UNIVERSITA' DEGLI STUDI DI VERONA

DEPARTMENT OF

DIAGNOSTICS AND PUBLIC HEALTH

GRADUATE SCHOOL OF

LIFE AND HELATH SCIENCES

DOCTORAL PROGRAM IN

APPLIED LIFE AND HELATH SCIENCES

WITH THE FINANCIAL CONTRIBUTION OF
(NAME IF THE FUNDING INSTITUTION)

UNIVERSITY OF VERONA

Cycle / year (1° year of attendance) 29°

TITLE OF THE DOCTORAL THESIS

DETERMINATION OF REFERENCE INTERVALS OF OXIDATIVE STRESS BIOMARKERS IN HEALTHY ITALIAN POPULATION

S.S.D. MED/01 STATISTICA MEDICA
(Please complete this space with the S.S.D. of your thesis – mandatory information)*

Coordinator: Prof./ssa Giovanni Malerba

Signature _____

Tutor: Prof./ssa Maria Elisabetta Zanolin

Signature _____

Doctoral Student: Dott./ssa Liliya Chamitava

Signature _____

* For the list of S.S.D. please refer to the Ministerial Decree of 4th October 2000, Attachment A "*Elenco dei Settori Scientifico – Disciplinari*" available at: http://www.miur.it/UserFiles/115.htm

Determination of reference intervals of oxidative stress biomarkers in healthy Italian population
Liliya Chamitava
PhD thesis
Verona, 4 July 2017

University of Verona
Research Office – National and International PhD programmes
ph: 045.802.8608 – fax 045.802.8411 – Via Giardino Giusti 2 – 37129 Verona

TABLE OF CONTENTS

SUMMARY.

Background. Recent interest to investigation of oxidative stress (OS) linked disorders is due to its high involvement in many chronic and acute pathologies: cardiovascular, lung, hematological, diabetes, cancer and many others. Oxidative stress is an imbalance between production of reactive oxygen species (ROS) and antioxidant defensive capacity of organism. Normally ROS are generated in human body and are considered to be natural byproducts of metabolism of oxygen. But in the framework of oxidative stress, ROS suppress antioxidant capacity in vivo, damaging DNA, lipids, and proteins. ROS are highly reactive molecules with a short half-life and therefore it is very difficult to detect them in human biological fluids. Instead, the products of their oxidizing reactions, denoted as biomarkers of oxidative stress, are used widely to detect the level of OS in a human body.

There are statistical techniques which allow estimating non-adjusted for determinants or adjusted for determinants reference intervals of biomarkers of oxidative stress.

Aim. In this work we aimed at identifying main demographic and laboratory determinants of urinary biomarkers of oxidative stress: DNA-derived 8-oxodG and lipid membrane-derived 8-isoprostane, and at defining their reference intervals, adjusted for main determinants, in a sample of healthy people from the general Italian population.

Methods. In current study, the data on 281 subjects from general Italian population, gathered during the Gene Environment Interactions in Respiratory Diseases (GEIRD) project who were never-and ex- (not smoking over the last year) smokers and who did not report respiratory symptoms at the screening questionnaire and at the clinical survey, either other comorbidities (heart disease, ictus, high blood pressure, diabetes and cancer) at the clinical interview and with normal lung function and allergologic test have been used.

Results. Non adjusted reference intervals of 8-oxodG and of 8-isoprostane were evaluated first in this work. The main determinants of both biomarkers of oxidative stress were 'distance from collection' (DFC, the period from the moment of urine collection and its laboratory processing), and season, the period of year when urine was collected and split into warm (*April - September*) and cold (*October - March*). The reference intervals of 8-oxodG and of 8-isoprostane stratified by season and adjusted for DFC were predicted using GAMLSS (generalized additive models for location, scale and shape) regression analysis and showed slight but statistically significant degradation of both biomarkers with increase of DFC in both seasons, except 8-oxodG biomarker during the warm season, which provided unchanged values with increased DFC.

Conclusion. To our knowledge it is for the first time when it was shown that both OS biomarkers 8-oxodG and 8-isoprostane should be evaluated in association with DFC and season, when urine has been collected. It is especially important in large epidemiological studies when long-term conservation of urine is stipulated. *(Semi)*parametric GAMLSS regression analysis is a new useful technique that can be used for estimating reference intervals of urinary biomarkers (8-oxodG and 8-isoprostane) from general adult population and adjusted for appropriate determinants.

ABSTRACT.

Sfondo. Il recente interesse per lo stress ossidativo (OS) è dovuto al suo elevato coinvolgimento in molte patologie croniche e acute cardiovascolari, polmonari, ematologiche, diabete, cancro e molte altre. Lo stress ossidativo è dovuto ad uno squilibrio tra la produzione di specie reattive dell'ossigeno (ROS) e la capacità antiossidante difensiva dell'organismo. Solitamente i ROS vengono generati nel corpo umano e sono considerati sottoprodotti naturali del metabolismo dell' ossigeno. Ma in termini di stress ossidativo, i ROS sopprimono la capacità antiossidante *in vivo*, danneggiando DNA, lipidi e proteine. I ROS sono molecole altamente reattive con una emivita breve e quindi sono molto difficili da misurare nei fluidi biologici umani. Invece, i prodotti delle loro reazioni ossidanti, chiamati marcatori di stress ossidativo, vengono ampiamente utilizzati per rilevare il livello di stress ossidativo in un corpo umano.

Ci sono tecniche statistiche che consentono di stimare intervalli di riferimento per biomarcatori di stress ossidativo, corretti o meno per eventuali determinanti.

Scopo. In questo lavoro, abbiamo individuato i principali fattori demografici e di laboratorio determinanti dei biomarcatori di stress ossidativo nelle urine: l'8-oxodG, derivato dal DNA, e l'8-isoprostano, derivato dai lipidi di membrana, e definire i loro intervalli di riferimento, aggiustando per i principali determinanti, in un campione di persone sane della popolazione generale italiana.

Metodi. Nello studio corrente, si sono utilizzati i dati su 281 soggetti della popolazione generale italiana, raccolti nel progetto GEIRD (Gene Environment Interactions in Respiratory Diseases): questi soggetti erano non-fumatori o ex-fumatori (non-fumatori nel corso dell'ultimo anno), non avevano riportato sintomi respiratori al questionario di screening e alla indagine clinica, né altre comorbidità (malattie cardiache, ictus, pressione alta, diabete e cancro) al colloquio clinico e presentavano una funzione polmonare e test allergologico normali.

Risultati. In un primo momento, si sono calcolati gli intervalli di riferimento di 8-oxodG e di 8-isoprostano non aggiustati. Si sono poi individuati i principali determinanti di entrambi i biomarcatori di stress ossidativo che sono risultati essere la 'distanza dal prelievo' (DFC, il periodo che va dal momento della raccolta delle urine e la sua analisi in laboratorio), e la stagione, il periodo dell'anno in cui l'urina è stata raccolta, diviso in caldo (aprile - settembre) e freddo (ottobre-marzo). Gli intervalli di riferimento di 8-oxodG e di 8-isoprostano stratificati per stagione e corretti per DFC sono stati calcolati tramite l'analisi di regressione GAMLSS (modelli additivi generalizzati per posizione, scala e forma): tale analisi ha rilevato un lieve ma statisticamente significativo degrado dei due biomarcatori all'aumentare della DFC in entrambe le stagioni, ad eccezione dell' 8-oxodG nella stagione calda, che è rimasto invariato all'aumentare di DFC.

Conclusione. Sulla base di quanto a noi noto, è la prima volta che è stata dimostrata un'associazione tra entrambi i biomarcatori di OS, 8-oxodG e 8-isoprostano, con DFC e con la stagione in cui l'urina è stata raccolta. Questa informazione è particolarmente importante, nei grandi studi epidemiologici, quando può essere necessaria una conservazione a lungo termine dell'urina. L'analisi di regressione (Semi)parametrica GAMLSS è una nuova tecnica utile che può essere utilizzata per stimare gli intervalli di riferimento dei biomarcatori urinarii (8-oxodG e 8-isoprostano) della popolazione adulta, aggiustati per i loro determinanti .

ABBREVIATIONS

AA, Arachidonic Acids
AIC, Akaike Information Criterion
BCCG, Box-Cox Cole and Green (Distribution)
BCT, Box-Cox *t* (Distribution)
BIC, Bayesian Information Criterion
CAT, Catalase
CB, Chronic Bronchitis
cdf, Cumulative Distribution Function
COPD, Chronic Obstructive Pulmonary Disease
CuZnSOD, Copper Zinc Superoxide Dismutase
DFC, Distance from Collection
df, Degrees of Freedom
DNA, Deoxyribonucleic Acid
ECRHS, European Community Respiratory Health Survey
edf, Effective Degrees of Freedom
ELISA, enzyme-linked immunosorbent assay
ESCODD, European Standards Committee on Oxidative DNA Damage
GA, Gamma (Distribution)
GAMLSS, Generalized Additive Models for Location, Scale and Shape
GEIRD, Gene Environment Interactions in Respiratory Disease
GPx, Glutathione Peroxidase
GSH, Reduced Glutathione
GSHPX, Glutathione Peroxidases
GSSG, Oxidized Glutathione
GST, Glutathione Transferase
HPLC, high pressure liquid chromatography
H_2O, Water
H_2O_2, Hydrogen Peroxide
IQR, Interquartile Range
LC-MS-MS, Liquid Chromatography-Mass Spectrometric
LOGNO, Log Normal (Distribution)
MnSOD, Manganese Superoxide Dismutase
MSR, Methionine Sulfoxide Reductase
NADPH oxidase, Nicotinamide Adenine Dinucleotide Phosphate-Oxidase
NO, Normal Distribution
NOS, Nitric Oxide Synthetase
OS, Oxidative Stress
O_2, Oxygen
O_3, Ozone
pdf, Probability Density Function

PUFA, Polyunsaturated Fatty Acids
PNA, Ribonucleic Acid
QR, Quantile Regression
RI, Reference Interval
ROS, Reactive Oxygen Species
RSH-Px, Thiol-Specific Peroxidase
SD, Standard Deviation
SE, Standard Error
SOD, Superoxide Dismutase

1 INTRODUCTION

Recent interest to investigation of oxidative stress (OS) linked disorders is due to its high involvement in many chronic and acute pathologies (1): cardiovascular, lung, hematological, diabetes, cancer and many others (Figure 1-1) (2). Oxidative stress is an imbalance between production of reactive oxygen species (ROS) and antioxidant defensive capacity of organism (3). The imbalance in oxidant-antioxidant system can be due to diminishing antioxidants or increasing production of ROS in human organism (4). ROS include a huge variety of different oxygen-centered radicals: superoxide $O_2^{\cdot}$-, hydroxyl $OH^{\cdot}$, peroxyl $RO_2^{\cdot}$, alkoxyl $RO^{\cdot}$, hydroperoxyl $HO_2^{\cdot}$, etc., as well as nonradical oxygen derivatives: hydrogen peroxide H_2O_2, hypochlorous acid HOCl, hypobromous acid HOBr, ozone O_3, etc. (4, 5).

Figure 1-1 ***Oxidative stress in different pathologies (2).***

Almost each cell in human organism is provided by antioxidant system which can be divided into primary protection preventing oxidant formation, secondary scavenging ROS, and tertiary removing or repairing oxidatively modified molecules (3, 4). Antioxidant system is complex and consists of endogenous and diet-derived molecules. One of the endogenous antioxidants is enzymatic superoxide dismutase (SOD), presented in mitochondria with manganese at active site (MnSOD), with zinc and copper at active site (CuZnSOD) in cytosol (3, 4). It converts O_2^{-} into H_2O_2 and then into water, H_2O, and oxygen, O_2 (4). But the most important antioxidants responsible for removing H_2O_2 from human cells are glutathione peroxidases (GSHPX). GSHPX utilize

H_2O_2 to oxidize reduced glutathione (GSH) to oxidized glutathione (GSSG) (4). GSH is a low-molecular nonenzymatic antioxidant and can scavenge different reactive species, e.g. HOCl. From dietary antioxidants the most important is α-Tocopherol with vitamin E activity. It is known because of its capacity to scavenge free radicals (4).

OS has both endogenous and exogenous origin (Figure 1-2) (6). Normally ROS are generated in human body and are considered to be natural byproducts of metabolism of oxygen (7). But in the framework of oxidative stress, ROS suppress antioxidant capacity *in vivo*, damaging DNA, lipids, and proteins (8). External oxidants, such as ozone, sulphur dioxide, nitrogen dioxide and also tobacco smoking or radiation, by-turn, injure lungs, which leads to oxidant-mediated cellular damage and worsening the OS derived disease (9). ROS are highly reactive molecules with a short half-life and therefore it is very difficult to detect them in human biological fluids. Instead, the products of their oxidizing reactions are used widely to detect the level of OS in an organism (9, 10).

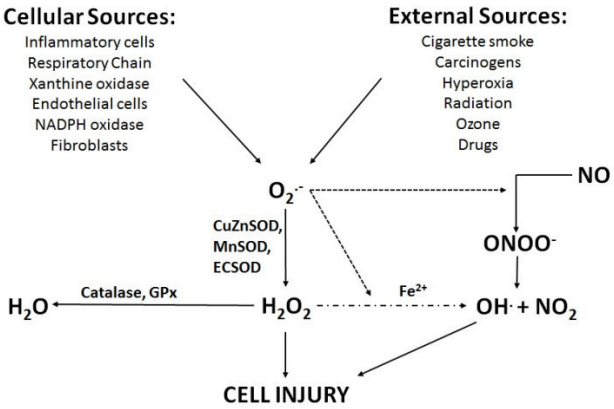


Figure 1-2. ***Generation of ROS and antioxidant system.*** *NADPH oxidase - nicotinamide adenine dinucleotide phosphate-oxidase, $O_2^{\cdot-}$ - superoxide anion, NO – nitric oxide, H_2O_2 – hydrogen peroxide, $HO^{\cdot}$ – hydroxyl radical, NO_2 – nitrogen dioxide, $ONOO^-$ - peroxynitrite, Fe^{2+} - ferrous ion, GPx – glutathione peroxidase, CuZnSOD - copper zinc superoxide dismutase, MnSOD - manganese superoxide dismutase, ECSOD - extracellular superoxide dismutase (6).*

Nowadays it is hypothesized that oxidative stress has hierarchical structure (Figure 1-3) (11). Mild oxidative stress induced by some environmental exposure can activate the transcription factor, nuclear erythroid 2 p45-related factor 2 (Nrf2), which is responsible for encoding more than 200 genes engaged in antioxidant, anti-inflammatory, cytoprotective, and detoxification functions with the involvement of catalase, superoxide dismutase (SOD)-3, heme oxygenase-1, glutathione-S-transferases, NAD(P)H-quinone oxidoreductase (NQO1), glutathione peroxidase, and glucuronosyltransferase-1a6 (UGT-1a6). In presence of slight oxidative stress antioxidants are able to restore cellular redox homeostasis. Stimuli which are present at high level of oxidative stress initiate

additional sets of intracellular proinflammatory signaling, engaging mitogen-activated protein kinase and nuclear factor-kB (NF-kB) and expressing inflammatory cytokines, chemokines, and adhesion molecules. In case of growing oxidative stress a cytotoxic response originating in the mitochondria and cellular apoptosis or necrosis can occur (11).

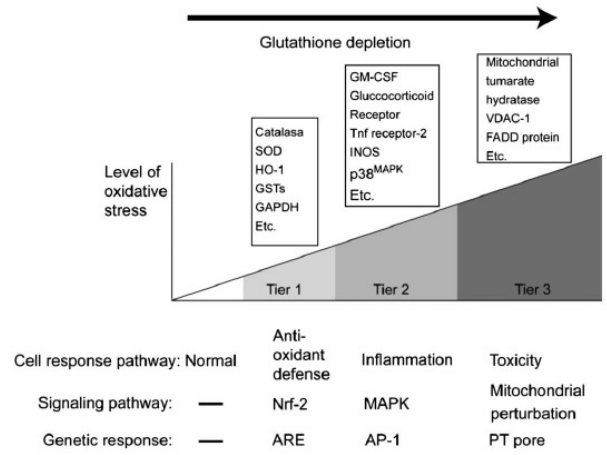


Figure 1-3. ***Hierarchical model of oxidative stress*** *(11).*

1.1 Oxidative Stress and Respiratory Health.

Due to anatomy, function and location lungs is an organ easily affected by exogenous ROS (3, 12) which could provoke onset of lung diseases (13) and, as a consequence, production of endogenous ROS. For example, an exposure to paraquat – a known herbicide, leads to lung injury inducing oxidative stress and lung inflammation (13). An accumulation of inflammatory cells in lower respiratory tract leads to increased oxygen burden (6). There is a number of lung cells engaged in a production of endogenous ROS, i.e. neutrophils, eosinophils, alveolar macrophages, peripheral monocytes-macrophages, mast cells, type II pneumocytes and so on (3). Such a production of ROS in inflammatory cells serves as a protection against invading organisms (13); but in chronic inflammatory conditions macrophages and neutrophils become a persistent source of oxidative DNA degradation which can lead to lung diseases (12), such as respiratory distress syndrome, asthma, chronic obstructive pulmonary disease (COPD), idiopathic pulmonary fibrosis, pneumoconiosis, etc. (6, 14). Recently it was shown that oxidative stress plays an important role in the pathogenesis of chronic bronchitis (CB) (15). Patients with CB had increased OS (malondialdehyde - a biomarker of lipid peroxidation) and decreased levels of antioxidants (SOD, glutathione) in respect to group of controls (15). Inflammatory and immune cells in patients with asthma generate increased number of ROS, which leads to tissue injury and further oxidative stress (6, 16). Both endogenous and exogenous ROS determine the severity of asthma (Figure 1-4) and COPD (11, 17). ROS and products of their oxidation reactions (lipid peroxidation) enhance

inflammatory response both in asthma and COPD in different ways: through impact on transduction mechanisms, activation of redox-sensitive transcriptions factors, and chromatin regulation which leads to pro-inflammatory gene expression. Moreover, ROS diminish activity of histone deacetylase-2 (HDAC-2) co-repressor, resulting on low efficacy of corticosteroids in COPD, severe asthma, and smoking asthmatics (17).

Oxidative stress can mediate such allergic respiratory diseases as rhinitis, and atopic dermatitis (18). House dust induces increased generation of H_2O_2 by nasal eosinophils in subjects with allergic rhinitis. Reduced levels of vitamin E, catalase, and glutathione peroxidase have been found in blood of subjects with skin allergic disorders (physical urticarias) (18).

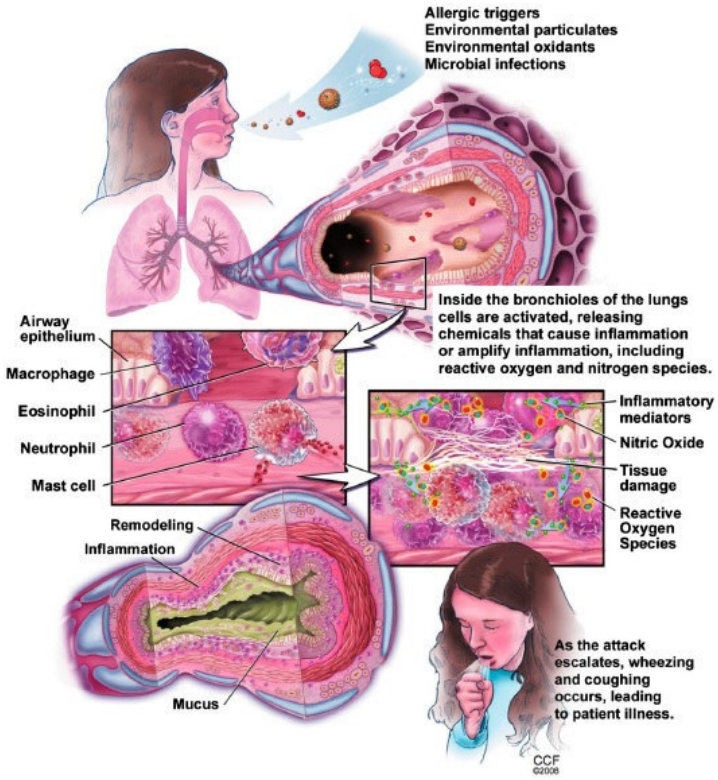


*Figure 1-4. **Asthma pathophysiology**. (11).*

Several studies show that oxidative stress is associated with lower lung function, such as forced expiratory volume in 1 second (FEV_1) or forced vital capacity (FVC). Thus an inverse association of a biomarker of lipid peroxidation, thiobarbituric acid-reactive substance (TBARS), with FEV_1 in a pilot group of nonsmoking subjects from general population has been found (14, 19). Larger cross-sectional study proved this finding, as well as showed inverse association of TBARS with FVC in a sample of men, suggesting gender differences in relations between oxidative stress and lung function (19).

