 0$, for which a.s.*

$$\sup_{0\le t\le T} |M(t) - \lambda t - \tau_* W_1(t)| = O(r(T)), \tag{2}$$

$$\sup_{0\le t\le T} |X(t) - mt - \sigma W_2(t)| = O(q(T)), \tag{3}$$

where $r(t) \uparrow \infty$, $r(t)/t \downarrow 0$, $q(t) \uparrow \infty$, $q(t)/t \downarrow 0$ as $t \to \infty$.

Let $\nu^2 = \sigma^2\lambda + m^2\tau_*^2$. Then $X(t)$ and $M(t)$ can be redefined on the one probability space together with a standard Wiener process $W(t)$ in such a way that a.s.

$$\sup_{0\le t\le T} |X(M(t)) - m\lambda t - \nu W(t)| = O(r(T) + q(T) + \ln T). \tag{4}$$

In forthcoming we use following notations and definitions. For a sequence of r.v. $\{X_i, i \ge 1\}$ denote

$$S(t) = \sum_{i=1}^{[t]} X_i, \quad t > 0, \quad [t] - \text{ entire of } t, \quad S(0) = 0.$$

Also suppose that $\{Z_i, i \ge 1\}$ is a sequence of non-negative i.i.d.r.v. independent of $\{X_i, i \ge 1\}$ with $EZ_1 = 1/\lambda > 0$,

$$Z(n) = \sum_{i=1}^{n} Z_i, \quad Z(0) = 0, \quad Z(x) = Z([x]),$$

and define the *renewal (counting) process* $N(t)$ associated with partial sums $Z(n)$ as

$$N(t) = \inf\{x \ge 0 : Z(x) > t\}.$$

Here and in the next sections consider *random sums (randomly stopped sums)* defined as

$$D(t) = S(N(t)) = \sum_{i=1}^{N(t)} X_i,$$

where r.v. $\{X_i, i \geq 1\}$ and renewal process $N(t)$ are given above.

Theorem 1 is rather convenient for investigation random sums. For successful application of this theorem to $D(t) = S(N(t))$ we need to have separately appropriate SIP-type results for usual sums $S(n)$ and renewal process $N(t)$. Historically most intensively strong approximation of usual partial sums was studied in the case of *independent* summands. Numerous investigations in this area were carried out by a number of authors, among them Kiefer, M.Csörgő, Révész, Komlós, Major, Tusnady, Berkes, Horváth (quantile Hungerian method), Stout, Phillip, Berkes (reconstruction method based on relationship between SIP and convergence in Prokhorov metrics), Horváth (inverse processes), Sakhanenko, Zaicev (sums of random vectors). For detail references see : Csörgő, Révész[5]; Csörgő and L. Horváth[6]; Zinchenko[16].

SIP for renewal processes also was studied. In the case $\tau^2 = varZ_1 < \infty$, $EZ_1 = 1/\lambda > 0$ Csörgő, Horvách, Steinebach, Alex, Deheuvels, Mason, van Zwet studied a.s. approximation of the type

$$\sup_{0 \leq t \leq T} |\lambda t - N(t) - \tau \lambda^{3/2} W(t)| = o(q(T)) \vee O(q(T)) \tag{5}$$

and proved that conditions, which provide (5) and corresponding optimal errors, are the same as for usual sums of i.i.d.r.v. $S(n)$ (see Alex and Steinebach[1], Csörgő and L. Horváth[6]):

Theorem 2. *(i) Let $E|Z_1|^p < \infty$, $p > 2$, then $N(t)$ can be constructed on the same probability space together with a Wiener process $\{W_1(t), t \geq 0\}$ in such a way that a.s.*

$$\sup_{0 \leq t \leq T} \left| \lambda t - N(t) - \tau \lambda^{3/2} W_1(t) \right| = o(T^{1/p}); \tag{6}$$

(ii) if $p = 2$ then right side of (6) is $o(T \ln \ln T)^{1/2}$; (iii) if $E \exp(uZ_1) < \infty$ for all $u \in (0, u_o)$, then right-hand side of (6) is $O(\ln T)$.