1.2 Biomarkers of Oxidative Stress.

Biomarkers of oxidative stress are used to evaluate presence of oxidative stress in human body as well as to measure antioxidant capacity of organism *in vivo* (20).

The products of proteins, lipids and DNA oxidation by ROS provide us with a huge variety of potential biomarkers of oxidative stress (21).

1.2.1 Protein oxidation products as biomarkers of OS.

Oxidation of proteins leads to scission of protein backbones and to deamination of some amino acids, forming protein carbonyls – protein-derived biomarkers of OS *in vivo* (Figure 1-5) (22, 23).

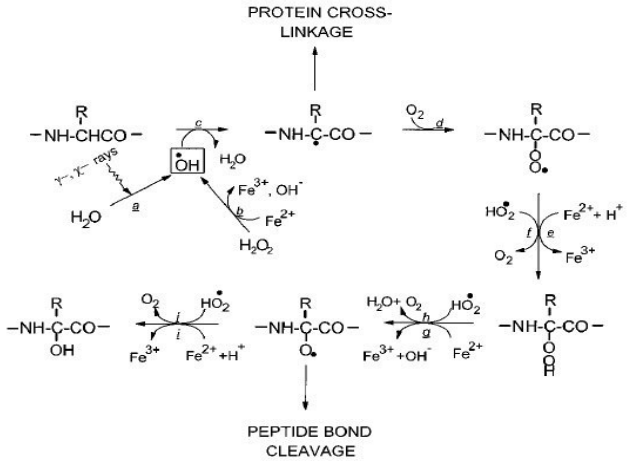


Figure 1-5. ROS-mediated oxidation of proteins (23).

The concentration of such protein degradation products is quite high because of different mechanisms of their formation. The list of amino-acids most exposed to oxidation is presented in the Table 1-1 (23).

Table 1-1. Amino acids exposed to oxidation and their products (23).

Amino acid	**Oxidation product**
Cysteine	Disulfides, cysteic acid
Methionine	Methionine sulfoxide, methionine sulfone
Tryptophan	2-, 4-, 5-, 6-, and 7-Hydroxytryptophan, nitrotryptophan, kynurenine, 3-hydroxykynurinine, formylkynurinine
Phenylalanine	2,3-Dihydroxyphenylalanine, 2-, 3-, and 4-hydroxyphenylalanine
Tyrosine	3,4-Dihydroxyphenylalanine, tyrosine-tyrosine cross-linkages, Tyr-O-Tyr, cross-linked nitrotyrosine
Histidine	2-Oxohistidine, asparagine, aspartic acid
Arginine	Glutamic semialdehyde
Lysine	a-Aminoadipic semialdehyde
Proline	2-Pyrrolidone, 4- and 5-hydroxyproline pyroglutamic acid, glutamic semialdehyde

Threonine	2-Amino-3-ketobutyric acid
Glutamyl	Oxalic acid, pyruvic acid

Normally the rate of intracellular protein oxidation should be balanced by the rate of oxidized proteins degradation, which depends on processes of ROS formation, as well as on proteases responsible for degradation of oxidized proteins (Figure 1-6) (23).

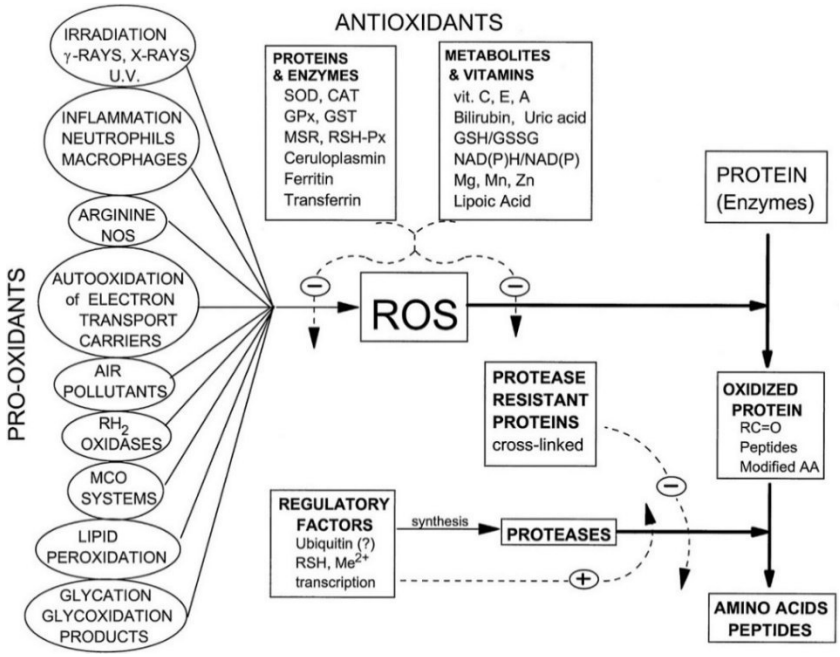


Figure 1-6. ***Pro-oxidant, oxidant and antioxidant balance in oxidized proteins formation.*** *MSR - methionine sulfoxide reductase; GPx - glutathione peroxidase; CAT - catalase; RSH-Px - thiol-specific peroxidase; NOS - nitric oxide synthetase; SOD - superoxide dismutase; GST - glutathione transferase (23).*

Usually the protein carbonyl content is measured in plasma, serum, cell lysates, or tissue homogenates, which makes this class of OS biomarkers less relevant in epidemiological studies.

1.2.2 Lipid oxidation products as biomarkers of OS.

Lipids, such as polyunsaturated fatty acids (PUFA), particularly linoleic and arachidonic acids (AA), and their esters are very susceptible to oxidation by ROS (20, 22). The products are conditioned by types of mechanism of peroxidation: enzymatic, non-enzymatic free radical and non-enzymatic non-free radical. Most investigated lipid peroxidation end products are such reactive aldehydes as 4-hydroxy-2-nonenal (4-HNE) and malondialdehyde (MDA); and products of radical mediated oxidation of AA - isoprostanes (IsoPs) (22).

4-HNE and MDA are known as biomarkers in a variety of chronic diseases (acute myocardial infarction, colon and breast cancer, ischemic stroke and post-stroke epilepsy, diabetes, etc.) (22). They are detectable in plasma cells and tissues with immunocyto- and immunohistochemistry. These methods require

individual sample elaborating, thus are time consuming and useless in routine clinical practice (22).

Isoprostanes are prostaglandin F_2-like compounds produced over the enzymatic free radical peroxidation of AA (Figure 1-7) (24). F_2-isoprostanes are recently developed biomarkers of oxidative stress, if compared with 4-HNE and MDA, but proved to be relevant biomarkers of OS because of their high sensitivity to OS, stability and detectability in urine (22).

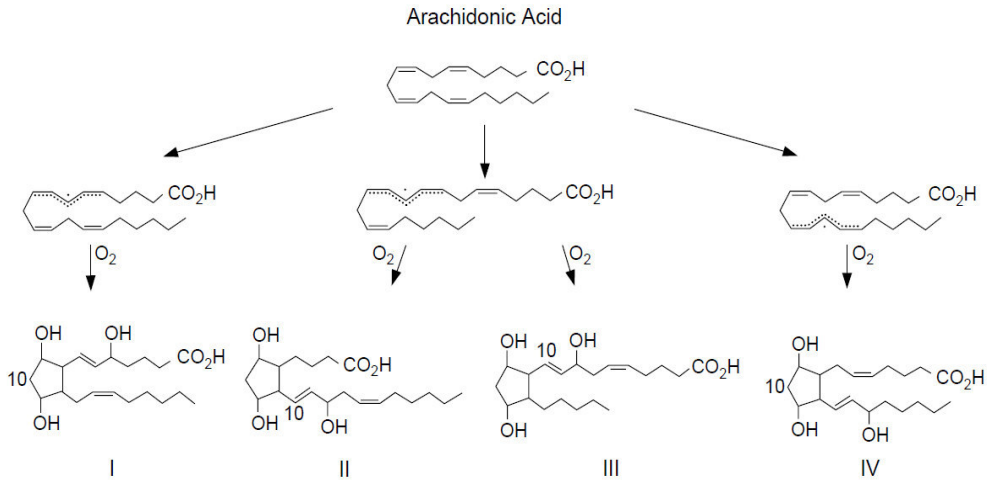


Figure 1-7. Peroxidation of arachidonic acid (AA) and formation of F_2-Isoprostanes (24). AA gets oxidized under the catalysis of free radicals, producing arachidonyl radical intermediates, which form then 4 isomers (I - IV) of F_2-like prostaglandins.

F_2-isoprostanes are chemically stable, are detectable in many biological fluids and tissues at quantitative amounts, do not depend on dietary lipids, and are stable and can be measured in urine which makes them reliable biomarkers of OS *in vivo* in epidemiological studies (25). Increased levels of F_2-isoprostanes are found in such pathologies as cardio-vascular disease, lung, liver, prostate cancers, age-related cognitive impairment, depression, cardiopulmonary bypass, hypertension etc. (22).

Isoprostanes are involved in the pathogenesis of respiratory diseases, such as COPD (26), which make them relevant biomarkers in the area of respiratory health. Oxidative stress leads to tissue damage, decrease of activity of antiproteases, hypersecretion of mucus, dysfunction of vascular barrier which in turn leads to edema of the bronchial wall, to bronchoconstriction by means of formation of isoprostanes from lipid peroxidation, and to higher lung inflammation through redox-sensitive transcription factors activation in leukocytes (Figure 1-8) (26).

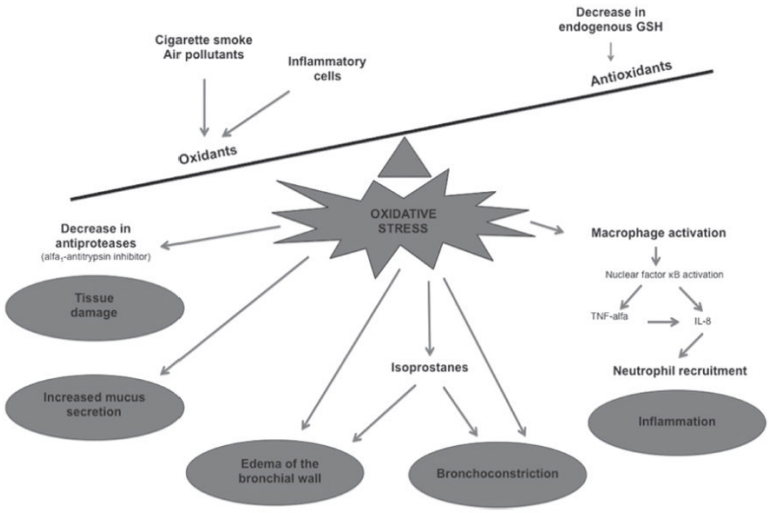
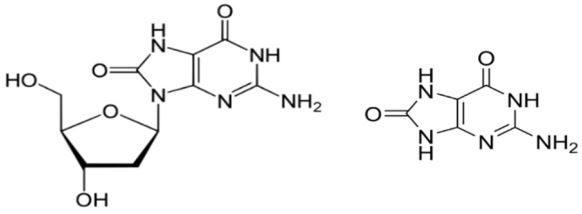


Figure 1-8. Oxidative stress pathways in COPD (26).

1.2.3 DNA oxidation products as biomarkers of OS.

There is a number of different endogenous and exogenous factors that lead to DNA damage. The most studied is the covalent modification of DNA, caused by ROS (27, 28). The investigation of DNA damage is mainly focused on identification of DNA lesion and its quantitative evaluation (27).

Out of more than 30 oxidative DNA degradation products, two of them are most-used as biomarkers of oxidative DNA degradation: 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodG) and its nuclease form 8-oxoguanine (8-oxoGua) (Figure 1-9) (27).



8-oxodG **8-oxoGua**

Figure 1-9. Main biomarkers of oxidative DNA damage: 8-oxodG and its nuclease form 8-oxoGua (27).

8-oxodG was discovered by Kasai and Nishimura in 1984, and was popular because of its abundance and easy detection. But recently it found a new birth because of its mutagenic potential, which rises from propensity to its mispairing with adenine while replicating, which leads to oxoG-A mispair (27). Direct oxidation of DNA and of 8-oxoguanine (GO) accumulated in DNA through the incorporation of 8-oxo-dGTP from the nucleotide pool leads to increased number of A:T to C:G or G:C to T:A transversion mutations after two rounds of replication. In the Figure 1-10 red line corresponds to mutagenic pathway. 8-oxo-

dGTP gets hydrolized by MTH1 to 8-oxo-dGMP and pyrophosphate in mammals defence systems. Then 8-oxo-dGMP is transformed to 8-oxodeozyguanosine (8-oxo-dG), which allows to avoid their incorporation into DNA Figure 1-10 (27, 29).

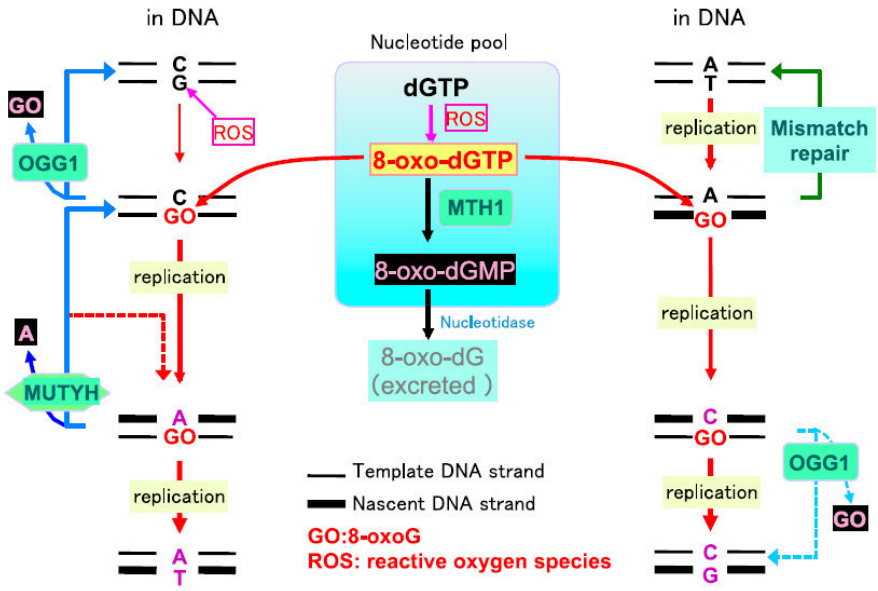


Figure 1-10. 8-oxodG excretion in mutagenic pathway in DNA (29). GO: 8-oxoG; OGG1: 8-Oxoguanine glycosylase; 8-oxo-dGTP: 8-Oxo-2'-deoxyguanosine-5'-Triphosphate; MTH1: the mammalian homolog of the Escherichia coli mutT gene which inhibits the incorporation of 8-oxodGTPase residues into DNA; 8-oxo-dGMP: 8-Oxo-2'-deoxyguanosine-5'-monophosphate; 8-oxo-dG: 8-oxo-7,8-dihydro-2'-deoxyguanosine; MUTYH: mutY DNA glycosylase.

The DNA glycosylase activity of OGG1 excises 8-oxoG opposite *cytosine (*Figure 1-10) mainly. OGG1 has a weak DNA-apurinic site lyase activity also. The human OGG1 gene is located on chromosome 3p25, a locus frequently lost in patients with lung cancer (29).

The ROS-induced damage can occur both in DNA and RNA. However, the mechanisms of their synthesis, maintenance and decay differ (Figure 1-11) (30). DNA integrity depends on its repair mechanisms, while RNA one on its surveillance and breakdown. If DNA integrity fails, the GC:TA transversion mutation takes place. RNA integrity failure induces formation of truncated or mutated proteins (Figure 1-11) (30).

All these factors induced a rise in development of new techniques for detection of 8-oxodG in cellular DNA, biological fluids, tissues, exhaled breath condensate and urine, such techniques as sensitive high pressure liquid chromatography (HPLC), Liquid Chromatography-Mass Spectrometry (LC-MS), enzyme-linked immunosorbent assay (ELISA) etc. (27, 31).

The major part of DNA damage investigations was dedicated to guanine oxidation (30). The main advantages of biomarkers derived from this oxidation

are that they are excreted in urine, quite stable in it, and can be easy detected with ELISA and HPLC methods. Both approaches are validated by European Standards Committee on Oxidative DNA Damage (ESCODD) (30). Nowadays the methodological gold standard is HPLC; however, in epidemiological studies the use of ELISA is appreciated, as far as it is a less costly and time saving method. 24-hour urine collection is the recommended procedure, but in large epidemiological surveys it is unpractical and morning spot creatinine-standardized samples are used instead (32).

It has been shown in several studies that OS can be a consequence of inflammatory response in respiratory disorders such as asthma, COPD, chronic bronchitis etc. (15, 33). The most studied biomarker of DNA damage in this condition is 8-oxodG, due to its abundance, ease of measurement and mutagenic potential. However, there is still lack of big epidemiological surveys conducted to provide an evidence of the 8-oxodG potential as a biomarker of OS in respiratory diseases.

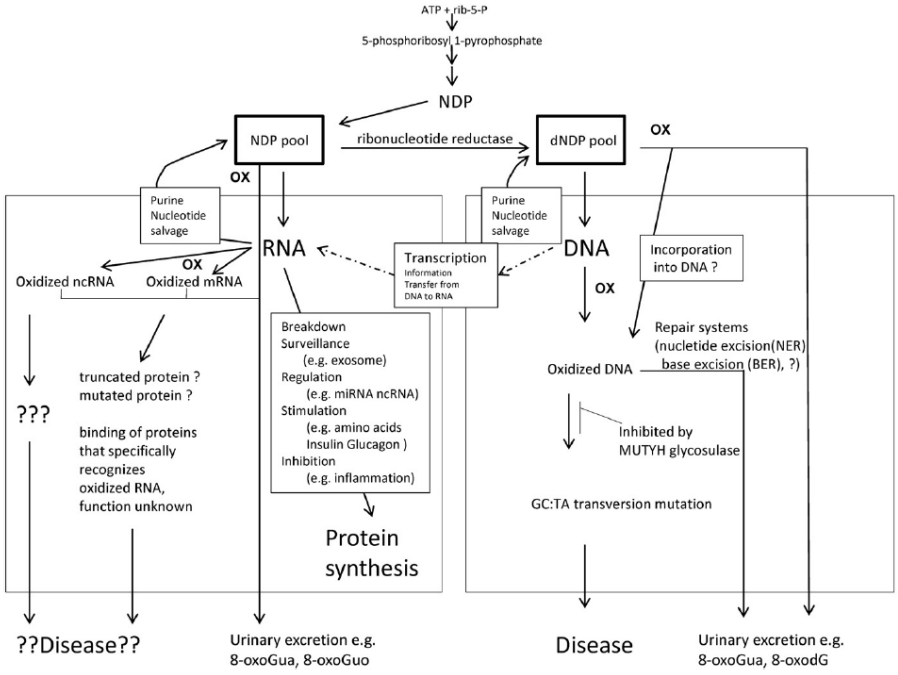


Figure 1-11. RNA and DNA potential sites of oxidation (OX). ATP: adenosine triphosphate; rib-5-P: ribulose 5-phosphate; NDP: nucleoside-diphosphate; dNDP: deoxynucleoside diphosphate; RNA: ribonucleic acid; DNA: deoxyribonucleic acid; ncRNA: non-coding RNA; mRNA: messenger RNA; miRNA: microRNA; NER: nucleotide excision repair; BER: base excision repair; MUTYH: mutY DNA glycosylase; GC:TA (GuanineCytosine:ThymineAdenine) (30).

1.3 Reference Intervals.

In medicine Reference Intervals (RI) are used to distinguish between healthy and probably diseased individuals. They provide a range of values of biochemical parameter, which characterizes a sample of healthy subjects with

similar age, sex, race, etc. (34). Sometimes RI are called Reference or Normal Ranges. However, these definitions are less accurate as reference ranges can refer to the entire range (100%) of normal values of the reference population (35), and the definition "Normal" can also refer to the normal (Gaussian) distribution of the observed data. Thus, in this work the definition 'Reference Interval' has been used.

Reference intervals represent two limits of the distribution range of the data under examination. The most-used in medical practice is a 95% RI, with lower (2.5%) and upper (97.5%) limits (34).

There are non-adjusted and adjusted for determinants (regression-based) methods to estimate RI.

1.3.1 Non-adjusted reference intervals.

Non-adjusted reference intervals refer to the direct evaluation of reference intervals from the empirical data set. If the data (measurements of biochemical in the human body) follow the normal distribution and there is an appropriate number of subjects in the reference group, the parametric approach can be used. Thus, 95% RI can be calculated as a population mean $\mu \pm 1.96$ standard deviations (SD), often taken as 2SD or 2σ (36). However, if the normal distribution or the transformation to normality fail, non-parametric approaches should be applied to the data set (34, 36). Non-parametric methods do not make any assumption on the distribution of the data. There is a straightforward approach in estimating 95% centile RI. In the ordered data the lower 2.5% and upper 97.5% limits get calculated directly (34). However, the number of observations in this method should be 120 or over, otherwise the RI calculated can be influenced by outliers (34).