Corollary 1. *If $N(t)$ is a Poisson process then a.s.*

$$\sup_{0 \leq t \leq T} |\lambda t - N(t) - \lambda^{1/2} W_1(t)| = O(\ln T). \tag{7}$$

Further development was connected with dependent r.v.: martingales, weakly dependent r.v., mixing and associated sequences. Investigations in this area was initiated by Strassen (1965), who studied the sums of martingales. During last two decades the interest to SIP for dependent r.v. remarkably increased, a lot of interesting results can be found in Philipp and Stout [13], Berkes and Philipp[3], Philipp[12], Yu [21], Wu [15], Dedecker, Doukhun and Merlevède [8] and in fundamental monograph by Bulinski and Shashkin[4]. Some of these results will be used in next sections.

3 SIP for random sums of dependent r.v.

Throughout this Section, unless otherwise stated, we suppose that inter-occurrence time intervals $\{Z_i\}$ for renewal process $N(t)$ have finite moments $E|Z_1|^p < \infty$ of order $p > 2$.

1. We start with the simple case of **Gaussian sequences** with n-th partial sum having variance close to n and with covariances converging to zero as distance between indices approaches infinity.

Theorem 3. *Let $\{X_i, i \geq 1\}$ be a Gaussian sequence centered at expectations. Suppose that uniformly in m*

$$E\left(\sum_{i=m+1}^{m+n} X_i\right)^2 = \sigma_1^2 n + O(n^{1-\theta}) \tag{8}$$

for some $\theta > 0$ and $\sigma_1 > 0$, also uniformly in m

$$EX_m X_{n+m} = O(n^{-2}), \tag{9}$$

then $\{X_i\}$ and $N(t)$ can be constructed on the same probability space together with a Wiener process $\{W(t), t \geq 0\}$ in such a way that a.s.

$$\sup_{0 \leq t \leq T} |S(N(t)) - \nu_1 W(t)| = O(T^{1/2-\vartheta}), \quad \nu_1^2 = \sigma_1^2 \lambda \tag{10}$$

for $0 < \vartheta < \vartheta_1, \vartheta_1 = \vartheta_1(\theta, p)$.

Proof. Philipp and Stout[13] proved that under conditions (8), (9) partial sums $S(t) = \sum_{i=1}^{[t]} X_i$ of Gaussian variables admit following a.s. approximation by a standard Wiener process $W_2(t)$

$$\sup_{0 \leq t \leq T} |S(t) - \sigma_1 W_2(t)| = O(q(T)), \quad q(T) = T^{1/2-\kappa} \tag{11}$$

for each $0 < \kappa < min(1/60, 4\theta/15)$. Now the proof of Theorem 3 immediately follows from Theorem 1 and form of error terms $q(t)$ and $r(t)$ in SIP-type results for $S(t)$ and renewal process $N(t)$ (see Theorem 2).

For the *stationary* Gaussian sequences moment assumptions on summands $\{X_i, i \geq 1\}$ in corresponding SIP for $S(t)$ are less restrictive, see Philipp and Stout[13], Berkes and Philipp[3], Morrow[11]. Thus we have

Theorem 4. *Let $\{X_i, i \geq 1\}$ be a stationary Gaussian sequence, $EX_1 = m$. Suppose that for some $\theta > 0$*

$$E\{(X_1 - m)(X_n - m)\} = O(n^{-1-\theta}), \tag{12}$$

$$0 < E(X_1 - m)^2 + 2\sum_{i>1} E(X_1 - m)(X_i - m) = \sigma_2^2 < \infty. \tag{13}$$

Then there exist Wiener process $\{W(t), t \geq 0\}$ such that a.s.

$$\sup_{0 \leq t \leq T} |S(N(t)) - mt\lambda - \nu_2 W(t)| = O(T^{1/2-\vartheta}), \quad \nu_2^2 = \sigma_2^2 \lambda + m^2 \tau^2 \lambda^3, \tag{14}$$

where $\vartheta \in (0, \vartheta_1)$ for some $\vartheta_1 = \vartheta_1(\theta, p)$.

Again the proof of Theorem 4 follows from Theorems 1, 2 and form of error term $q(t)$ in SIP-type theorems for $S(t)$ due to Morrow[11], who established the possibility to approximate $S(t)$ by a standard Wiener $W_2(t)$ process in such a way that a.s.

$$\sup_{0\le t\le T}|S(t)-mt-\sigma_2 W_2(t)|=O(T^{1/2-\vartheta}),\quad \vartheta=\min(1,\theta)/500. \tag{15}$$

2. On the next step consider stationary **φ-mixing sequences**.

Given a sequence $\{X_i, i\ge 1\}$, let F_a^b denote the σ-field generated by $X_a, X_{a+1},\dots,X_b$, $a<b<\infty$, and F_b^∞ – the σ-field generated by $X_b, X_{b+1},\dots$. Then the sequence is said to be φ-mixing if there exist a sequence $\{\varphi(n)\}$ of real numbers, $\varphi(n)\downarrow 0$ as $n\to\infty$, such that for each $t\ge 1$, $n>0$, $A\in F_1^t$, $B\in F_{t+n}^\infty$

$$|P(AB)-P(A)P(B)|\le\varphi(n)P(A) \tag{16}$$

Theorem 5. *Let $\{X_i, i\ge 1\}$ be a strictly stationary φ-mixing sequence with $EX_1=m$, $E|X_1|^{2+\delta}<\infty$. Suppose*

$$\sum_{n=1}^{\infty}\phi^{1/2}(n)<\infty \tag{17}$$

and

$$0<\lim_{n\to\infty}n^{-1}E\left(\sum_{i=1}^{n}(X_i-m)\right)^2=\sigma_3^2<\infty. \tag{18}$$

Then $\{X_i\}$ and $N(t)$ can be constructed on the same probability space together with a Wiener process $\{W(t), t\ge 0\}$ in such a way that a.s.

$$\sup_{0\le t\le T}|S(N(t))-mt\lambda-\nu_3 W(t)|=O(T^{1/2-\vartheta}),\quad \nu_3^2=\sigma_3^2\lambda+m^2\tau^2\lambda^3, \tag{19}$$

where $\vartheta\in(0,\vartheta_1)$ for some $\vartheta_1=\vartheta_1(\delta,p)$

As above proof easily follows from Theorem 1 and possibility of strong approximation of the partial sums $S(t)$ of mentioned φ-mixing r.v. by a standard Wiener process $W_2(t)$ (see Philipp and Stout[13]) in such a way that a.s.

$$\sup_{0\le t\le T}|S(t)-mt-\sigma_3 W_2(t)|=O(T^{1/2-\vartheta}),\quad \forall\vartheta<\delta/(24+12\delta). \tag{20}$$

Remark 1. If inter-occurrence intervals $\{Z_i, i\ge 1\}$ have only second moment, then conclusions of Theorems 3-5 hold with error-term $o((T\ln\ln T)^{1/2})$. Such error term also appears in (19) and (20) if only finiteness of second moment of summands is assumed.

Remark 2. Statements (10),(14), (19) become more simple for summands centered at expectations, in this case $\nu_i^2=\sigma_i^2\lambda$, $i=1,2,3$.

3. Finally, we consider the case of **associated summands**.

Definition. R.v. $X_1, \ldots, X_n$ are **associated**, if for any two coordinate-wise nondecreasing functions $f, g : R^n \to R^1$,

$$Cov\big(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)\big) \geq 0$$

whenever the covariance is defined. A sequence $\{X_i, i \geq 1\}$ is associated, if every finite sub-collection is associated.