1.3.2 Adjusted (regression-based) reference intervals.

Regression analysis allows evaluating associations between response variable and its explanatory variables. Regression-based RI can be estimated with (semi)parametric and non-parametric methods.

If regression residuals follow normal distribution, parametric regression analysis can be used to estimate adjusted RI. The linear regression analysis is a most-used way to estimate adjusted parametric RI (37,38). If one of main assumptions of linear regression as *linearity*, *normality of residuals distribution*, *homoscedasticity, independence of errors* is not met, non-parametric or semi-parametric regression analysis can be used (39).

Estimating RI, based on mean of the population is not always reasonable because of the skewed data and because of the possible outliers in the data set (37). Quantile regression analysis provides huge advantage as addresses this issue;

allows predicting regression coefficients at the quantiles of interest, e.g. 2.5 and 97.5 percentiles to calculate 95% RI; and allows building centile (growth) curves (38, 39).

The alternative regression models can be used to address an issue of the skewed or kurtotic data. Generalized Additive Models for Location, Scale and Shape (GAMLSS), developed by Professors M. Stasinopoulos and R. Rigby, address these issues as allow building huge number (~ 80) of distribution families for outcome variable, dealing with skewed and kurtotic data as well (40). The GAMLSS is parametric method as it still requires parametric distribution function of the response variable, and is semi-parametric because non-parametric smoothing functions can be involved in the modelling of the parameters (41).

1.3.3 Reference intervals of oxidative stress biomarkers (8-oxodg and 8-isoprostane) in literature.

Both biomarkers (8-oxodG and 8-isoprostane) can be detected in different biological matrices: urine, plasma, cerebrospinal fluid and in exhaled breath condensate (30, 42). The 8-oxodG is one out of more than 30 oxidatively modified bases of DNA, and is selected as the main biomarker of oxidative stress because of its higher stability in urine over time (27, 32, 43). The global oxidative stress in human body is measured through urinary excretion of biomarkers of OS, as this method reflects the average rate of the entire oxidation of organism (30). The urinary excretion of 8-oxodG represents global oxidative stress, to DNA, and can be defined as general oxidative stress (30). As in case of 8-oxodG, 8-isoprostane is detected in urine because of its short-life in plasma and relevant stability in urine (10). Urinary detection of biomarkers is relatively inexpensive, easily collected and non-invasive, which makes this approach widespread (32, 44).

There are chemical (e.g. HPLC – high pressure liquid chromatography, with electrochemical or mass spectrometry detection methodologies, etc.) and biological (e.g. ELISA, Comet assay etc.) techniques for examination of urinary biomarkers of oxidative stress, such as 8-oxodG and 8-isoprostane (30). Nowadays the methodological gold standard is HPLC measurement of OS biomarkers in 24h urine samples. However, in large epidemiological studies, it is quite complicated to collect 24h urine and the use of 'spot' morning urine samples is feasible if biomarker measures are corrected by creatinine, as they are correlated with 24h measures ($r=0.60$) and have decreased intra-individual variability (30, 32, 45). The ELISA method is appreciated in large epidemiological studies due to its lower cost, lower time consuming and ease of reproduction in comparison with HPLC approach (44).

To our knowledge, there are still no works dedicated to investigation of adjusted *regression-based* reference intervals of 8-oxodG and 8-isoprostane.

Methodological approaches of detecting non-adjusted reference intervals are described in the paragraph 1.3.1. There is a strong homogeneity around works dedicated to 8-oxodG and 8-isoprostane RI detection. All of them were independent of explanatory covariates (non-adjusted reference intervals). This homogeneity can be considered as a disadvantage or a lack of variety of methodological approaches in 8-oxodG and 8-isoprostane RI determination, because different authors estimated RI of 8-oxodG (1, 10, 28, 46-48) or 8-isoprostane (10, 46) in subsamples stratified by age or gender, and some of them showed statistically significant difference in RI levels (most often using just a t-test or its non-parametric equivalent Mann-Whitney U test), while others instead did not find it. The investigation of RI of 8-oxodG and 8-isoprostane in age- and gender- related subsamples are most-studied in literature and the variability in age- and gender RI of OS biomarkers make them relevant demographic covariates to be included in regression models to check their joint influence on both biomarkers, 8-oxodG and 8-isoprostane, and in the presence of other potential determinants.

Epidemiological studies are usually spread in time and, although the ELISA method has lower cost in comparison with HPLC detection, it is still quite costly. The ELISA kit gets used when there are urine samples from 90 subjects. It can take a long time from the moment of first patient sample up to last (90th) patient urine sample gathering (especially in epidemiological studies) essential for opening the ELISA kit. Moreover, not all laboratories in hospitals have educated staff to perform enzyme-linked immunosorbent assays like ELISA for 8-oxodG or in order to avoid inter-laboratory variability in measurements; they just collect urine samples from patients, conserve for unspecified time and then send them all together to a single laboratory to be examined, which can be located in different cities. Obviously, such a process can take a long time, i.e. time from the moment of urine collection to its laboratory elaboration. The correction for such variable as DFC of 8-oxodG and 8-isoprostane values can help avoiding the problem of inter-individual variability in these biomarkers, because different results can be given not by health condition of each individual but simply just by the different time passed from the moment of urine gathering and its analysis.

In our collaboration with medical practitioners who deal with 8-oxodG and 8-isoprostane levels determination, we were asked to consider such environmental factor as season when urine was collected, that can influence OS biomarkers values due to seasonal differences in air temperatures and pollution (49, 50).

In this work, we aimed at identifying main demographic, environmental and laboratory determinants of urinary biomarkers of OS (8-oxodG and 8-isoprostane), and at defining their reference intervals, adjusted for main determinants, in a sample of healthy people from the general Italian population.

2 METHODS.

2.1 Study Design and the population.

The Gene Environment Interactions in Respiratory Diseases (GEIRD) project is an Italian multi centre multi case-control study which (51) aimed at building multicenter database and biobank according to phenotypes of asthma, allergic rhinitis and COPD, and in conformity with their inflammatory and genetic profiles and risk factors (52). A part of GEIRD is nested in ECRHS (European Community Respiratory Health Survey) study and data from ECRHS cohort are already available to ECRHS epidemiologists. To have an access to entire GEIRD data, researchers from around the world can contact GEIRD steering Committee to collaborate (www.geird.org) (51, 52).

This project was developed to cover a lack of large-scale epidemiological, genetic investigations in the field of three main respiratory disorders such as asthma, COPD and rhinitis. It was urged to collect information on biomarkers of inflammation and oxidative stress, individual and ecological exposures, diet, early-life factors, smoking habits, genetic traits and medication use in large and precisely defined cases of asthma, allergic rhinitis and COPD (52). The GEIRD involved 8 Italian research centres (Ancona, Palermo, Pavia, Salerno, Sassari, Terni, Turin, and Verona as a coordinating centre) (51). The study design included two main stages: screening phase, and case-control selection and phenotyping.

In the first stage new random samples or pre-existing randomly sampled Italian ECRHS cohorts from the general population (20-64 years of age, male/female = 1/1) were mailed the respiratory symptoms questionnaire (53, 54). Over the following stage 2, all recipients reporting symptoms relating to asthma, chronic obstructive pulmonary disease (COPD) or chronic bronchitis (CB), and a random sample of people with rhinitis and with no respiratory symptoms were invited to undergo a detailed clinical interview and lung function tests for accurate phenotyping, and to provide blood and urine samples (53, 54). The consent of each participant, as well as ethical committee approval have been received for all centres.

In this work, data from 4 Italian centres, i.e. Verona, Pavia, Turin, and Sassari, collected between 2007 and 2013 years have been analyzed. Out of 16569 people, who were selected to participate in GEIRD stage 1, 9741 (59%) answered to the screening questionnaire. Of these, 4981 (51%) were selected to attend GEIRD stage 2 and 2259 (45%) participated in the clinical survey (Figure 2-1).

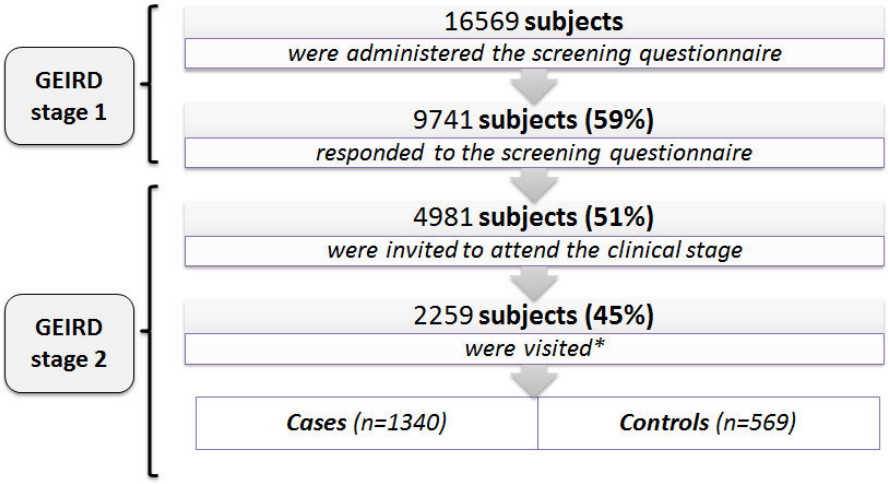


*Figure 2-1. Number of participants in GEIRD study of Verona, Pavia, Turin and Sassari. *Data available for 1909 subjects. Cases included asthma, rhinitis, COPD and chronic bronchitis subjects. Controls at this point are subjects with no respiratory diseases.*

2.2 Clinical and Laboratory Measurements.

2.2.1 Lung Function and Allergologic Tests.

The forced spirometry was applied to subjects in accordance with the American Thoracic Society reproducibility criteria (54, 55). In accordance with Quanjer PH et al., the FEV_1 % predicted (Forced expiratory volume in the 1st second), the LLN (lower limit of normal), and the FEV_1/FVC (forced vital capacity) ratio (Tiffeneau-Pinelli index) have been calculated (54, 56). If prebronchodilator FEV_1/FVC > LLN and >70% and FEV_1> 70% predicted, and skin prick tests did not cause a reaction, subjects were recognized as controls. This group of controls with further inclusion criteria (Figure 2-2) has been used in current analysis.

2.2.2 Urine collection.

Participants were asked to collect their first urine of the morning (on awakening) in a clean container, not necessarily sterile. The container had a capacity to contain at least 10 ml of urine. If the patient had not brought their own urine, the collection was performed in clinic filled in questionnaire with the time of sampling, the number of cigarettes smoked and medicines taken from the moment of awakening (if a case) (53).

The container with the urine sample was stored at 4°C for up to 24 hours. Then, from the sample were derived 4 rates 1 ml each for subsequent analysis. On each rate was applied the code of the centre and the code of identification of the subject. The rates stored in freezer (-80^{0}C). The storage at this temperature can last for a maximum of one year (53).

2.2.3 Biomarkers evaluation.

The 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodG) and the 8-isoprostane, both standardized by creatinine (ng/mg) (32), were evaluated in spot urine samples with the immunosorbent assay kit ELISA (Cosmo Bio LTD, Tokyo Japan, and Cayman Chemical, Ann Arbor, MI, USA, respectively). In the assessment of creatinine concentration (mg/ml) another ELISA kit has been used (Cayman Chemical, Ann Arbor, MI, USA).

2.2.4 Subjects

For the current analysis, the sample containing all never-and ex- (not smoking over the last year) smokers, who did not report respiratory symptoms at the screening questionnaire and at the clinical survey, either other comorbidities (heart disease, ictus, high blood pressure, diabetes and cancer) at the clinical interview and with normal lung function and allergologic test has been used. Overall 281 subjects have been investigated (Figure 2-2).

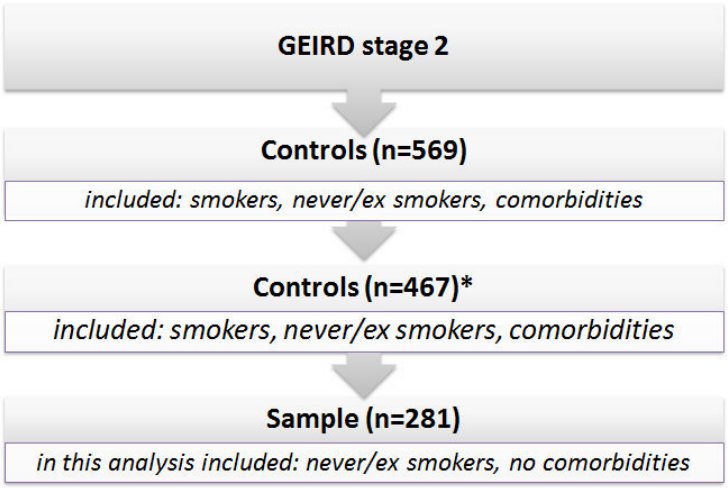


*Figure 2-2. Inclusion criteria in the sample for current analysis. *Data on biomarkers of oxidative stress are available for 467 subjects.*

2.3 Statistical analyses.

The descriptive statistical analysis was used to define main characteristics of the studied sample. The Shapiro-Wilk normality test was used to check the distribution of both biomarkers (8-oxodG and 8-isoprostane). The logarithmic transformation was applied to both biomarkers to normalize the data.

2.3.1 Non-adjusted reference intervals determination.

The parametric approach was used first to define the non-adjusted 95% RI. To the log-transformed data of 8-oxodG and 8-isoprostane the following formula

has been applied:

$$\mathbf{RI} = \bar{\mathbf{x}} \pm \mathbf{z} * \mathbf{SD},$$

where $\bar{x}$ is a sample mean, z-score=1.96; SD is a sample standard deviation.

The results were back transformed into normal values within exponential function (34, 36).

To evaluate non-parametric non-adjusted reference intervals of 8-oxodG and of 8-isoprostane, the empirical data were ordered and 2.5% and 97.5% limits were evaluated directly to produce the 95% RI (34).

2.3.2 Assessing main determinants of 8-oxodG and 8-isoprostane and adjusted RI estimation.

Such subject- , laboratory- , and collection-related explanatory variables as age, gender, DFC (distance from collection - the period from the moment of urine collection and its laboratory processing) and season (the period of year split into warm (*April - September*) and cold (*October - March*)) have been evaluated in the current study.

Univariable and multivariable regression models were used to define main explanatory variables of 8-oxodG and 8-isoprostane.

Linear Regression.

Parametric univariable linear regression analysis within log-transformed data of 8-oxodG was used first to define unadjusted associations with potential determinants (age, gender, DFC and season). The associations were then analyzed in multivariable model adjusted for four potential determinants simultaneously.

The interaction term between DFC and season was tested to check the effect modification on 8-oxodG distribution.

The linear regression estimates of log(8-oxodG) on DFC have been analyzed in warm and cold seasons.

Bootstrapping Quantile Regression.

The non-parametric bootstrapping quantile regression (QR) analysis (57, 58) was used to analyze associations between age, gender, DFC, season and biomarkers of OS.

The median (50% quantile) QR was used for estimating central location changes in distribution of a response variable (59). As in case of linear regression, where fitted coefficients are interpreted as estimates of the change in the mean of the response distribution resulting from a one-unit increase in a continuous

covariate or the change of the value from 0 to 1of a binary covariate, similarly the quantile regression estimates are the changes in specified quantiles (e.g. 0.025, 0.05, 0.5 (median), 0.975, 0.99) of the response variable provided by a one unit change in the continuous explanatory variable or the change of the value from 0 to 1of a dummy covariate. (59). The conditional quantiles can be used to evaluate how an explanatory variable predicts the conditional off-central location in the distribution of the response variable (59).

In case of 95% reference interval estimation, we were interested in the lower (0.025 quantile) and upper (0.975 quantile) distribution changes of 8-oxodG and 8-isoprostane on predictors (age, gender, DFC and season) change. The median (0.5 quantile) regression has been used for central tendency estimates.

The unconditional associations between subject-related (age and gender), collection-related (season), and laboratory-related (DFC) explanatory variables and outcome variables, 8-oxodG and 8-isoprostane, were examined using univariable bootstrapping quantile regression models (57, 58).

The multivariable bootstrapping quantile regression analysis adjusted for age, gender, season and DFC was performed to evaluate association of the explanatory variables with outcome biomarkers 8-oxodG and 8-isoprostane. Independent variables were inserted in the model in accordance with their values of statistical significance in univariable analysis, or their biologic plausibility. Age was inserted first in the model as far as it is considered that oxidative stress is associated with aging (60) and was followed by gender, another important demographic variable well-described in literature regarding 8-oxodG and 8-isoprostane (1, 10, 28, 56). Afterwards were added DFC (32, 61, 62) and season (49, 50).

The covariates that did not provide statistical significance ($p > 0.05$) have been excluded from further analyses.

Multivariable median bootstrapping QR models with interaction term between DFC and season have been checked as well. The stratified by season multivariable 2.5% and 97.5% bootstrapping QR models were used then to repeat all the analyses in subsamples.

500 bootstrap replications were used in all QR models.

Descriptive statistical analysis linear and quantile regression analyses have been performed with a STATA/IC 14.0 for Windows.

GAMLSS Regression.

The (semi)parametric GAMLSS in R (Version 0.99.902 – © 2009-2016 RStudio, Inc.) regression analysis was used to analyze the data on 8-oxodG and 8-

isoprostane (40, 63).

Generalized additive models for location, scale and shape (GAMLSS) are an extension of the generalized linear models, generalized additive models and of generalized linear mixed models as they allow modelling both skewness and kurtosis for the response variable (Y) and provide linear predictors for the distribution parameters: μ (location, mean, median, mu), σ (scale, variability, sigma), ν (shape, skewness, nu) and τ (shape, kurtosis, tau) (64). It is possible to fit with GAMLSS both the additive and the multiplicative models using identity or log links respectively (65).

The GAMLSS analysis is based on LMS (lambda, mu, sigma) approach which is widely used in constructing growth reference charts (66). LMS includes three parameters: the skewness (lambda) serves for modelling deviations of variables from normality; the median (mu) explains the change of outcome variable with an explanatory variable; and the coefficient of variation (sigma) models the dissemination of values around the mean value (66).

The principle of LMS method is that the distribution of the outcome variable (y) depends on varying parameters of explanatory variable, i.e. on λ, μ, and σ such that the transformed outcome:

$$z = \frac{(Y/\mu)^{\lambda} - 1}{\lambda \times \sigma}$$

is a z-score with a distribution N(0,1) (64, 65).

Rigby and Stasinopoulos proposed LMST (lamda, mu, sigma, tau) method within the GAMLSS regression analysis, where T (tau) is a kurtosis – *t* distribution degrees of freedom (67). There is a huge variety (~80) of different distributions that can be modelled within the GAMLSS. Some of them, like normal (NO) or gamma (GA), allow modelling just two parameters, i.e. mean (μ, or median, which depends on the distribution chosen) and coefficient of variation (σ); others, like Box-Cox-Cole and Green (BCCG) or power exponential (PE) model three parameters, i.e. mean (μ, or median), coefficient of variation (σ) and skewness (ν); and others, like Box-Cox-*t* (BCT) or Box-Cox power exponential (BCPE), can model all four parameters, i.e. mean (μ, or median), coefficient of variation (σ), skewness (ν), and kurtosis (τ). The list of some distribution forms within the **gamlss** package is described in the Table 2-1.

Table 2-1 Continuous distributions implemented within the gamlss package in R.

Distributions	**R Name**	**μ**	**σ**	**ν**	**τ**
beta	BE()	logit	logit	-	-
beta inflated (at 0)	BEOI()	logit	log	logit	-
beta inflated (at 1)	BEZI()	logit	log	logit	-
beta inflated (at 0 and 1)	BEINF()	logit	logit	log	log
Box-Cox Cole and Green	BCCG()	identity	log	identity	-
Box-Cox power exponential	BCPE()	identity	log	identity	log
Box-Cox-*t*	BCT()	identity	log	identity	log
exponential	EXP()	log	-	-	-
exponential Gaussian	exGAUS()	identity	log	log	-
exponential gen. beta type 2	EGB2()	identity	identity	log	log
gamma	GA()	log	log	-	-
generalized beta type 1	GB1()	logit	logit	log	log
generalized beta type 2	GB2()	log	identity	log	log
generalized gamma	GG()	log	log	identity	-
generalized inverse Gaussian	GIG()	log	log	identity	-
generalized y	GT()	identity	log	log	log
Gumbel	GU()	identity	log	-	-
inverse Gaussian	IG()	log	log	-	-
Johnson's SU (μ the mean)	JSU()	identity	log	identity	log
Johnson's original SU	JSUo()	identity	log	identity	log
logistic	LO()	identity	log	-	-
log normal	LOGNO()	log	log	-	-
log normal (Box-Cox)	LNO()	log	log	fixed	-
NET	NET()	identity	log	fixed	fixed
normal	NO()	identity	log	-	-
normal family	NOF()	identity	log	identity	-
power exponential	PE()	identity	log	log	-
reverse Gumbel	RG()	identity	log	-	-
skew power exponential type 1	SEP1()	identity	log	identity	log
skew power exponential type 2	SEP2()	identity	log	identity	log
skew power exponential type 3	SEP3()	identity	log	log	log
skew power exponential type 4	SEP4()	identity	log	log	log
shash	SHASH()	identity	log	log	log
skew *t* type 1	ST1()	identity	log	identity	log
skew *t* type 2	ST2()	identity	log	identity	log
skew *t* type 3	ST3()	identity	log	log	log
skew *t* type 4	ST4()	identity	log	log	log
skew *t* type 5	ST5()	identity	log	identity	log
t Family	TF()	identity	log	log	-
Weibull	WEI()	log	log	-	-
Weibull (PH)	WEI2()	log	log	-	-
Weibull (μ the mean)	WEI3()	log	log	-	-
zero adjusted IG	ZAIG()	log	log	logit	-

The link functions (mu.link, sigma.link, nu.link and tau.link) for each distribution family, presented in this table, help the fitting procedure. The default link functions can be usual glm() functions plus logshifted, logitshifted and own (allows to define own link function) (41).

Each parameter (μ, σ, ν, τ) of a GAMLSS distribution is modelled through:

$$g_k(\theta_k) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\,\gamma_{jk},$$

where $\boldsymbol{g_k}$ is the link function, $\boldsymbol{\theta_k}$ is one of the modelled parameters μ, σ, ν or τ, $\boldsymbol{\eta_k}$ is the linear predictor, $\boldsymbol{X_k\beta_k}$ are linear terms, $\sum_{j=1}^{J_k} \boldsymbol{Z_{jk}\gamma_{jk}}$ are additive terms (41, 65). Such additive terms like cubic smoothing spline (*cs*), or penalized B- spline allow modelling complex effects (nonlinear relationships) of explanatory variable on the dependent variable (64). *cs* allows the dependent variable varying smoothly (non-linearly) as a function of an explanatory variable (64, 68). The complexity of the spline curve is expressed through effective degrees of freedom (edf), the lowest edf=0 corresponds to linear term, while higher edf correspond to more complex curve (68).

The choice of the best fitted model involves minimizing Bayesian Information Criterion (BIC= -2 ln $f(y|\widehat{\theta_k})$ + k ln n, where $f(y|\widehat{\theta_k})$ is a candidate or approximating model, k is the number of free parameters to be estimated, and n is the sample size) (68, 69).

The modelling within **gamlss** consists of the following steps (65):

1) Are all explanatory variables (e.g. age, gender, DFC, season) required in the model?

2) Is the relationship of chosen covariates additive or multiplicative—should their effects be added together or multiplied together?

3) What is the best way to model them: linear or log?

4) Is the chosen continuous variable effect linear, quadratic or more complex?

5) Which is the distribution of the residuals, and does it vary with explanatory variables?

The development of the GAMLSS regression model for 8-oxodG and 8-isoprostane was done using 4 cofactors: age, gender, DFC and season, to identify which of them should be included in the final model. Three of the most tried and tested distribution families have been used: NO (normal,), BCCG (Box-Cox Cole and Green) and BCT (Box-Cox-*t*) in the entire sample (41, 63, 65, 70).

The development of the model was performed first in the framework of each distribution family, using logarithmic or identity link functions for mean

(median) μ, variability σ, skewness ϑ, and kurtosis τ of continuous independent variables (age and DFC). The μ functions of response variables (8-oxodG and 8-isoprostane) and of continuous explanatory variables (age, DFC) were modified using logarithmic transformation to check whether this could improve the model fitting. Minimizing the Bayesian Information Criterion (BIC) (69) and in accordance with normalized residual distribution, the best model was identified for each distribution family. Final choice in between models with different distribution families was made in accordance with same criteria (64). The models that fitted the data worse have been excluded from further analyses.

The multivariable GAMLSS regression was performed then using the best fitted models for each biomarker with the interaction term between DFC and season.

When stratifying by season, GAMLSS models adjusted for DFC were developed in subsamples of 8-oxodG and 8-isoprostane. The new development of models was performed first in the framework of each distribution family, i.e. BCCG, BCT, GA (gamma), LOGNO (log-normal), applying polynomials, cubic and penalized B-splines to mean (median) μ, variability σ, skewness ϑ, and kurtosis τ parameters of DFC (63, 65). The best model for each biomarker was chosen minimizing the Bayesian Information Criterion (BIC) and in accordance with normalized residual distribution, biologically plausible centile curves (decreasing levels of OS biomarkers with increasing DFC) and significance of link functions (in the variety of models built for each biomarker, it would be preferable to choose the model with statistically significant median, variability, and skewness link functions). BIC criterion is the main measure to consider an evidence against the model *i*. It is considered powerful with ΔBIC > 2 (69):

BIC_i - BIC_{min}	Evidence Against Model *i*
0 - 2	Not worth more than a bare mention
2 – 6	Positive
6 – 10	Strong
>10	Very strong

If ΔBIC is weak (<2) other parameters of fitted models should be considered to make a decision about the better model, i.e. residuals distribution, degrees of freedom (df) used to fit the model (the less df were used the simpler and the better model is), and finally centile curves and significance of link functions.

Outliers for each biomarker were defined using Influence “bubble” plots (71, 72). The “bubble” plots represent Studentized residuals by hat values (most common measure of leverage, which shows how far away the independent variable values of an observation are from those of the other observations), with the areas of the circles representing the observations proportional to Cook's

distances (Cook's D, is an estimate on how a data point can influence the data when performing a regression analysis, the larger the area of the Cook's D circle is, the higher the influence of the point is). Vertical reference red lines are drawn at twice and three times the average hat value, horizontal reference lines at -2, 0, and 2 on the Studentized-residual scale (71, 72). Most influential observations on such a graph have high absolute values of Studentized residuals, high hat values and large Cook's D area.

The best developed GAMLSS models for each biomarker per each season were used then to build models without extreme values in accordance with influential plots for outliers. Then the best developed models and models without extreme values built on their basis were compared with models containing cubic splines for μ link function of DFC, and with models without outliers and containing cubic splines for μ link function of DFC, to check which of them fit the data better in each case.

Reference intervals of 8-oxodG and 8-isoprostane as functions of DFC were determined with best GAMLSS regression model for each biomarker in two seasons, warm and cold, per each two weeks of DFC.

GAMLSS Distribution Families used in current analyses.

Normal (or Gaussian) distribution (NO) (63).

The normal (Gaussian) probability density function (pdf), denoted in GAMLSS as NO(μ,σ), is

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

Equation 2-1,

where for $-\infty < y < \infty$, mean of Y is given by $E(Y) = \mu$ and the standard deviation of Y by $Var(Y) = \sigma^2$, and $-\infty < \mu < \infty$ and $\sigma > 0$.

Log Normal distribution (LOGNO) (63).

The log-normal distribution is used in case of positively skewed data. The pdf of LOGNO is

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left\{-\frac{[\log(y)-\mu]^2}{2\sigma^2}\right\}$$

Equation 2-2,

where for $y > 0$, $\mu > 0$, $\sigma > 0$ and $w = \exp(\sigma^2)$, $E(Y) = \omega^{1/2}e^{\mu}$ and $Var(Y) = \omega(\omega-1)e^{2\mu}$.

Gamma distribution (GA) (63).

The gamma distribution is appropriate for positively skewed data. The pdf of the gamma distribution is

$$f_Y(y|\mu,\sigma) = \frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1}\, e^{-y/(\sigma^2\mu)}}{\Gamma(1/\sigma^2)}$$

Equation 2-3,

where for $y > 0$, $\mu > 0$, $\sigma > 0$, $\sigma^2 = 1/\alpha$ and $\mu = \alpha\beta$, $E(Y) = \mu$ and $Var(Y) = \mu^2\sigma^2$.

Box-Cox Cole and Green distribution (BCCG) (63).

The Box-Cox Cole and Green distribution, BCCG(μ,σ,ν) is appropriate both for positively and negatively skewed data.

For positive random variable Y>0 transformed trough random variable Z with standard normal distribution:

$$Z = \begin{cases} \frac{1}{\sigma\nu}\left[\left(\frac{Y}{\mu}\right)^{\nu} - 1\right], & \text{if } \nu \neq 0 \\ \frac{1}{\sigma}\log\left(\frac{Y}{\mu}\right), & \text{if } \nu = 0 \end{cases}$$

Equation 2-4,

where for $0 < Y < \infty$, $\mu > 0$, $\sigma > 0$ and $-\infty < \nu < \infty$, the conditions

$-1/(\sigma\nu) < Z < \infty$ if $\nu > 0$ and

$-\infty < Z < -1/(\sigma\nu)$ if $\nu < 0$

are required.

The pdf of Y in BCCG is

$$f_Y(y) = \frac{y^{\nu-1}\exp(-\frac{1}{2}z^2)}{\mu^{\nu}\sigma\sqrt{2\pi}\Phi(\frac{1}{\sigma|\nu|})}$$

Equation 2-5,

where Z is given by Equation 2-4 and Φ() is the cumulative distribution function (cdf) of a standard normal distribution.

Box-Cox t distribution (BCT) (63, 70).

For positive random variable Y>0 with BCT(μ,σ,ν,τ) transformed trough random variable Z with standard normal distribution (Equation 2-4), Z follows truncated t distribution with $\tau>0$ degrees of freedom and treated as a continuous parameter.

Thus for Y>0 with BCT(μ,σ,ν,τ), pdf is given by:

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{y^{\nu-1} f_T(z)}{\mu^{\nu}\sigma F_T(\frac{1}{\sigma|\nu|})}$$

Equation 2-6,

where for $y > 0$, $\mu > 0$, $\sigma > 0$ and $-\infty < \nu < \infty$, and where z is given by (Equation 2-4) and $f_T(t)$ and $F_T(t)$ are the pdf and cdf respectively of a random variable T with a standard t distribution with degrees of freedom parameter $\tau > 0$, i.e. $T \sim t_T = TF(0,1,\tau)$.

3 RESULTS.

3.1 Characteristics of the sample.

The sample of ex/no-smokers with no respiratory diseases and comorbidities contained the data on 281 subjects. Out of them were 275 observations of the 8-oxodG biomarker and 227 observations of 8-isoprostane biomarker (Table 3-1). The mean age was 44.5 years and 57% of them were females. The median DFC was 387 days and the urine was collected from half of the subjects in the warm season (Table 3-1).

Table 3-1. Characteristics of the sample.

	N	MEDIAN (IQR)	MEAN(SD)
RESPONSE VARIABLES			
8-oxodG, ng/mg$_{creat}$	275*	3.89 (1.91–7.95)	6.75 (10.60)
8-isoprostane, ng/mg$_{creat}$	227*	0.60 (0.24–1.57)	1.09 (1.30)
EXPLANATORY VARIABLES			
Age, years	281	44.80 (38.5–51.1)	44.5 (9.23)
DFC, days	255*	387 (316–611)	433.6 (190.9)
Gender, females (%)	161(57%)	-	-
Season, warm (%)	127(50%)*	-	-

**Missing data: 8-oxodG N= 6, 8-isoprostane N=54, DFC and Season N= 26. DFC – Distance From Collection - the period from the moment of urine collection and its laboratory processing; Season – warm (April - September), cold (October - March)*

Both biomarkers showed highly skewed frequency distributions (Figure 3-1) (A – 8-oxodG; B – 8-isoprostane), that was also confirmed with Shapiro-Wilk test for normality ($p<0.001$ 8-oxodG; $p< 0.001$ 8-isoprostane).

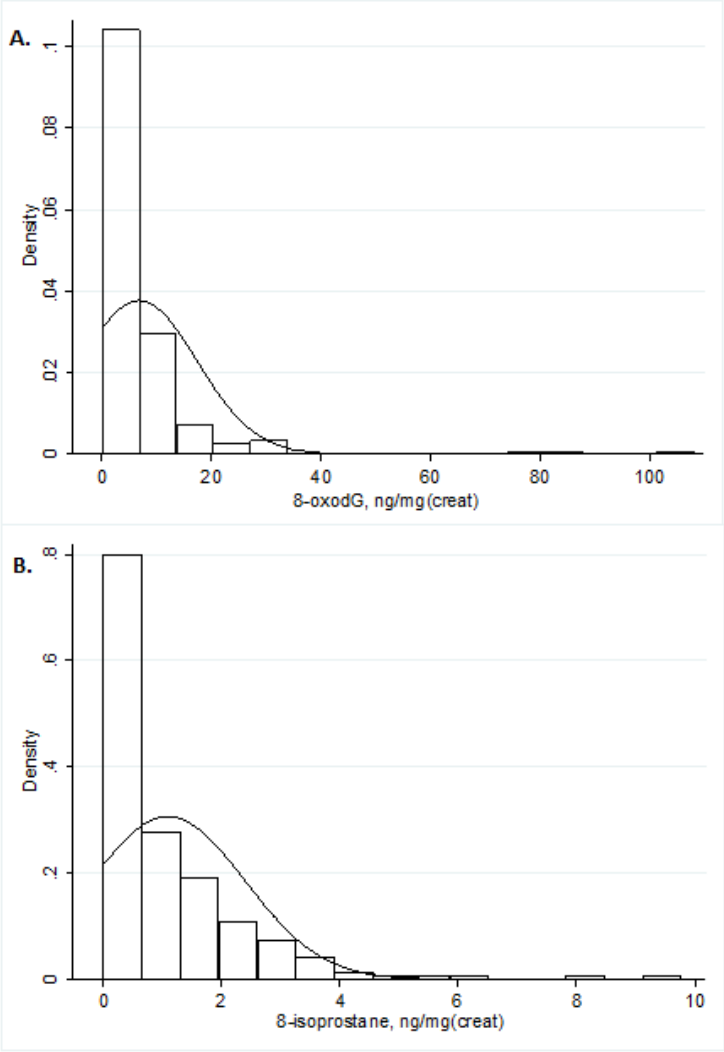


Figure 3-1. Frequency distribution of 8-oxodG (A) and 8-isoprostane (B).

The logarithmic transformation led to normalization of distribution in case of 8-oxodG (Figure 3-2, A) biomarker (p=0.11463, Shapiro-Wilk test), while in case of 8-isoprostane it provided visual normalization (Figure 3-2, B) but was not confirmed by Shapiro-Wilk test (p<0.001).

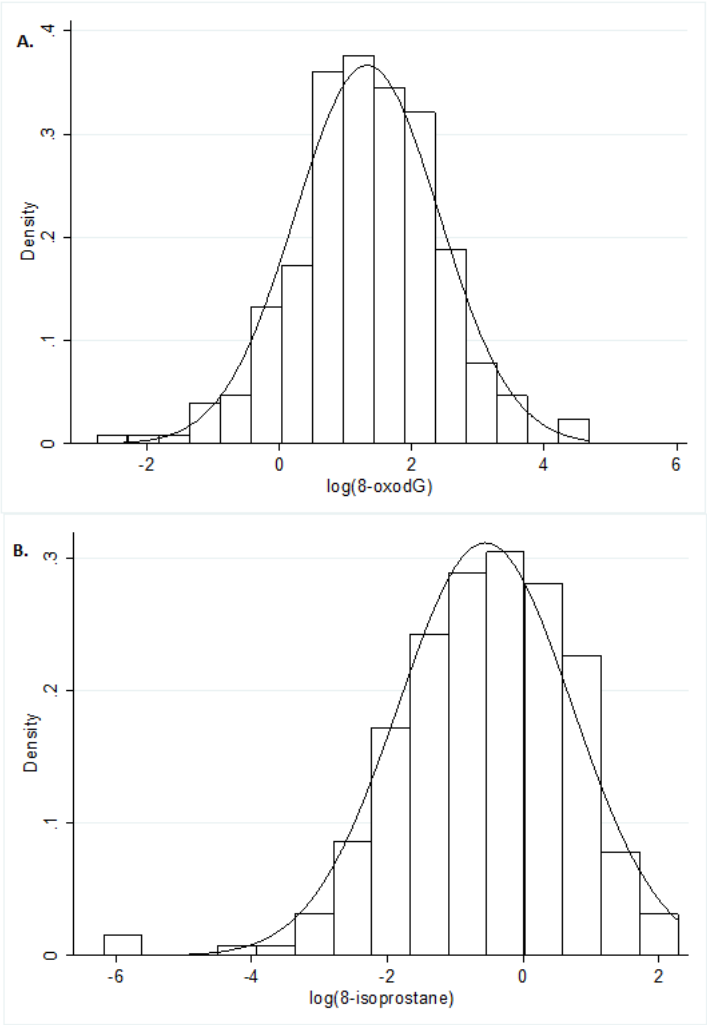


Figure 3-2. Frequency distribution of logarithmically transformed 8-oxodg (A) and 8-isoprostane (B).

3.2 Unadjusted Reference Intervals.

Reference intervals of the biomarkers, defined by parametric method, were: 0.45 – 31.85 ng/mg$_{creat}$ for 8-oxodG and 0.05 – 7.04 ng/mg$_{creat}$ for 8-isoprostane.

Non-parametric method provided the following RI for the biomarkers: 0.34 – 28.60 ng/mg$_{creat}$ for 8-oxodG and 0.05 – 4.33 ng/mg$_{creat}$ for 8-isoprostane, which differs by upper limit from RI for 8-isoprostane evaluated with parametric approach.

3.3 Main determinants of 8-oxodG and 8-isoprostane and adjusted RI estimation.

Linear Regression.

As far as the Shapiro-Wilk test for normality failed in the sample of 8-isoprostane, the linear regression analysis for this biomarker has not been used.

The unadjusted linear regression analysis showed significant negative association of logarithmically transformed 8-oxodG with DFC (coef.β=-0.0009, p=0.012; Table 3-2). Other unadjusted associations of explanatory variables with log(8-oxodG) did not provide significant results (Table 3-2).

Table 3-2. Unadjusted linear regression estimates of logarithmic 8-oxodG with explanatory variables one by one in the sample.

	log(8-oxodG)	
	coef.β	**p-value**
DFC, *days*	-0.0009	**0.012**
Season, *warm vs cold*	0.2400	0.070
Age, *years*	-0.0001	0.988
Gender, *women vs men*	-0.1605	0.227

In the multivariable analysis, including four variables simultaneously, the negative significant association of log(8-oxodG) with DFC has been confirmed (coef.β=-0.0013, p<0.001; Table 3-3). Adjusted for all explanatory variables linear regression analysis showed also increased value of log(8-oxodG) during the warm season in respect to cold season (coef.β=0.4299, p=0.002; Table 3-3).). In the multivariable analysis age and gender did not evidence statistically significant associations with log(8-oxodG) and, thus, were excluded from further analyses (Table 3-3).

Table 3-3. Adjusted linear regression estimates of logarithmic 8-oxodG with explanatory variables in the sample.

	log(8-oxodG)	
	coef.β	**p-value**
DFC, days	-0.0013	**<0.001**
Season, *warm vs cold*	0.4299	**0.002**
Age, *years*	0.0026	0.699
Gender, *women vs men*	-0.2063	0.118

The interaction term between DFC and season showed statistically significant effect on log(8-oxodG) data (Table 3-4).

Table 3-4 Adjusted linear regression estimates of logarithmic 8-oxodG with DFC, season and interaction term between DFC and season in the sample.

	log(8-oxodG)	
	coef.β	**p-value**
DFC, days	-0.0029	**<0.001**
Season, *warm vs cold*	-0.8681	**0.008**
DFC*Season	0.0030	**<0.001**

In regard with significant interaction term between season and DFC, the stratified linear regression analysis was performed and showed significant decrease of log(8-oxodG) values with increasing DFC in the cold season (coef.β= -0.0029, p-value<0.001, Table 3-5) and no significant association of DFC with log(8-oxodG) values in the warm season (coef.β=0.0001, p=0.846, Table 3-5).

Table 3-5. Linear regression estimates of logarithmic 8-oxodG with DFC during the warm and cold seasons.

	log(8-oxodG)			
	Warm Season		Cold Season	
	coef.β	p-value	coef.β	p-value
DFC, days	0.0001	0.846	-0.0029	**<0.001**

The assumption on normality of residuals in linear regression of log(8-oxodG) during the warm season was not met (Jarque-Bera normality test p<0.001, Figure 3-3 A), while during the cold season the residuals met the assumption on normality (Jarque-Bera normality test p=0.54, Figure 3-3 B).