Useful SIP-type results for partial sums of associated summands were obtained by Wu[15], Hu[21], Bulinski and Shashkin[4]:

Theorem 6. *Let $\{X_i, i \geq 1\}$ be a strictly stationary associated sequence, $EX_1 = m$. Suppose that $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$ and Cox-Grimmett coefficient*

$$u(n) = \sup_{k \geq 1} \sum_{j:|j-k| \geq n} Cov(X_j X_k) = O(e^{-\theta n}) \tag{21}$$

for some $\theta > 0$. Denote

$$E(X_1 - m)^2 + 2 \sum_{i>1} E(X_1 - m)(X_i - m) = \sigma^2 > 0. \tag{22}$$

Then there exist standard Wiener process $\{W_2(t), t \geq 0\}$ such that for some $\theta = \theta(\theta, \delta) > 0$ a.s.

$$\sup_{0 \leq t \leq T} |S(t) - mt - \sigma W_2(t)| = O(q(T)) = O(T^{1/2-\theta}). \tag{23}$$

Substituting this $q(T)$ and corresponding $r(T)$ in Theorem 1 we easily obtain

Theorem 7. *Let $\{X_i, i \geq 1\}$ be a strictly stationary associated sequence satisfying all conditions of the previous Theorem 6, inter-occurrence intervals $\{Z_i, i \geq 1\}$ be i.i.d.r.v. with $0 < EZ_1 = 1/\lambda < \infty$, $\tau^2 = VarZ_1 < \infty$. Denote $\nu^2 = \sigma^2\lambda + m^2\tau^2\lambda^3$. Then $\{X_i\}$ and $N(t)$ can be constructed on the same probability space together with a Wiener process $\{W(t), t \geq 0\}$ in such a way that a.s.*

$$\sup_{0 \leq t \leq T} |S(N(t)) - mt\lambda - \nu W(t)| = O(\varrho(T)), \tag{24}$$

where error term is $\varrho(T) = T^{1/2-\vartheta_1}$ for some $\vartheta_1 = \vartheta_1(\delta, p)$, when $E|Z_1|^p < \infty$ for $p > 2$, and right-hand side of (24) is $o((T \ln \ln T)^{1/2})$, when Z_1 has only second moment.

Corollary 2 (SIP for Poisson random sums). *Theorems 3-5, 7 hold if $N(t)$ is a homogeneous Poisson process with intensity $\lambda > 0$.*

4 The rate of growth and fluctuations of the random sums

One of the main reasons of increasing interest to SIP-type theorems is a possibility to investigate with their help the asymptotic behavior of usual and random sums. Indeed, using SIP with appropriate error term, one can easily extend the results about the asymptotic behavior of the Wiener process on the rate of growth of random sums $D(t) = S(N(t))$. Formalizing this idea and extending the approach due to Philipp and Stout[13], we formulate rather general theorems (again not obligatory connected with random sums).

Theorem 8. *Suppose that random process $D(t)$ admits a.s. approximation by a standard Wiener process $W(t)$ with an error term $O(t^{1/p}), p > 2$, i.e. a.s.*

$$\sup_{0\le t\le T} |D(t) - Mt - \nu W(t)| = O(T^{1/p}), \quad M \in R^1, \nu > 0, \tag{25}$$

then a.s.

$$\limsup_{t\to\infty} \frac{|D(t) - Mt|}{\sqrt{2t \ln\ln t}} = \nu. \tag{26}$$

Statement (26) is a straightforward consequence of the classical LIL for a Wiener process and form of error term in (25). On the other hand, from Chung's LIL for Wiener process it easily follows

Theorem 9. *Let $D(t)$ be as in previous Theorem, then a.s.*

$$\liminf_{t\to\infty} \left(\frac{8 \ln\ln T}{\pi^2 T}\right)^{1/2} \sup_{0\le t\le T} |D(t) - MT| = \nu. \tag{27}$$

The rate of growth of increments of a Wiener process on intervals whose length a_t grows, but not faster than t, is well described by Erdös-Rényi-Csörgő-Révész law (Csörgő and Révész[5]), which leads to following