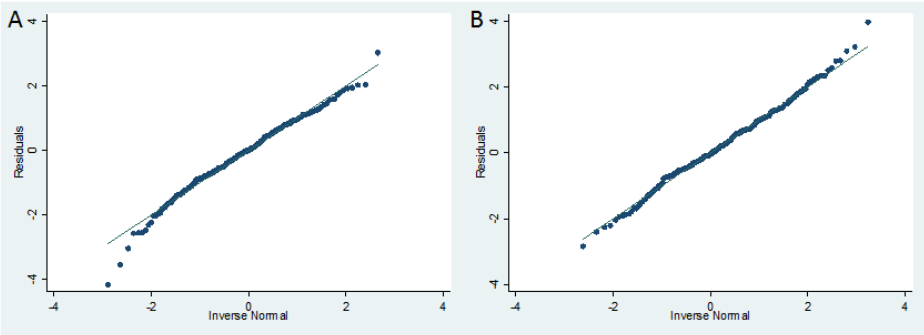


Figure 3-3. Quantiles of linear regression residuals against the quantiles of the normal distribution in the warm (A) and in the cold (B) subsamples of log(8-oxodG).

In accordance with assumption on normal distribution of residuals not met and with absence of significant association of DFC with log(8-oxodG) in the warm season, further use of linear regression in this case has been declined. For the homogenity of methods to be applied, linear regression analysis was excluded in the cold subsample of log(8-oxodG) as well.

Bootstrapping Quantile Regression.

The univariable median bootstrapping quantile regression analysis showed that in the warm season (April - September) there was a statistically significant higher level of 8-isoprostane in comparison with the cold (March - September) season (coef.β=0.4508, p=0.003;

Table 3-6). There was also evidence of a slight, but statistically significant loss of concentration of 8-oxodG with the increase of DFC (coef.β=-0.0031, p=0.026;

Table 3-6); this significant negative association was also confirmed at 2.5% quantile of 8-oxodG (coef.β=-0.001, p=0.018;

Table 3-6) and of 8-isoprostane (coef.β=-0.0002, p=0.003;

Table 3-6).

Table 3-6. Unadjusted bootstrapping quantile regression estimates at 2.5%, 50% and 97.5% quantiles of 8-oxodG and 8-isoprostane with all explanatory variables one by one in the sample.

	8-oxodG, ng/mg$_{creat}$			**8-isoprostane, ng/mg$_{creat}$**		
	coef.β (SE)	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**
Quantiles, *%*	**2.5**	**50**	**97.5**	**2.5**	**50**	**97.5**
DFC, *days*	-0.0010 * (0.0004)	**-0.0031 *** (0.0014)	0.0010 (0.0217)	-0.0002** (0.0001)	-0.0001 (0.0005)	-0.0012 (0.0058)
Season, *warm vs cold*	0.3349 (0.2676)	0.8952 (0.4990)	-2.8025 (14.6308)	0.0050 (0.0356)	**0.4508**** (0.1520)	0.6621 (2.0026)
Age, *years*	0.0134 (0.0137)	-0.0122 (0.0273)	0.1621 (0.7720)	0.0025 (0.0014)	0.0055 (0.0079)	0.0611 (0.0687)
Gender, *women vs men*	-0.3330 (0.1952)	0.0058 (0.5645)	-4.2056 (25.2150)	-0.0115 (0.0364)	0.1882 (0.1423)	-2.2124 (1.9137)

** - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001; SE-standard error*

In the Table 3-7Table 3-7 coefficient estimates of multivariable analysis, adjusted for age, gender, DFC and season are presented at 2.5%, 50% and 97.5% quantiles of response variables (8-oxodG or 8-isoprostane). The median quantile regression provided statistically significant daily decrease of 8-oxodG level by 0.0037 ng/mg$_{creat}$ (p=0.004; Table 3-7) and of 8-isoprostane by 0.0011 ng/mg$_{creat}$ (p=0.001;Table 3-7); and statistically significant higher levels of 8-oxodG

(coef.β=1.2242, p=0.021; Table 3-7) and of 8-isoprostane (coef.β=0.5093, p<0.001;Table 3-7) during the warm season respectively to the cold one. In case of 8-isoprostane these associations were confirmed at 97.5% quantile both for DFC (coef.β=-0.0041, p=0.024;Table 3-7) and for season (coef.β=1.5758, p=0.015;Table 3-7).

Table 3-7. Multivariable bootstrapping quantile regression estimates at 2.5%, 50% and 97.5% quantiles of 8-oxodG and 8-isoprostane, adjusted for DFC, season, age and gender in the sample.

	8-oxodG, ng/mg$_{creat}$			**8-isoprostane, ng/mg$_{creat}$**		
	coef.β (SE)	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**
Quantiles, *%*	**2.5**	**50**	**97.5**	**2.5**	**50**	**97.5**
DFC, *days*	-0.0009 (0.0009)	**-0.0037** ** (0.0013)	0.0087 (0.0214)	-0.0001 (0.0001)	**-0.0011**** (0.0003)	-0.0041* (0.0018)
Season, *warm vs cold*	0.2411 (0.3247)	**1.2242*** (0.5259)	-5.0104 (8.9784)	-0.0153 (0.0312)	**0.5093***** (0.1267)	1.5758* (0.6404)
Age, *years*	-0.0026 (0.0178)	-0.0145 (0.0237)	-0.0739 (0.3266)	0.0021 (0.0015)	0.0056 (0.0068)	0.0622 (0.0325)
Gender, *women vs men*	-0.1077 (0.3291)	0.0789 (0.5121)	-11.0921 (7.4052)	0.0012 (0.0263)	0.1954 (0.1249)	-1.4882 (1.7254)

** - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001; SE-standard error*

As age and gender did not provide statistically significant associations with both biomarkers, they were removed from the regression model and further analyses.

An interaction effect between DFC and season was checked in the multivariable median bootstrapping quantile regression models for both biomarkers. Statistically significant interaction between DFC and season was found for 8-oxodG biomarker (coef.β=0.0068, p=0.007, Table 3-8). In case of 8-isoprostane the interaction was not statistically significant (p=0.105), but both significant associations with DFC (coef.β=-0.0008, p=0.05; Table 3-8) and season (coef.β=1.4779, p=0.017) were confirmed in this model.

Table 3-8. Median multivariable bootstrapping quantile regression estimates of 8-oxodG and 8-isoprostane, adjusted for DFC, season and interaction between DFC and season in the sample.

	8-oxodG, ng/mg$_{creat}$	**8-isoprostane, ng/mg$_{creat}$**
	coef.β (SE)	**coef.β (SE)**
Quantiles, *%*	**50**	**50**
DFC, *days*	**-0.0074***** (0.0018)	**-0.0008*** (0.0004)
Season, *warm vs cold*	-1.8268 (1.2939)	**1.4779*** (0.6138)

DFC*Season	**0.0068**** (0.0025)	-0.0016 (0.0010)

** - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001*

The stratified by season bootstrapping quantile regression analysis, adjusted for DFC, was then performed for 8-oxodG and for the homogeneity of the analyses for 8-isoprostane as well.

Over the cold season in the median regression (50%), the statistically significant reduction of 8-oxodG level with the daily increase of DFC was still present (coef.β=-0.0074, p<0.001; Table 3-9). It was also saved for 8-isoprostane during the cold season at statistically significant level (coef.β=-0.0008, p=0.003; Table 3-9), and showed significantly lower level of 8-isoprostane with increasing DFC at 2.5% quantile (coef.β=-0.0002, p=0.002; Table 3-9).

Table 3-9. Multivariable bootstrapping quantile regression estimates at 2.5%, 50% and 97.5% quantiles of 8-oxodG and 8-isoprostane, adjusted for DFC during the cold season in the sample.

	8-oxodG, ng/mg$_{creat}$			**8-isoprostane, ng/mg$_{creat}$**		
	coef.β (SE)	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**
Quantiles, *%*	**2.5**	**50**	**97.5**	**2.5**	**50**	**97.5**
DFC, *days*	-0.0014 (0.0009)	**-0.0074***** (0.0018)	0.0441 (0.0493)	-0.0002** (0.0001)	**-0.0008*** (0.0004)	-0.0058 (0.0057)

** - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001; SE-standard error*

During the warm season the negative association between DFC and 8-oxodG was not present in the median bootstrapping quantile regression (coef.β=-0.0006, p=0.694, Table 3-10). 8-isoprostane proved statistically significant negative association with DFC during the warm season in the median bootstrapping quantile regression (coef.β=-0.0024, p=0.007Table 3-10).

Table 3-10. Bootstrapping quantile regression estimates at 2.5%, 50% and 97.5% quantiles of 8-oxodG and 8-isoprostane, adjusted for DFC during the warm season in the sample.

	8-oxodG, ng/mg$_{creat}$			**8-isoprostane, ng/mg$_{creat}$**		
	coef.β (SE)	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**	**coef.β (SE)**
Quantiles, *%*	**2.5**	**50**	**97.5**	**2.5**	**50**	**97.5**
DFC, *days*	0.0010 (0.0013)	-0.0006 (0.0016)	0.0288 (0.0444)	-0.0003 (0.0004)	**-0.0024**** (0.0009)	-0.0023 (0.0047)

** - p-value <0.05; ** - p-value <0.01; *** - p-value <0.001; SE-standard error*

Centile curves were built after stratified by season bootstrapping QR for 2.5%, 50% and 97.5% quantiles for both biomarkers.

There is a similar trend in 2.5% and 50% curves, which showed decrease of 8-oxodG values against DFC in cold season, but upper 97.5% curve showed

increasing progress with the increase of DFC in this model (Figure 3-4).

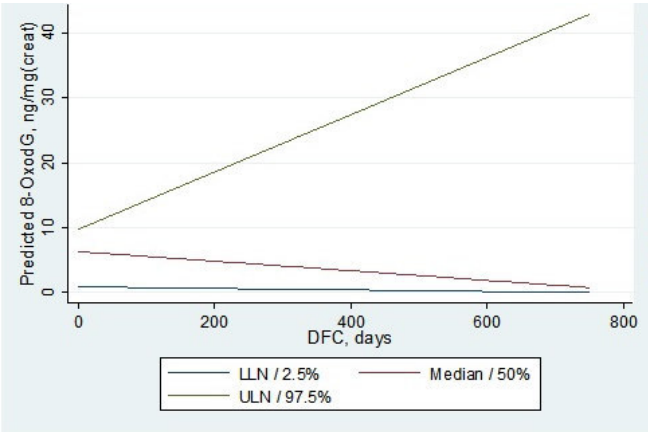


Figure 3-4. Linear predictions of 8-OxodG values against DFC in cold season.

On the graph with linear predictions for 8-oxodG values in the warm season a difference was seen between median (50%) curve which showed quite constant central trend against DFC, while 2.5% curve showed slight linear increase and 97.5% curve showed sharp increase of 8-oxodG values with increase of DFC (Figure 3-5).

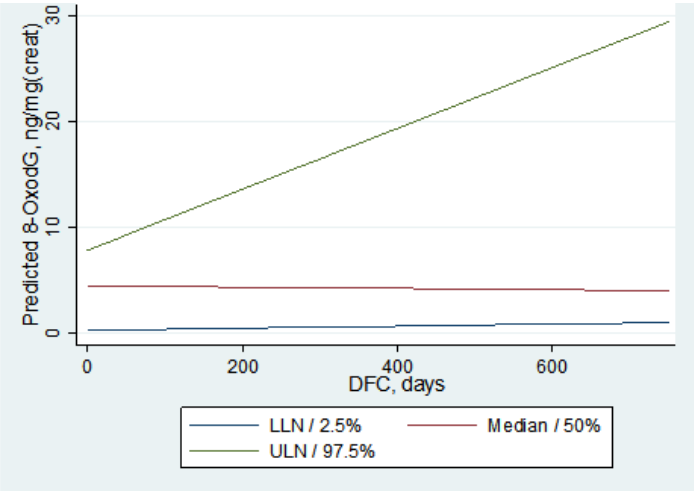


Figure 3-5. Linear predictions of 8-OxodG values against DFC in warm season.

Quantile regression centile curves for 8-isoprostane in cold season showed linear decrease of 8-isoprostane values with increasing distance from collection at all quantiles (Figure 3-6).

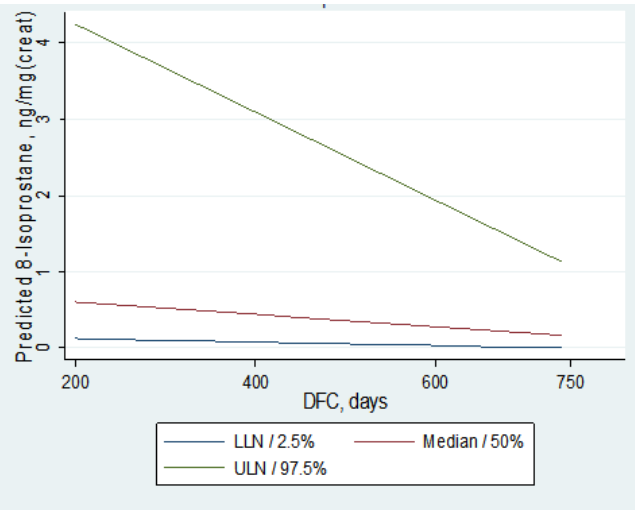


Figure 3-6. Linear predictions of 8-isoprostane values against DFC in cold season.

In the warm season 2.5%, 50% and 97.5% centile curves showed linear decrease of the 8-isoprostane values against DFC values (Figure 3-7, A and B).

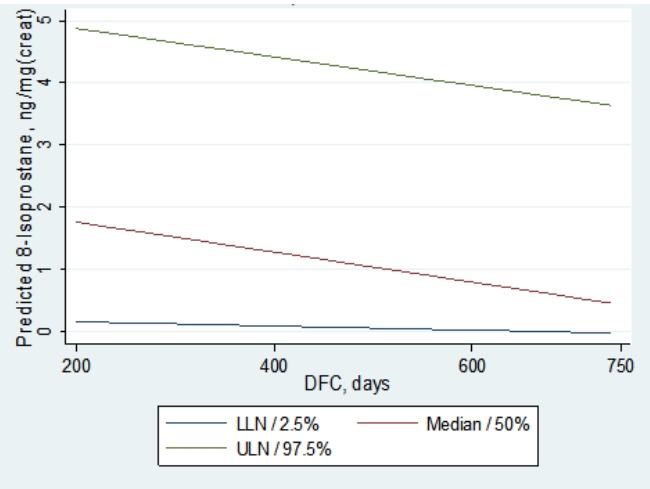


Figure 3-7. Linear predictions of 8-isoprostane values against DFC in warm season.

The quantile regression method was not considered in further analyses because of the lack of statistical techniques to check the goodness of fit of the estimated QR models.

GAMLSS Regression.

The best GAMLSS regression models for entire sample inside of each distribution family (BCCG, BCT and NO) for each biomarker are presented in the Table 3-11. Thus, in case of NO distribution, the best model was represented by identity mu.link function for 8-oxodG with log DFC; sigma.link function was logarithmic by default. This combination provided the AIC=1707.59 and BIC=1721.66 with df=4 (Table 3-11). The range of AIC for all tried models was 1711.55 - 1707.59, and 1732.650- 1721.66 for BIC criterion. However, none of

the models provided good fitting to the data from the point of view of residuals distribution (Figure 3-8), and significance of the cofactors: log(DFC) (coef.β=-0.629, p=0.444), season *warm vs cold* (coef.β=1.005, p=0.306).

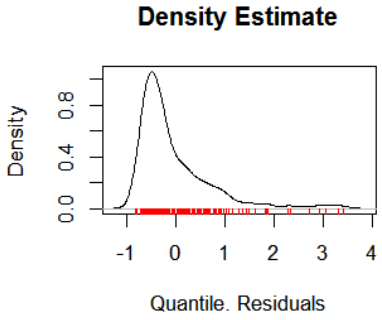


Figure 3-8. The residual distribution of 8-oxodG in the sample: quantile residuals plotted against DFC in NO family.

Table 3-11. Development of the GAMLSS model for 8-oxodG and for 8-isoprostane in the sample.

Model	Distribution	Linear predictor for μ					Linear predictor for σ	Linear predictor for ν	Linear predictor for τ	df	BIC	AIC
		Link Function	DFC	Age	Season	Gender	Link Function	Link Function	Link Function			
8-oxodG, ng/mg$_{creat}$												
1	NO	identity	log	-	binary	-	log	-	-	3	1721.66	1707.59
2	BCCG	log	linear	linear	binary	binary	log	identity	-	7	11079.18	11054.56
3	BCT	identity	linear	-	binary	-	log	identity	log	6	1365.80	1344.69
8-isoprostane, ng/mg$_{creat}$												
1	NO	identity	linear	-	binary	-	log	-	-	4	645.28	632.08
2	BCCG	log	linear	linear	binary	binary	log	identity	-	7	9909.684	9886.60
3	BCT	identity	linear	-	binary	-	log	identity	log	6	412.140	392.320

The best NO model for 8-isoprostane with identity mu.link function had logarithmic sigma.link function and DFC on a linear scale. The AIC of this model was 632.080, and BIC=645.280 with 4 degrees of freedom for the fit (Table 3-11). The range of AIC for all built models was 635.08 - 632.08, and 651.57 - 645.28 for BIC criterion. In case of 8-isoprostane neither normal nor log-normal distribution regression models provided normal residual distributions: mean=$3.59*10^{-06}$, variance=1.01, coef. of skewness=3.19, coef. of kurtosis=20.54. However, both explanatory variables showed significant association with 8-isoprostane: DFC (coef.β=-0.002, p=0.015), season *warm vs cold* (coef.β=0.577, p=0.003).

The best, among built for 8-oxodG BCCG models, had log mu.link function with log sigma.link and identity nu.link functions and had DFC and age variables expressed on a linear scale and included two binary variables, season and gender. The AIC of this model was equal to 11054.56 and varied up to 11163.76, BIC was equal to 11079.18 and its biggest value was 11181.34, the best model had 7 degrees of freedom for the fit (Table 3-11). None of the cofactors in the model were significantly associated with 8-oxodG: DFC (coef.β=$-8.014*10^{-17}$, p=1.00), age (coef.β=$-4.662*10^{-16}$, p=1.00), season *warm vs cold* (coef.β=$-5.698*10^{-15}$, p=1.00), gender *women vs men* (coef.β=$1.336*10^{-14}$, p=1.00). The residual distribution of 8-oxodG did not show the good fitting to the data (Figure 3-9).

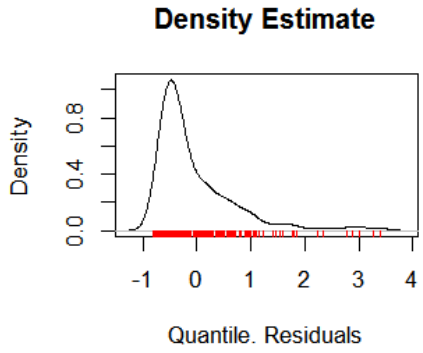


Figure 3-9. The residual distribution of 8-oxodG in the sample: quantile residuals plotted against DFC in BCCG family.

The identical best BCCG model was developed for 8-isoprostane, that was characterized by log mu.link and sigma.link functions and by identity nu.link function, by linear age and DFC cofactors and two binary variables, season and gender. The AIC of this model was equal to 9886.60 and varied up to 10369.37, BIC was equal to 9909.68 and its biggest value was 10385.86, the best model had 7 degrees of freedom for the fit (Table 3-11). As in case of 8-oxodG, none of the

explanatory variables in the model provided statistically significant associations with 8-isoprostane: DFC (coef.β=-1.278*10^{-17}, p=1.00), age (coef.β=-1.368*10^{-15}, p=1.00), season *warm vs cold* (coef.β=-1.039*10^{-14}, p=1.00), gender *women vs men* (coef.β=-1.053*10^{-14}, p=1.00). The residuals' distribution was not normalized as well: mean=-6.60, variance= 0.01, coef. of skewness=-0.55, coef. of kurtosis=3.94.

Within the BCT distribution family for 8-oxodG the lowest BIC value (1365.80) and AIC=1344.69 were reached in the model with 6 degrees of freedom for the fit and two explanatory variables, i.e. DFC and season. The range of BIC values in all built models was 1382.74 – 1365.80, and the range of AIC values was between 1354.605 and 1342.344. The model was characterized by identity mu.link and nu.link functions and by log sigma.link and tau.link functions; DFC was expressed on a linear scale (Table 3-11).

The identical model was developed for 8-isoprosatne within the BCT distribution family. The model had two explanatory variables, i.e. DFC and season, identity mu and nu link functions and log link functions for sigma and tau; DFC was on a linear scale (Table 3-11) and df=6. This model had BIC=412.140 and varied from 424.76 to 412.140, AIC was equal to 392.320 and its range was 398.33 - 387.13.

In both cases for 8-oxodG and for 8-isoprostane biomarkers, the best models were built with BCT distribution: lowest BIC criteria and normalized residuals distributions.

BCT distribution models from Table 3-11 with interaction term between DFC and season were built for both biomarkers to control for effect modification. This effect was found in case of 8-oxodG: DFC*season (coef.β=0.0087, p=5.25*10^{-05}) (Table 3-12). It was not observed for 8-isoprostane: DFC (coef.β=1.252*10^{-03}, p=6.80*10^{-07}), season *warm vs cold* (coef.β=4.222*10^{-01}, p=0.402), DFC*season (coef.β=2.828*10^{-05}, p=0.972) Table 3-12.