Theorem 10. *Suppose that $D(t)$ satisfies (25) and function $a_T, T \ge 0$ satisfies following conditions: (i) $0 < a_T < T$, (ii) T/a_T does not decrease in T. Also assume that $a_T > c_1 T^{2/p}/\ln T$ for some $c_1 > 0$. Then a.s.*

$$\limsup_{T\to\infty} \frac{|D(T + a_T) - D(T) - Ma_T|}{\gamma(T)} = \nu, \tag{28}$$

where

$$\gamma(T) = \{2a_T(\ln\ln T + \ln T/a_T)\}^{1/2}.$$

Since random sums $D(t) = S(N(t))$ of dependent r.v., introduced in Section 3, satisfy (25) with $M = \lambda m$, $\nu^2 = \sigma^2\lambda + m^2\tau^2\lambda^3$ and $1/p = (1/2) - \vartheta$ for some $\vartheta > 0$, Theorems 8 -10 yield following Corollaries:

Corollary 3 (Classical LIL for random sums, associated summands). *Let $\{Z_i\}$ be i.i.d.r.v. with $0 < EZ_1 = 1/\lambda < \infty$, $\tau^2 = VarZ_1 < \infty$, $\{X_i\}$ constitute the strictly stationary associated sequence with mean $EX_1 = m$ and covariance, satisfying sufficient conditions for SIP (Theorem 7), then a.s.*

$$\limsup_{t\to\infty} \frac{|S(N(t)) - m\lambda t|}{\sqrt{2t\ln\ln t}} = \nu, \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2\tau^2. \tag{29}$$

Corollary 4 (Classical LIL for random sums, ϕ-mixing summands). *Statement analogous to (28) holds with corresponding σ and ν for strictly stationary ϕ-mixing summands satisfying all conditions of Theorem 5.*

Corollary 5 (Chung's LIL for random sums). *Let $\{X_i\}$ and $\{Z_i\}$ be as in Corollary 3 (or as in Corollary 4), then a.s.*

$$\liminf_{t\to\infty} \left(\frac{8\ln\ln T}{\pi^2 T}\right)^{1/2} \sup_{0\le t\le T} |S(N(t)) - m\lambda t| = \nu, \quad \nu^2 = \lambda\sigma^2 + \lambda^3 m^2\tau^2. \tag{30}$$

When summands are centered at expectation such statements become simpler with $\nu^2 = \lambda\sigma^2$, but the most simple form they have when $N(t)$ is a Poisson process.

Corollary 6. *Corollaries 3 – 5 hold when $N(T)$ is a homogeneous Poisson process with intensity $\lambda > 0$.*

Next theorem gives the partial answer on the question: how big are increments of the random sums of dependent r.v.? A number of results in this area (but for independent summands) are presented by Zinchenko and Safonova[17], Frolov[9], Martikainen and Frolov[10].

Theorem 11. *Let $N(T)$ be homogeneous Poisson process with intensity $\lambda > 0$ and let $\{X_i\}$ be the strictly stationary associated sequence with mean $EX_1 = m$ and covariance, satisfying sufficient conditions for SIP (Theorem 7). Suppose that function $a_T, T \ge 0$ satisfies all conditions of Theorem 10 and $a_T > c_1 T^{2/p}/\ln T$ for some $c_1 > 0$, $1/p = (1/2) - \vartheta$, $\vartheta > 0$.*

Then a.s.

$$\limsup_{T\to\infty} \frac{|S(N(T + a_T)) - S(N(T)) - m\lambda a_T|}{\gamma(T)} = \nu, \tag{31}$$

where

$$\nu^2 = \lambda(\sigma^2 + m^2), \quad \gamma(T) = \{2a_T(\ln\ln T + \ln T/a_T)\}^{1/2}.$$

5 Final remarks

It is well known, that random sums often appear as useful models in various areas: in queuing theory, in risk theory, in financial mathematics [7]. For

instance, in popular collective risk model the risk process, which describes the evolution of reserve capital, is defined as

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i = u + ct - S(N(t)), \quad c > 0, \quad u > 0, \tag{32}$$

where random sum $D(t) = S(N(t))$ is interpreted as a total claim amount arising during time interval $[0, t]$ and its increments $D(t + a_t) - D(t)$ - as claim amount arising during time interval $[t, t + a_t]$. Main known results concerning $U(t)$ and $D(t)$ are focused on the case of *independent* claim sizes $\{X_i, i \geq 1\}$. Our approach allows to study the case of *dependent* claims too. Thus, certain results about approximation of ruin probabilities, bounds for rates of growth and fluctuations of total claim amounts can be obtained similar to how it was done in [20] for independent summands.