Table 3-12. GAMLSS regression estimates of 8-oxodG and 8-isoprostane, adjusted for DFC, season and interaction between DFC and season in the sample.

	8-oxodG, ng/mg$_{creat}$		**8-isoprostane, ng/mg$_{creat}$**	
Determinants	**coef.β**	**p-value**	**coef.β**	**p-value**
DFC, *days*	-0.0082	<0.001	1.252e-03	<0.001
Season, *warm vs cold*	-2.7831	0.019	4.222e-01	0.402
DFC*Season	0.0087	**<0.001**	2.828e-05	0.972

The development of new GAMLSS regression models stratified by season and adjusted for DFC, was then performed for 8-oxodG and for the homogeneity of the analyses for 8-isoprostane as well.

From around the distribution families tried for each biomarker per each season, the two best models for each biomarker per season were chosen for further consideration and processing (Table 3-13).

Thus, for 8-oxodG during the cold season two better models were built with LOGNO and BCCGo distributions (Table 3-13). For modelling LOGNO gamlss regression for 8-oxodG with log mu.link function, DFC was on a linear scale, linear + quadratic terms for σ parameter with $edf_{\mu,\sigma DFC=}0$ were used (edf – effective degrees of freedom, characterize the complexity of the spline curve). This combination of modelled parameters allowed reducing BIC (684.90) of the LOGNO model by ~5 units in comparison with BCCGo model (689.88) , as well as degrees of freedom for the fit from 8 of BCCGo model to 5 for the LOGNO model (Table 3-13). The LOGNO model also showed a slightly better normalization of quantile residuals distribution (Table 3-14), which can be also seen on Density Estimate and Normal Q-Q plots (Figure 3-10). Centile curves showed similar trends for 2.5% and 50% lines, but the 97.5% curves were different, and in case of LOGNO distribution it showed more appropriate biological plausibility (Figure 3-11). LOGNO gamlss regression model was chosen for 8-oxodG during the cold season for further analyses.

Table 3-13 Development of the GAMLSS model for 8-oxodG and for 8-isoprostane over warm and cold seasons in the sample with linear predictors for median μ, variability σ and skewness ν.

Model	Distribution	Linear predictor for μ			Linear predictor for log σ		Linear predictor for ν			df	BIC	AIC
		Link Function	$edf_{\mu DFC}$	p_{DFC}	$edf_{\sigma DFC}$	p_{DFC}	Link Function	$edf_{\nu DFC}$	p_{DFC}			
8-oxodG, ng/mg$_{creat.}$ Cold season.												
1	**LOGNO**	log	0	linear	0	linear+quadratic	-	-	-	5	684.90	670.64
2	**BCCGo**	log	0	linear	0	linear+quadratic	identity	0	linear+quadratic	8	689.88	667.06
8-oxodG, ng/mg$_{creat.}$ Warm season.												
1	**LOGNO**	log	0	linear	0	linear+quadratic	-	-	-	5	677.21	663.23
2	**BCCGo**	log	3	pb	2	pb	identity	2	pb	7	682.32	662.84
8-isoprostane, ng/mg$_{creat.}$ Cold season.												
1	**BCCGo**	log	0	linear	0	quadratic	identity	0	quadratic	6	183.17	167.19
2	**GA**	identity	0	linear	0	linear+quadratic	-	-	-	5	195.19	181.88
8-isoprostane, ng/mg$_{creat.}$ Warm season.												
1	**BCCGo**	log	2	pb	2	pb	identity	2	pb	6	234.86	219.54
2	**GA**	identity	0	linear	0	linear+quadratic	-	-	-	5	235.58	222.81

p_{DFC} - polynomials or splines fitted in a GAMLSS formula,
pb - a penalized B-spline fitted in a GAMLSS formula,
edf – characterizes the complexity of the spline curve; edf=0 corresponds to the linear term, the higher it is, the more complex is the curve (65).

Table 3-14. Summary of the Quantile Residuals in LOGNO and BCCGo models of 8-oxodG during the cold season.

	LOGNO	BCCGo
mean	0.0110	0.0154
variance	1.0078	1.0379
coef. of skewness	-0.2199	-0.0976
coef. of kurtosis	2.6595	2.2741

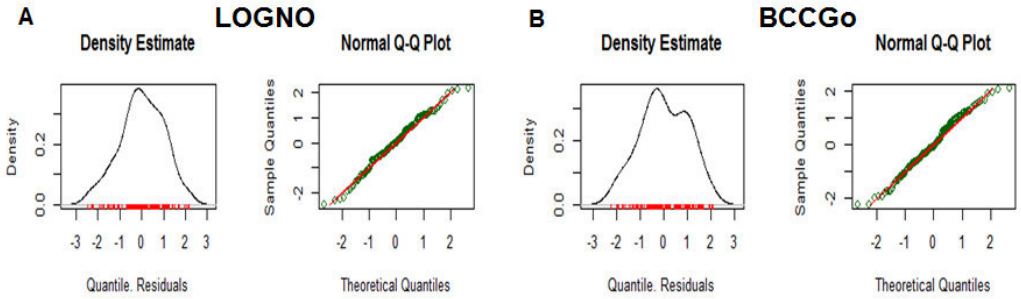


Figure 3-10. The residual distribution of 8-oxodG in cold season: the density estimate with rug plot and the quantile-quantile plot for LOGNO (A) model and for BCCGo (B) model.

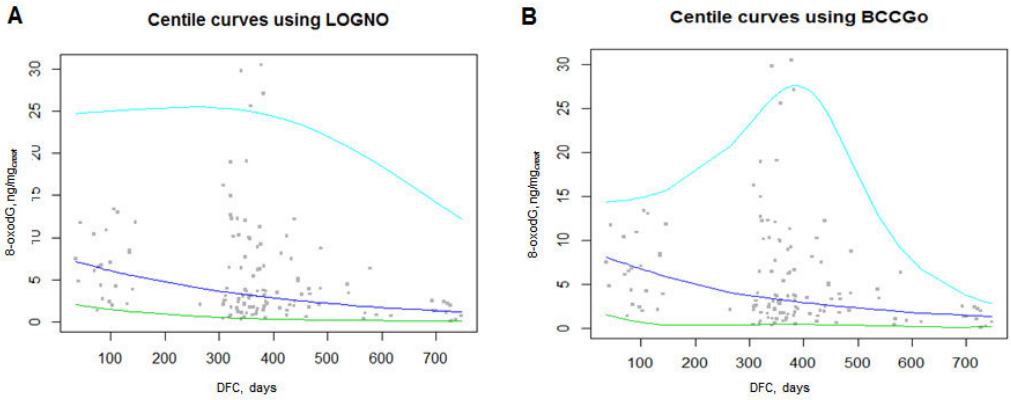


Figure 3-11. Observed 8-oxodG values in cold season with three fitted model centile curves (2.5%, 50%, 97.5%) from LOGNO (A) and BCCGo (B) models against DFC.

LOGNO and BCCGo gamlss regression models showed also better fitting in 8-oxodG during the warm season. As in the cold season, LOGNO distribution family allowed reducing BIC by ~5 units from 682.32 (BCCGo) to 677.21 (LOGNO), and degrees of freedom for the fit from 7 (BCCGo) to 5 (LOGNO) (Table 3-13). Adding penalized B-splines into the BCCGo model, did not improve the model, and instead increased $edf_{\mu DFC}$=3 by 3 units in comparison with linear term for LOGNO distribution $edf_{\mu DFC}$=0, as well as $edf_{\sigma DFC}$=2 by 2 units in comparison with LOGNO model $edf_{\sigma DFC}$=0, and added $edf_{\nu DFC}$=2 for ν parameter. The residuals were distributed similarly in both models and showed values of

mean, variance and coefficient of skewness close to normal (standardized) values, with a slightly higher coefficient of kurtosis in case of BCCGo model (Table 3-15, Figure 3-12). As in case of 8-oxodG in cold season, the 97.5% centile curve was more biologically plausible in case of LOGNO model for 8-oxodG in warm season (Figure 3-13).

Table 3-15. Summary of the Quantile Residuals in LOGNO and BCCGo models of 8-oxodG during the warm season.

	LOGNO	BCCGo
mean	-0.0084	-0.0003
variance	1.0083	1.0083
coef. of skewness	-0.0618	0.0098
coef. of kurtosis	3.4513	3.5655

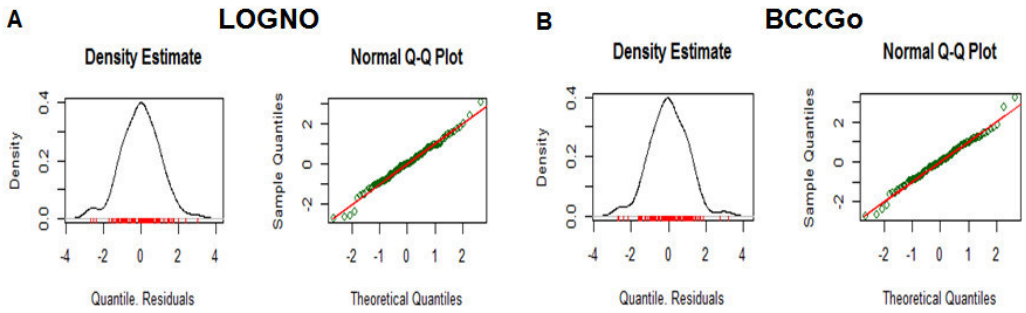


Figure 3-12. The residual distribution of 8-oxodG in warm season: the density estimate with rug plot and the quantile-quantile plot for LOGNO (A) model and for BCCGo (B) model.

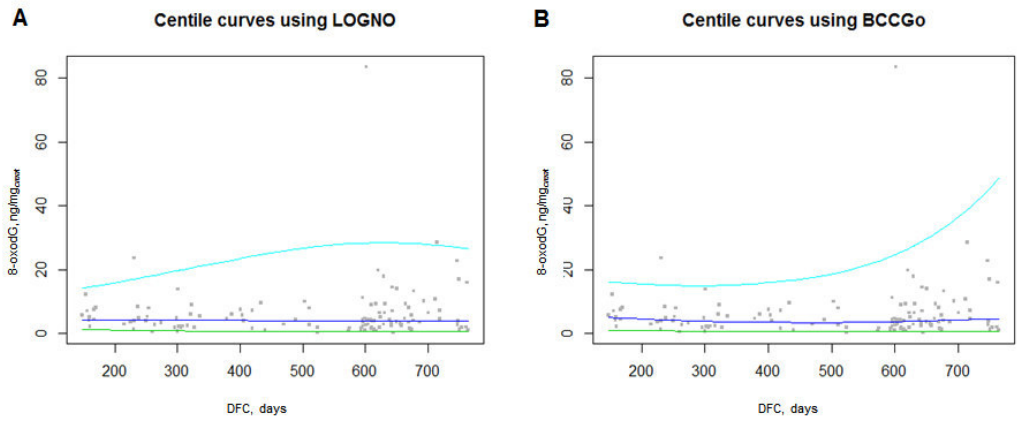


Figure 3-13. Observed 8-oxodG values in warm season with three fitted model centile curves (2.5%, 50%, 97.5%) from LOGNO (A) and BCCGo (B) models against DFC.

8-isoprostane in cold season was modelled better using BCCGo and GA distribution families. The BCCGo model showed better (lower) BIC criterion, BIC=183.17 in comparison with GA model, BIC=195.19. The BCCGo model had one unit more (df=6) degrees of freedom for the fit in respect to df=5 for GA model (Table 3-13). In case of BCCGo model log mu.link function was used, while in GA model the identity one. In both models all effective degrees of

freedom $edf_{\mu,\sigma,(\nu)DFC}$ were equal to 0, as no additive terms were applied. The BCCGo model allowed achieving also better standardization of the residuals distribution (Table 3-16, Figure 3-14). Centile curves in both models showed similar trends for 2.5% and 50% lines, with sharper extrapolation of 97.5% line to the 8-isoprostane values axis in BCCGo model (Figure 3-15). Summarizing all mentioned factors the BCCGo model for 8-isoprostane during the cold season has been chosen for further analyses.

Table 3-16. Summary of the Quantile Residuals in GA and BCCGo models of 8-isoprostane during the cold season.

	GA	BCCGo
mean	-0.0313	-0.0020
variance	0.9735	1.0116
coef. of skewness	0.9069	0.0093
coef. of kurtosis	4.1965	2.2605

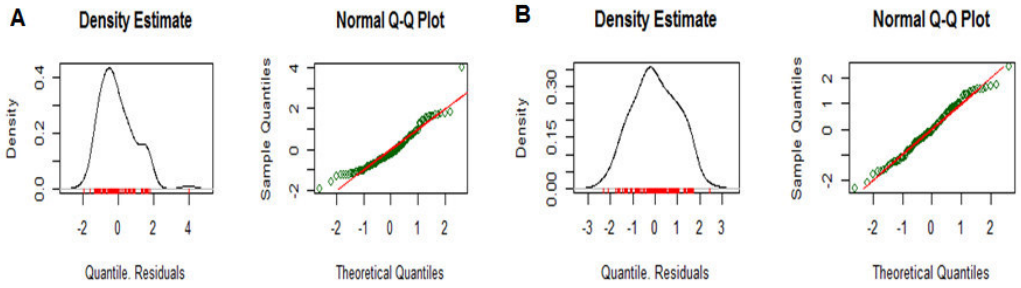


Figure 3-14. The residual distribution of 8-isoprostane in cold season: the density estimate with rug plot and the quantile-quantile plot for GA (A) model and for BCCGo (B) model.

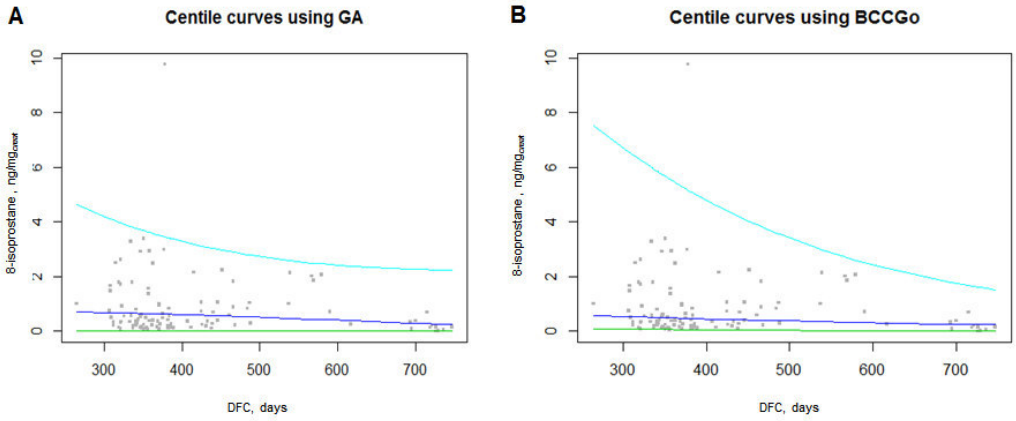


Figure 3-15. Observed 8-isoprostane values in cold season with three fitted model centile curves (2.5%, 50%, 97.5%) from GA (A) and BCCGo (B) models against DFC.

The GA and BCCGo distribution models showed also better results for 8-isoprostane in warm season. Both models showed close to each other values of BIC criterion, BIC=234.86 for BCCGo and BIC=235.58 for GA distribution families. The BCCGo model had one unit more (df=6) degrees of freedom for the fit in respect to df=5 for GA model (Table 3-13). In case of BCCGo model

penalized B-splines were used to fit the data better, which increased $edf_{\mu DFC}$=2 by 2 units in comparison with linear term for GA distribution $edf_{\mu DFC}$=0, as well as $edf_{\sigma DFC}$=2 by 2 units in comparison with GA model $edf_{\sigma DFC}$=0, and added $edf_{\nu DFC}$=2 for ν. Adding these penalized B-splines in BCCGo model did not improve it a lot neither from point of view of BIC criterion (0.72 units lower value in comparison with GA model), nor from point of view of residuals distribution (Table 3-17, Figure 3-16), meanwhile added one degree of freedom for the fit: df=6 for BCCGo model in respect to df=5 for GA model (Table 3-13). The 97.5% centile curve for GA model also showed slightly more plausible trend (Figure 3-17). Thus, the GA model for 8-isoprsotane in warm season has been chosen for further analyses.

Table 3-17. Summary of the Quantile Residuals in GA and BCCGo models of 8-isoprostane during the warm season.

	GA	BCCGo
mean	-0.0313	-0.0020
variance	0.9735	1.0116
coef. of skewness	0.9069	0.0093
coef. of kurtosis	4.1965	2.2605

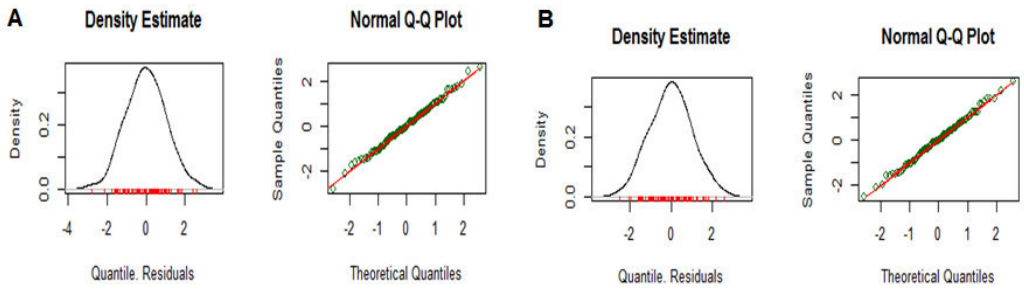


Figure 3-16. The residual distribution of 8-isoprostane in warm season: the density estimate with rug plot and the quantile-quantile plot for GA (A) model and for BCCGo (B) model.

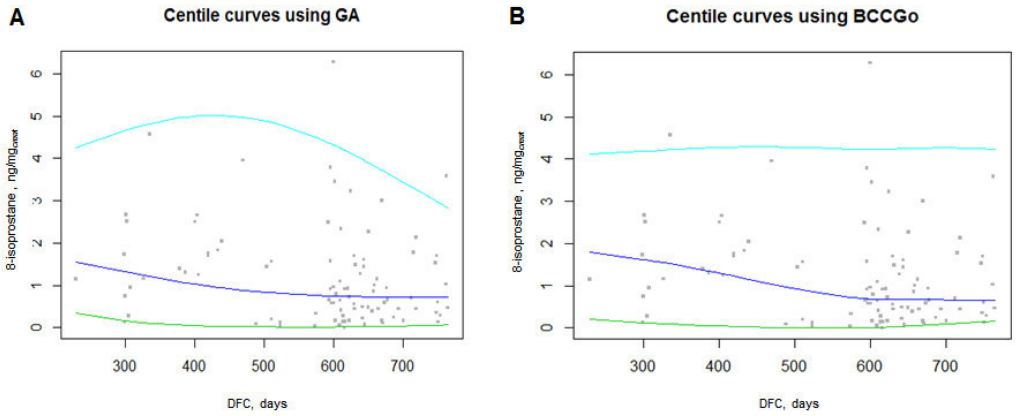


Figure 3-17. Observed 8-isoprostane values in warm season with three fitted model centile curves (2.5%, 50%, 97.5%) from GA (A) and BCCGo (B) models against DFC.

Some centile curves built after GAMLSS fitted models showed overestimating the outcome values (for example like in case of 8-isoprostane in cold season, Figure 3-15 B). The new models with excluded extreme values were tested to see whether it could improve the fitting. As an alternative approach, to avoid removing outliers, the models with additive terms, like cubic smoothing splines, were built to check which of them fit the data better in each case (two biomarkers in two seasons). Extreme values were defined with influential plots for outliers (Figure 3-18 - Figure 3-21).

The analysis for outliers in the subsample of 8-oxodG during the warm period showed outliers in terms of all three parameters, i.e. Studentized residuals, hat values and Cook's distance: the point 83.49 ng/mg$_{creat}$ (Figure 3-18) was characterized by higher values of Studentized residual, hat values and Cook's distance and that is why was deleted for checking GAMLSS models with no outliers.

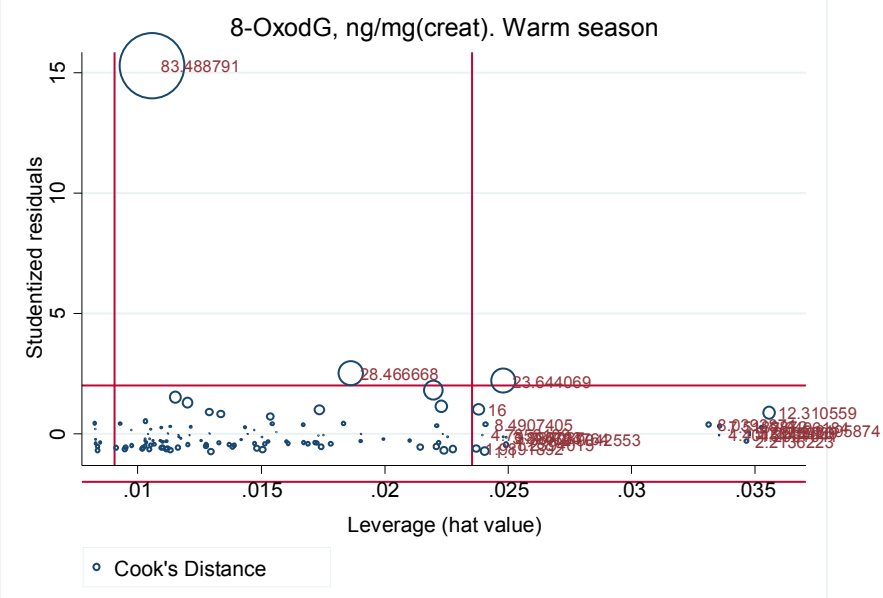


Figure 3-18. Outliers in the sample of 8-oxodG during the warm season.

During the cold season the subset of 8-oxodG values was characterized by 4 points with higher levels of Studentized residuals, hat values and Cook's distances, i.e. 25.57, 27.06, 29.18 and 30.15 ng/mg$_{creat}$ (Figure 3-19), that were then removed.

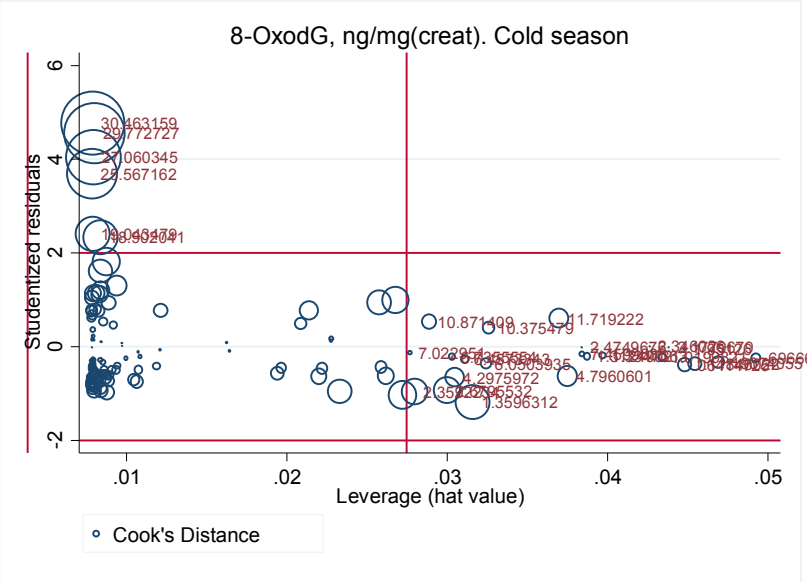


Figure 3-19. Outliers in the sample of 8-oxodG during the cold season.

8-isoprostane over the warm season showed outliers in terms of all three parameters as well, i.e. Studentized residuals, hat value and Cook's distance, the point 6.27 ng/mg$_{creat}$ (Figure 3-20) showed higher values and, consequently, has been removed.

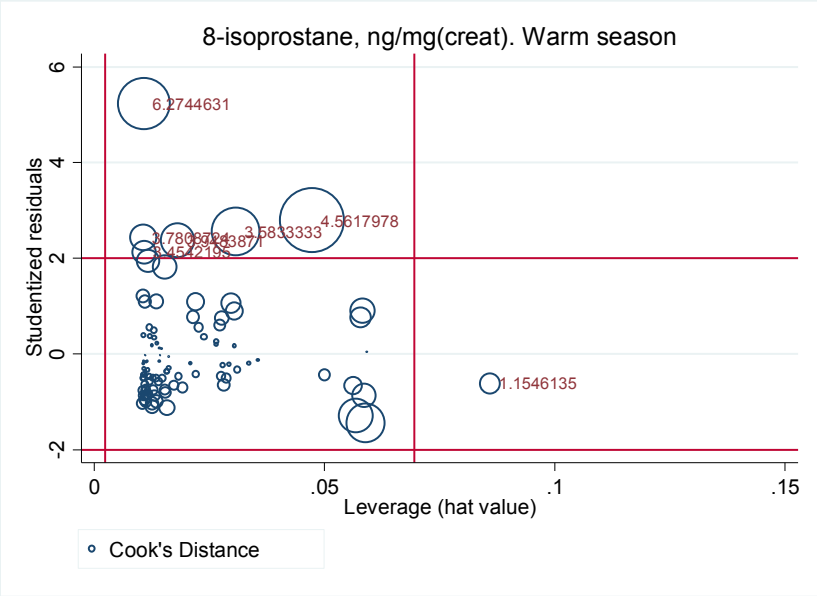


Figure 3-20. Outliers in the sample of 8-isoprostane during the warm season.

Over the cold season 8-isoprostane was characterized by higher Studentized residual, hat value and Cook's distance values for the point 9.77 ng/mg$_{creat}$ (Figure 3-21) and, thus, this observation has been removed.

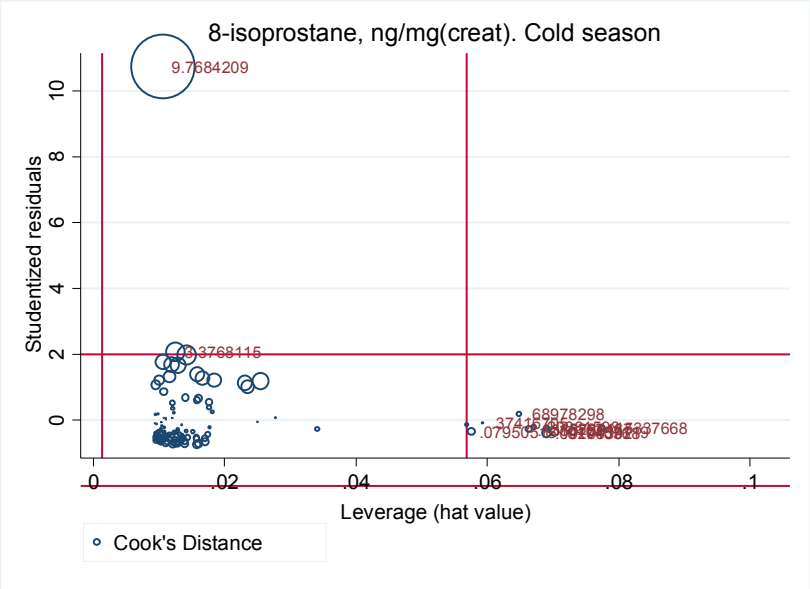


Figure 3-21. . Outliers in the sample of 8-isoprostane during the cold season.

These models were compared then with models containing smoothing cubic splines for parameters of DFC, and with models without outliers and containing cubic splines for parameters of DFC, to check which of them fit the data better in each case.

In all cases, i.e. for 8-oxodG in cold and warm season, as well as for 8-isoprostane in cold and warm season, the model without outliers showed better fitting from point of view of BIC criteria, residuals distribution and biological plausibility of centile curves (Table 3-18).

Table 3-18. Comparison between standard, no outliers, cubic splines and no outliers + cubic splines models for 8-oxodG and 8-isoprostane during the cold and the warm seasons.

Model	Distribution	df	BIC	Residuals distribution			
				mean	variance	coef. of skewness	coef. of kurtosis
8-oxodG. Cold season							
Standard*	LOGNO	5	684.90	0.0110	1.0078	-0.2199	2.6595
No outliers	LOGNO	5	631.37	0.0045	1.0081	-0.3605	2.5457
Cubic splines	LOGNO	6	685.20	0.0048	1.0079	-0.2701	2.5177
No outliers + Cubic splines	LOGNO	6	633.41	0.0024	1.0081	-0.3892	2.4649
8-oxodG. Warm season							
Standard*	LOGNO	5	677.21	-0.0084	1.0083	-0.0618	3.4513
No outliers	LOGNO	5	656.51	-0.0026	1.0084	-0.2983	3.1647
Cubic splines	LOGNO	6	677.34	-0.0037	1.0083	-0.0522	3.4685
No outliers + Cubic splines	LOGNO	6	656.00	-0.0021	1.0084	-0.2885	3.1013
8-isoprostane. Cold season							
Standard*	BCCGo	6	183.17	-0.0020	1.0116	0.0093	2.2605
No outliers	BCCGo	6	169.73	-0.0024	1.0112	0.0020	2.1098
Cubic splines	BCCGo	7	177.59	-0.0017	1.0173	-0.0074	2.2703
No outliers + Cubic splines	BCCGo	7	163.77	-0.0018	1.0174	-0.0168	2.1109
8-isoprostane. Warm season							
Standard*	GA	5	235.58	-0.0002	1.0095	0.0346	2.9073
No outliers	GA	5	224.23	-0.0006	1.0081	-0.0770	2.8085
Cubic splines	GA	6	238.98	-0.0003	1.0094	0.0243	2.9271
No outliers + Cubic splines	GA	6	227.14	0.0018	1.0111	-0.0977	2.7759

Standard - under the standard model here is considered the best developed GAMLSS model and presented in the Table 3-11.*

Afterwards with all 'No outliers' models for each biomarker per each season different combinations of linear functions and quadratic polynomials for μ, σ and ν were tried to define the best model with no outliers term.

Such in case of 8-oxodG during the cold and the warm seasons within LOGNO distribution family that allows modelling just two parameters: μ and σ, 12 models have been built with different combinations of linear and quadratic terms for μ and σ parameters. Same 12 combinations of terms were used for 8-isoprostane during the warm season within GA distribution (see Appendix I Table 1).

In case of 8-isoprostane in cold season within BCCGo distribution (which allows modelling μ, σ and ν parameters), 48 models with different combinations of linear and quadratic terms were built to fit the data (see Appendix I Table 2).

Using all procedures described above and considering such factors as BIC criteria, residuals distribution, degrees of freedom for the fit and centile curves, the best final models have been chosen for each biomarker per each season (Table 3-19).

Removing quadratic polynomial for σ parameter in 8-oxodG model during the cold season reduced BIC criterion (627.02, Table 3-19) by 4.4 units in comparison with previous 'No outliers' model BIC=631.37 (Table 3-18), as well as degrees of freedom by one unit from 5 (Table 3-18) to 4 (Table 3-19). This model also provided slightly better residuals distribution (Figure 3-22) and more plausible 97.5% centile curve (Figure 3-23). Statistically significant mean negative association of DFC with log(8-oxodG) presented in this model (coef.β=-0.003, $p=3.52*10^{-6}$).

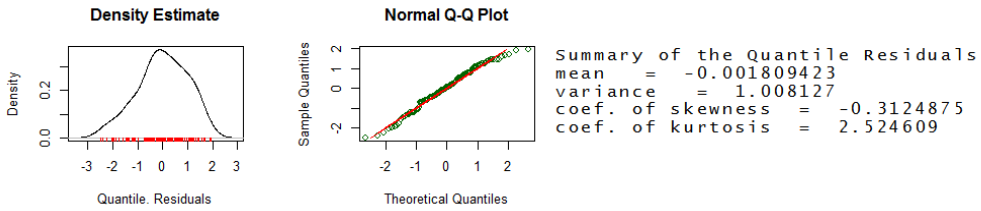


Figure 3-22. The residual distribution of 8-oxodG in cold season: the density estimate with rug plot and the quantile-quantile plot for LOGNO model.

Table 3-19. Best GAMLSS models for 8-oxodG and for 8-isoprostane over warm and cold seasons in the sample with linear predictors for median μ, variability σ and skewness ν.

Model	Distribution	Linear predictor for μ			Linear predictor for log σ		Linear predictor for ν			df	BIC	AIC
		Link Function	$edf_{\mu DFC}$	p_{DFC}	$edf_{\sigma DFC}$	p_{DFC}	Link Function	$edf_{\nu DFC}$	p_{DFC}			
8-oxodG, ng/mg$_{creat.}$ Cold season.												
1	**LOGNO**	log	0	linear	0	linear	-	-	-	4	627.02	615.74
8-oxodG, ng/mg$_{creat.}$ Warm season.												
1	**LOGNO**	log	0	linear	0	-	-	-	-	3	649.96	641.60
8-isoprostane, ng/mg$_{creat.}$ Cold season.												
1	**BCCGo**	log	0	linear	0	-	identity	0	-	4	166.32	155.71
8-isoprostane, ng/mg$_{creat.}$ Warm season.												
2	**GA**	identity	0	linear	0	-	-	-	-	3	217.49	209.86

p_{DFC} - polynomials or splines fitted in a GAMLSS formula,
edf – characterizes the complexity of the spline curve; edf=0 corresponds to the linear term, the higher it is, the more complex is the curve (65).

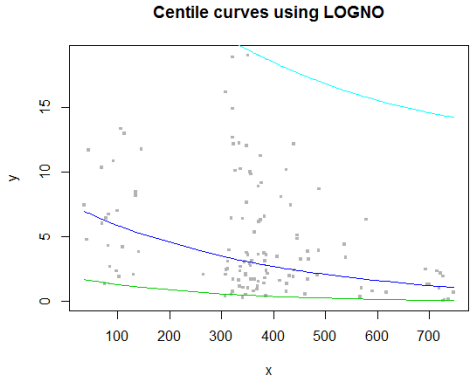


Figure 3-23. Observed 8-oxodG values in cold season with three fitted model centile curves (2.5%, 50%, 97.5%) from LOGNO model against DFC.

In the warm season for 8-oxodG the best model with no terms for sigma parameter the BIC criterion was reduced by 6.6 units from 656.51 in 'No outliers' model (Table 3-18) to 649.96 (Table 3-19). It also showed a slightly better residuals distribution (Figure 3-24) and plausible 97.5% centile curve (Figure 3-25), which was also in agreement with no statistically significant association between DFC and log(8-oxodG) in warm season (coef.β=1.502*10^{-5}, p=0.971).

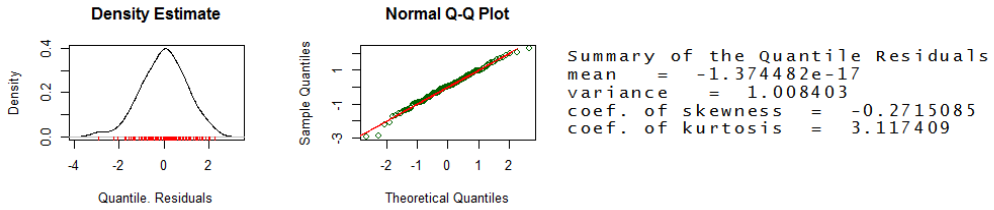


Figure 3-24. The residual distribution of 8-oxodG in warm season: the density estimate with rug plot and the quantile-quantile plot for LOGNO model.

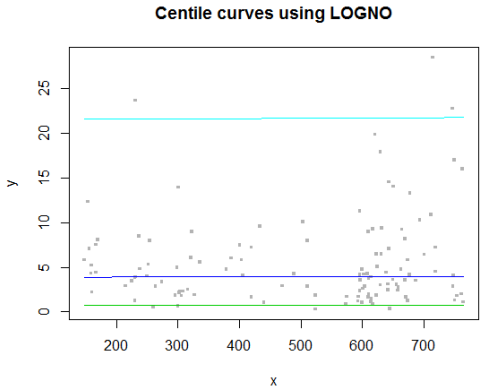


Figure 3-25. Observed 8-oxodG values in warm season with three fitted model centile curves (2.5%, 50%, 97.5%) from LOGNO model against DFC.

The best model for 8-isoprostane during the cold season was the BCCGo model without terms for μ, σ and ν parameters. It allowed reducing BIC criterion by 3.4 units from 169.73 for BCCGo model 'No outliers' in the Table 3-18 to 166.32 in the Table 3-19. The current model provided 2 degrees of freedom for the fit less (df=4, Table 3-19) in comparison with previous 'No outliers' model (df=6, Table 3-18). Quantile residuals distribution was also slightly improved in the current model (Figure 3-26). The 97.5% centile curve had biologically plausible trend (Figure 3-27). The statistically significant negative association was seen in this model between DFC and log(8-isoprstane) (coef.β=-0.002, p=0.038).

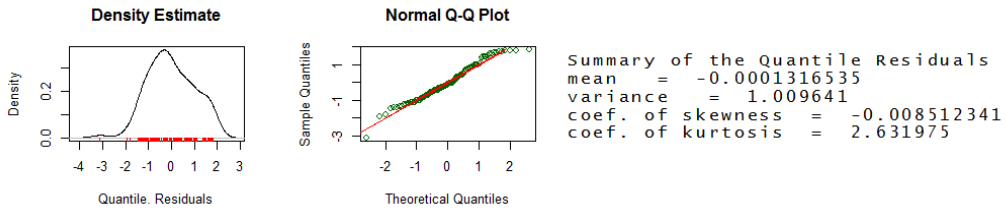


Figure 3-26. The residual distribution of 8-isoprostane in cold season: the density estimate with rug plot and the quantile-quantile plot for BCCGo model.

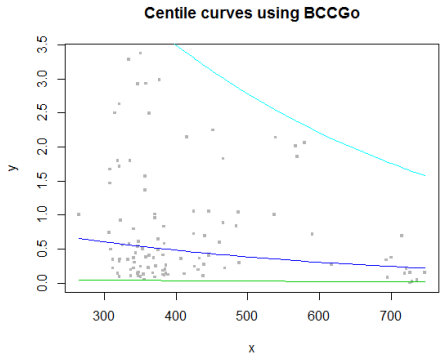


Figure 3-27. Observed 8-isoprostane values in cold season with three fitted model centile curves (2.5%, 50%, 97.5%) from BCCGo model against DFC.

In the warm season for 8-isoprostane the best model had GA distribution with no terms for μ and σ parameters. The BIC criterion was reduced by 6.7 units from 224.3 in the 'No outliers' model (Table 3-18) to 217.49 in the current model (Table 3-19), as well degrees of freedom for the fit by 2 units from df=5 (Table 3-18) to df=3 (Table 3-19). This model had standardized residuals distribution, the coefficient of kurtosis was slightly better (Figure 3-28) in comparison with 'No outliers' model in the Table 3-18, and plausible 97.5% centile curve (Figure 3-29). The statistically significant negative association was found in this model between 8-isoprostane and DFC (coef.β=-0.002; p=0.025)

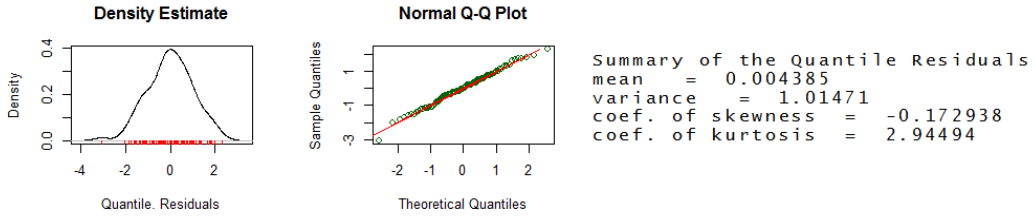


Figure 3-28. The residual distribution of 8-isoprostane in warm season: the density estimate with rug plot and the quantile-quantile plot for GA model.

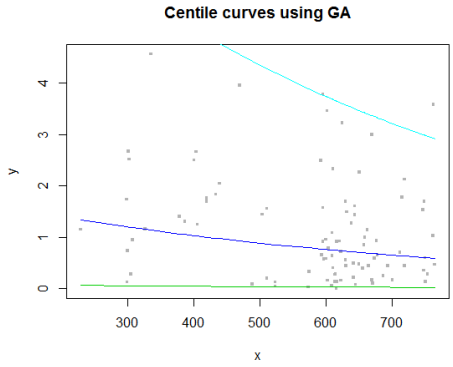


Figure 3-29. Observed 8-isoprostane values in warm season with three fitted model centile curves (2.5%, 50%, 97.5%) from GA model against DFC.

The 95% reference intervals were estimated using the best developed model for each biomarker per each season, per each two weeks of DFC.

Thus during the cold season 8-oxodG values decreased from 1.68 in first two weeks to 0.08 ng/mg$_{creat}$ in last two weeks for lower 2.5% limit of 95% RI and from 28.85 in first two weeks to 14.19 ng/mg$_{creat}$ in last two weeks for upper 97.5% limit of 95% RI (Table 3-20).

During the warm season 8-oxodG values were quite constant: 95% RI = 0.70 - 21.52 ng/mg$_{creat}$ in first two weeks and 95% RI = 0.71 – 21.75 ng/mg$_{creat}$ in last two weeks (Table 3-21).

There was also a decrease of 8-isoprostane values during the cold season: 95% RI = 0.05 – 5.17 ng/mg$_{creat}$ in first two weeks and 95% RI = 0.02 – 1.55 ng/mg$_{creat}$ in last two weeks (Table 3-22).