References

1. M. Alex, J. Steinebach. Invariance principles for renewal processes and some applications. *Teor. Jmovirn. Mat. Stat.* 50, 22–54, 1994.
2. R. Balan. A strong invariance principle for associated random fields *Ann. Probab.* 33, 823–840, 2005.
3. I. Berkes, W. Philipp. Approximation theorems for independent and weakly dependent random vectors.*Ann. Probab.* 7, 29–54, 1979.
4. A.V. Bulinski, A.P. Shashkin. *Limit theorems for associated random fields and related systems*, Fizmatlit, Moscow, 2008.
5. M. Csörgő, P. Révész. *Strong Approximation in Probability and Statistics*, Acad. Press, New York, 1981.
6. M. Csörgő, L. Horváth. *Weighted Approximation in Probability and Statistics*, Wiley, New York, 1993.
7. B.V. Gnedenko, V.Yu. Korolev. *Random Summation: Limit Theorems and Applications*, CRT Press, Boca - Raton, Florida, 1996.
8. J. Dedecher, P. Doukhan, T. Merlevède. Rates of convergence in the strong invariance principle under projective criteria. *Electron J. Probab.* 17, 1–31, 2012.
9. A. Frolov. Strong limit theorems for increments of compound renewal processes. *Journal of Math. Sciences* 152, 944–957, 2008.
10. A. Martikainen, A. Frolov. On the Chung law for compound renewal processes. *Journal of Math. Sciences*, 145, 2, 4866–4870, 2007.
11. G. Morrow. Invariance principle for Gaussian sequences *Z. Wahr. verw. Geb.*, 52, 115–126,1980
12. W. Philipp. Invariance principle for independent and weakly dependent r.v In:*Dependence in Probability and Statistics: A Survey of Recent Results (E. Eberlein and M.S.Taqqu, eds)*, Birkhäuser, Boston, 1986.
13. W. Philipp, W. Stout. Almost sure invariance principle for partial sums of weakly dependent random variables. *Mem. of Amer. Math. Society*, 161, 1975.
14. V. Strassen, *Almost sure behavior for sums of independent r.v. and martingales,* Pros. 5th Berkeley Symp.**2**, 315–343, 1965
15. W. Wu. Strong invariance principle for dependent random variables. *Annals of Probability* 35, 6, 2294–2330, 2007.
16. N. Zinchenko. The Skorokhod representation and strong invariance principle. *Teor. Jmovirn. Mat. Stat.* 60, 51–63, 2000.

17. N. Zinchenko, M. Safonova. Erdös-Renyi type law for random sums with applications to claim amount process. *Journal of Numerical and Applied Mathematics* 1(96), 246–264, 2008.
18. N. Zinchenko. Strong invariance principle for a superposition of random processes. *Theory of Stoch. Processes* 16(32), 130–138, 2010.
19. N. Zinchenko. Almost sure approximation of the superposition of the random processes. *Methodology and Computing in Applied Probability*, 17, 1, 235–250, 2015, DOI 10.1007/s 11009-013-9350-y.
20. N. Zinchenko. Strong approximation of random sums with applications in queuning and risk theories. *Proceedings of 3-d International Conference "Stochastic Modeling Techniques and Data Analysis" (SMTDA2014)*, Lisbon, 877– 888, 2014.
21. H. Yu. A strong invariance principle for associated sequences. *Annals of Probability*, 24, 4, 2079–2097, 1996.