During the warm season 8-isoprostane values decreased as well: 95% RI = 0.06 – 6.54 ng/mg$_{creat}$ in first two weeks and 95% RI = 0.03 – 2.96 ng/mg$_{creat}$ in last two weeks (Table 3-23).

It should be noticed that 8-isoprostane values were processed in the laboratory starting from DFC = 230 days (or from DFC = 15 weeks).

Table 3-20. Predicted values of 8-oxodG ng/mg$_{creat}$ as a function of DFC during the cold season per each 2 weeks.

DFC		95% Reference Intervals	
Weeks	**Days**	**Lower Limit (2.5%)**	**Upper Limit (97.5%)**
2	36	1.68	28.85
4	51	1.59	28.24
6	66	1.50	27.66
8	81	1.41	27.09
10	96	1.34	26.54
12	111	1.26	26.01
14	126	1.19	25.50
16	141	1.12	25.00
18	156	1.06	24.51
20	171	1.00	24.05
22	186	0.94	23.59
24	201	0.89	23.15
26	216	0.84	22.73
28	231	0.79	22.32
30	246	0.74	21.92
32	261	0.70	21.53
34	276	0.66	21.16
36	291	0.62	20.80
38	306	0.58	20.45
40	321	0.55	20.11
42	336	0.52	19.78
44	351	0.49	19.46
46	366	0.46	19.16
48	381	0.43	18.86
50	396	0.40	18.58
52	411	0.38	18.30
54	426	0.36	18.03
56	441	0.33	17.78
58	456	0.31	17.53
60	471	0.29	17.29
62	486	0.28	17.06
64	501	0.26	16.83
66	516	0.24	16.62
68	531	0.23	16.41
70	546	0.21	16.22
72	561	0.20	16.02
74	576	0.19	15.84
76	591	0.17	15.67
78	606	0.16	15.50
80	621	0.15	15.34
82	636	0.14	15.18
84	651	0.13	15.04

86	666	0.12	14.89
88	681	0.12	14.76
90	696	0.11	14.63
92	711	0.10	14.51
94	726	0.09	14.40
96	741	0.09	14.29
98	756	0.08	14.19

Table 3-21. Predicted values of 8-oxodG ng/mg$_{creat}$ as a function of DFC during the warm season per each 2 weeks.

DFC		**95% Reference Intervals**	
Weeks	**Days**	**Lower Limit (2.5%)**	**Upper Limit (97.5%)**
2	36	0.70	21.52
4	51	0.70	21.52
6	66	0.70	21.52
8	81	0.70	21.53
10	96	0.70	21.53
12	111	0.70	21.54
14	126	0.71	21.54
16	141	0.71	21.55
18	156	0.71	21.55
20	171	0.71	21.56
22	186	0.71	21.56
24	201	0.71	21.57
26	216	0.71	21.57
28	231	0.71	21.58
30	246	0.71	21.58
32	261	0.71	21.59
34	276	0.71	21.59
36	291	0.71	21.60
38	306	0.71	21.60
40	321	0.71	21.61
42	336	0.71	21.61
44	351	0.71	21.62
46	366	0.71	21.62
48	381	0.71	21.63
50	396	0.71	21.63
52	411	0.71	21.64
54	426	0.71	21.64
56	441	0.71	21.65
58	456	0.71	21.65
60	471	0.71	21.66
62	486	0.71	21.66
64	501	0.71	21.67
66	516	0.71	21.67
68	531	0.71	21.68

70	546	0.71	21.68
72	561	0.71	21.69
74	576	0.71	21.69
76	591	0.71	21.70
78	606	0.71	21.70
80	621	0.71	21.71
82	636	0.71	21.71
84	651	0.71	21.71
86	666	0.71	21.72
88	681	0.71	21.72
90	696	0.71	21.73
92	711	0.71	21.73
94	726	0.71	21.74
96	741	0.71	21.74
98	756	0.71	21.75

Table 3-22. Predicted values of 8-isoprostane ng/mg$_{creat}$ as a function of DFC during the cold season per each 2 weeks.

DFC		**95% Reference Intervals**	
Weeks	**Days**	**Lower Limit (2.5%)**	**Upper Limit (97.5%)**
2	230	0.05	5.17
4	245	0.05	5.00
6	260	0.05	4.83
8	275	0.05	4.66
10	290	0.04	4.51
12	305	0.04	4.35
14	320	0.04	4.21
16	335	0.04	4.06
18	350	0.04	3.93
20	365	0.04	3.79
22	380	0.04	3.66
24	395	0.04	3.54
26	410	0.03	3.42
28	425	0.03	3.30
30	440	0.03	3.19
32	455	0.03	3.08
34	470	0.03	2.98
36	485	0.03	2.88
38	500	0.03	2.78
40	515	0.03	2.69
42	530	0.03	2.60
44	545	0.02	2.51
46	560	0.02	2.42
48	575	0.02	2.34
50	590	0.02	2.26
52	605	0.02	2.19

54	620	0.02	2.11
56	635	0.02	2.04
58	650	0.02	1.97
60	665	0.02	1.90
62	680	0.02	1.84
64	695	0.02	1.78
66	710	0.02	1.72
68	725	0.02	1.66
70	740	0.02	1.60
72	755	0.02	1.55

Table 3-23. Predicted values of 8-isoprostane ng/mg$_{creat}$ as a function of DFC during the warm season per each 2 weeks.

DFC		**95% Reference Intervals**	
Weeks	**Days**	**Lower Limit (2.5%)**	**Upper Limit (97.5%)**
2	230	0.06	6.54
4	245	0.06	6.39
6	260	0.06	6.25
8	275	0.06	6.11
10	290	0.06	5.97
12	305	0.06	5.84
14	320	0.06	5.71
16	335	0.06	5.58
18	350	0.05	5.45
20	365	0.05	5.33
22	380	0.05	5.21
24	395	0.05	5.10
26	410	0.05	4.98
28	425	0.05	4.87
30	440	0.05	4.76
32	455	0.05	4.66
34	470	0.05	4.55
36	485	0.04	4.45
38	500	0.04	4.35
40	515	0.04	4.25
42	530	0.04	4.16
44	545	0.04	4.06
46	560	0.04	3.97
48	575	0.04	3.88
50	590	0.04	3.80
52	605	0.04	3.71
54	620	0.04	3.63
56	635	0.04	3.55
58	650	0.03	3.47
60	665	0.03	3.39
62	680	0.03	3.32

64	695	0.03	3.24
66	710	0.03	3.17
68	725	0.03	3.10
70	740	0.03	3.03
72	755	0.03	2.96

4 DISCUSSION

The results of this study were obtained from the Gene Environment Interactions in Respiratory Diseases survey (a part of GEIRD is nested in ECRHS study) that had a multicase-control design and allowed studying all inflammatory diseases and controls simultaneously (51). The GEIRD study produced a multicentre database with phenotypes of asthma, COPD, rhinitis and subjects with no respiratory diseases (controls, which were used in current analysis) that were characterized by their inflammatory and genetic profiles and risk factors simultaneously (51). The phenotyping protocol included the identification protocol, sub-phenotyping protocol and along with many clinical tests, also tests related to oxidative stress and inflammation, such as urine collection for 8-oxodG and 8-isopostane and plasma measurements for glutathione, exhaled nitric oxide, exhaled alveolar air, exhaled breath condensate measurements and induced sputum cell count (51).

Urinary 8-xodG and 8-isoprostane are two sensitive, noninvasive, easy-detectable and quite stable in urine markers of *in vivo* oxidative damages of DNA and lipids and hence of oxidative stress (10, 73). There is still a lack of researches on reference intervals of these two biomarkers of oxidative stress. However the unadjusted RI obtained from adult Italian population in current study were similar to findings of Sakano N. et al (10). Our unadjusted results were also in line with findings of Ogino K et al (46). It should be noticed that it is not always possible to compare obtained results with those presented in literature because of the different techniques used for urine OS biomarkers measurements, i.e. HPLC or ELISA. It was found that ELISA provides higher levels of 8-oxodG in respect to HPLC detection (74). That is also why it was very important to define reference intervals of 8-oxodG and of 8-isoprostane detected with ELISA method.

In this work we departed from unadjusted RI assessment. Both methods, parametric and non-parametric, used for this reason showed similar results of 8-oxodG reference intervals; while they were a bit different by upper (97.5%) limits in 8-isoprostane. As far as 8-isoprostane normal frequency distribution was not achieved with logarithmic transformation, and there was a satisfactory number of subjects in the sample (>120) (34), we suppose the RI obtained with non-parametric approach to be more precise. The unadjusted reference intervals of 8-oxodG, defined in our research were quite similar to those reported by Lily Wu et al. (1), and by Tamura S. et al. (48), and a little bit higher by upper limit of those reported by Ogino et al. (46) and higher by upper limit of RI defined by Andreoli et al. (28) and by Aleksandra Topic et al. (47). Reference interval of 8-isoprostane given in the work of Ogino K et al. (46) was much higher by upper limit than in our work.

It is very important in large epidemiological studies to define explanatory variables which can predict values of biomarkers at statistically significant level, especially when long-term conservation of biological liquid, containing those biomarkers is stipulated. To define main predictors which influence levels of 8-oxodG and 8-isoprostane linear, quantile and GAMLSS regression analyses have been used in this work. All three methods showed total agreement on evaluation of main predictors of OS biomarkers: no age- and gender-related differences, but significant influence of DFC and of season. However, only GAMLSS regression analysis was finally used to derive adjusted reference intervals of both biomarkers. Linear regression analysis was excluded because of residual distribution assumptions not met. Quantile regression approach was rejected because of the lack of diagnostic tools to define the goodness of the fit of the model (75).

Generally, it is considered that oxidative stress is associated with aging (60). Thus Sakano N. et al. (10) found statistically significant difference in 8-oxodG levels in two healthy Japanese age groups less or over 40 years old. Ogino K et al. also showed significantly different 8-oxodG values in healthy Japanese subjects younger and older than 45 years (46). Aleksandra Topic et al. found that 8-oxodG values were lower in younger subjects rather than in older. In our study we did not found a significant difference neither in the 8-oxodG, nor in the 8-isoprostane values associated with age. The group of scientists from Parma (28) investigated healthy Italian subjects and did not find significant association of 8-oxodG values with age as well. Concerning 8-isoprostane, an absence of significant association with age has been proved also by Sakano et al. (10) and Ogino et al. (46).

There are also inconsistent data on OS biomarkers values in gender-related subpopulations. In current work we did not find differences in 8-oxodG and 8-isoprostane values associated with gender. Andreoli et al. (28) showed that gender did not influence 8-oxodG distributions as well. Topic et al. also did not find gender-related differences in 8-oxodG values (47). No difference in 8-oxodG levels between gender groups was observed in the study of Sakano et al. (10). While Lily Wu et al. (1) showed that normal values of 8-oxodG in females were higher than in males. K. Ogino et al. found instead that men had significantly higher levels of 8-oxodG in respect to women (46), while in the same research they showed that 8-isoprostane values were independent of gender. Meanwhile Sakano et al. found higher mean values of 8-isoprostane in men in respect to women (10).

To our knowledge there are no studies on assessment of RI of 8-oxodG and 8-isoprostane associated with such explanatory variable as Season, when urine has been collected. There are some studies on clinical examination of urine change of concentration of OS biomarkers with time, but not on large

epidemiological surveys as in the current study including such variable as DFC. In our study these two predictors proved to be influential on values of both biomarkers, and hence should be considered while constructing reference intervals of both OS biomarkers. These variables can be very important in large longitudinal epidemiological studies, which could require more time for urine conservation before its laboratory elaboration. Such environmental factors as season when urine was collected can influence biomarkers values due to seasonal differences in air temperatures.

Some studies report high stability of 8-oxodG in urine (32, 76, 77), but there is inconsistent data whether its level should decrease or increase with time. Thus Yuki Matsumoto et al. say that the concentration of urinary 8-oxodG can elevate with time because of progressive oxidation (76). However, Shigenaga et al. (73) showed in their research that there is no additional formation of 8-oxodG when urine was stored for 19 days at 4ºC. In our study we found a slight but significant degradation of 8-oxodG in urine with time that was collected during the cold season. When urine was collected during the warm season the concentrations of 8-oxodG remained stable over the entire period.

Some authors report the stability of 8-isoprostane in urine as well (61, 62, 78, 79). But none of these works speaks about long-term, over one year, periods of urine conservation before analysis. In our study we showed slight but statistically significant loss of concentration of this biomarker in urine in entire sample as well as in subsamples stratified by seasons, warm and cold, when urine was collected.

The explanation of decreasing concentrations of biomarkers stored in urine over long periods can be due to their degradation, especially if conserved inappropriately. But further investigations of both urine biomarkers in association with environmental, laboratory and human derived predictors, such as DFC and season are needed to check this hypothesis.

For the estimation of adjusted reference intervals, the GAMLSS regression analysis has been used.

Linear regression was excluded as both biomarkers did not met assumption on normality of residuals distribution.

Quantile regression is a very useful technic as it is ‘distribution free’ regression analysis (75), and allows estimating regression coefficients at quantiles of interest, e.g. at 2.5% and 97.5% needed for assessing adjusted 95% RI. However, quantile regression has a lack of techniques to measure goodness of fit, residual diagnostic plots and statistics for model comparison and adequacy checking, while working with a single statistical model in isolation is not a good practice (75). Summarizing said above, it was decided to reject this approach too.

The GAMLSS regression analysis has a number of advantages and main of them are that it makes assumptions on distribution and provides about 80 distributions for response variable; and that it has platform to fit, compare and check many models (75).

The range of two weeks for estimating reference intervals was chosen based on the medical experience with RI. Thus, one-week reference interval does not show clear difference in values, while one-month reference interval could lose difference in values of biomarker of interest.

To our knowledge it is for the first time when it was shown that both OS biomarkers 8-oxodG and 8-isoprostane should be evaluated in association with DFC and season, when urine has been collected. It is especially important in large epidemiological studies when long-term conservation of urine is stipulated. *(Semi)*parametric GAMLSS regression analysis is a new useful technique that can be used for estimating reference intervals of urinary biomarkers (8-oxodG and 8-isoprostane) from general adult population and adjusted for appropriate determinants.

5 BIBLIOGRAPHY

1. Wu LL, Chiou CC, Chang PY, Wu JT. Urinary 8-OHdG: a marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics. Clin Chim Acta. 2004;339(1-2):1-9.

2. Online Resources for Disorders Caused by Oxidative Stress. Available at: www.oxidativestressresource.org [Accessed February 2017].

3. Tkaczyk J, Vizek M. Oxidative stress in the lung tissue--sources of reactive oxygen species and antioxidant defence. Prague Med Rep. 2007;108(2):105-14.

4. Halliwell B. Role of free radicals in the neurodegenerative diseases: therapeutic implications for antioxidant treatment. Drugs Aging. 2001;18(9):685-716.

5. Park HS, Kim SR, Lee YC. Impact of oxidative stress on lung diseases. Respirology. 2009;14(1):27-38.

6. Rahman I, Biswas SK, Kode A. Oxidant and antioxidant balance in the airways and airway diseases. Eur J Pharmacol. 2006;533(1-3):222-39.

7. Devasagayam TP, Tilak JC, Boloor KK, Sane KS, Ghaskadbi SS, Lele RD. Free radicals and antioxidants in human health: current status and future prospects. J Assoc Physicians India. 2004;52:794-804.

8. Sakano N, Wang DH, Takahashi N, Wang B, Sauriasari R, Kanbara S, et al. Oxidative stress biomarkers and lifestyles in Japanese healthy people. J Clin Biochem Nutr. 2009;44(2):185-95.

9. Ciencewicki J, Trivedi S, Kleeberger SR. Oxidants and the pathogenesis of lung diseases. J Allergy Clin Immunol. 2008;122(3):456-68; quiz 69-70.

10. Sakano N, Takahashi N, Wang DH, Sauriasari R, Takemoto K, Kanbara S, et al. Plasma 3-nitrotyrosine, urinary 8-isoprostane and 8-OHdG among healthy Japanese people. Free Radic Res. 2009;43(2):183-92.

11. Sahiner UM, Birben E, Erzurum S, Sackesen C, Kalayci O. Oxidative stress in asthma. World Allergy Organ J. 2011;4(10):151-8.

12. Bargagli E, Olivieri C, Bennett D, Prasse A, Muller-Quernheim J, Rottoli P. Oxidative stress in the pathogenesis of diffuse lung diseases: a review. Respir Med. 2009;103(9):1245-56.

13. Villegas L, Stidham T, Nozik-Grayck E. Oxidative Stress and Therapeutic Development in Lung Diseases. J Pulm Respir Med. 2014;4(4).

14. Schunemann HJ, Muti P, Freudenheim JL, Armstrong D, Browne R, Klocke RA, et al. Oxidative stress and lung function. Am J Epidemiol. 1997;146(11):939-48.

15. Nagaraj, Priyadarshini MD. Systemic oxidative stress and antioxidant status in Chronic Bronchitis patients. J Sci Innov Res. 2014;3(2):234-8.

16. Ahmad A, Shameem M, Husain Q. Relation of oxidant-antioxidant imbalance with disease progression in patients with asthma. Ann Thorac Med. 2012;7(4):226-32.

17. Kirkham P, Rahman I. Oxidative stress in asthma and COPD: antioxidants as a therapeutic strategy. Pharmacol Ther. 2006;111(2):476-94.

18. Bowler RP, Crapo JD. Oxidative stress in allergic respiratory diseases. J Allergy Clin Immunol. 2002;110(3):349-56.

19. Ochs-Balcom HM, Grant BJ, Muti P, Sempos CT, Freudenheim JL, Browne RW, et al. Oxidative stress and pulmonary function in the general population. Am J Epidemiol. 2005;162(12):1137-45.

20. Niki E, Yoshida Y. Biomarkers for oxidative stress: measurement, validation, and application. J Med Invest. 2005;52 Suppl:228-30.

21. Blumberg J. Use of biomarkers of oxidative stress in research studies. J Nutr. 2004;134(11):3188S-9S.

22. Frijhoff J, Winyard PG, Zarkovic N, Davies SS, Stocker R, Cheng D, et al. Clinical Relevance of Biomarkers of Oxidative Stress. Antioxid Redox Signal. 2015;23(14):1144-70.

23. Berlett BS, Stadtman ER. Protein oxidation in aging, disease, and oxidative stress. J Biol Chem. 1997;272(33):20313-6.

24. Morrow JD, Frei B, Longmire AW, Gaziano JM, Lynch SM, Shyr Y, et al. Increase in circulating products of lipid peroxidation (F2-isoprostanes) in smokers. Smoking as a cause of oxidative damage. N Engl J Med. 1995;332(18):1198-203.

25. Montuschi P, Barnes PJ, Roberts LJ, 2nd. Isoprostanes: markers and mediators of oxidative stress. FASEB J. 2004;18(15):1791-800.

26. Santus P, Corsico A, Solidoro P, Braido F, Di Marco F, Scichilone N. Oxidative stress and respiratory system: pharmacological and clinical reappraisal of N-acetylcysteine. COPD. 2014;11(6):705-17.

27. Topic A. Oxidatively Modified DNA Lesions and Systems for Reparation: sources, clinical significance and methods of analysis. OMICS Group eBooks Available at: http://wwwesciencecentralorg/ebooks/oxidatively/ [Accessed March 2014].

28. Andreoli R, Mutti A, Goldoni M, Manini P, Apostoli P, De Palma G. Reference ranges of urinary biomarkers of oxidized guanine in (2'-deoxy)ribonucleotides and nucleic acids. Free Radic Biol Med. 2011;50(2):254-61.

29. Nakabeppu Y. Cellular levels of 8-oxoguanine in either DNA or the nucleotide pool play pivotal roles in carcinogenesis and survival of cancer cells. Int J Mol Sci. 2014;15(7):12543-57.

30. Poulsen HE, Nadal LL, Broedbaek K, Nielsen PE, Weimann A. Detection and interpretation of 8-oxodG and 8-oxoGua in urine, plasma and cerebrospinal fluid. Biochim Biophys Acta. 2014;1840(2):801-8.

31. Borrill ZL, Roy K, Singh D. Exhaled breath condensate biomarkers in COPD. Eur Respir J. 2008;32(2):472-86.

32. Barregard L, Moller P, Henriksen T, Mistry V, Koppen G, Rossner P, Jr., et al. Human and methodological sources of variability in the measurement of urinary 8-oxo-7,8-dihydro-2'-deoxyguanosine. Antioxid Redox Signal. 2013;18(18):2377-91.

33. Al-Afaleg NO, Al-Senaidy A, El-Ansary A. Oxidative stress and antioxidant status

in Saudi asthmatic patients. Clin Biochem. 2011;44(8-9):612-7.

34. Horn PS, Pesce AJ. Reference intervals: an update. Clin Chim Acta. 2003;334(1-2):5-23.

35. Clinical Laboratory and Standards Institute. “How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline—Second Edition” CLSI document C28-A2, Wayne, PA: 2000.

36. Bertholf RL. Statistical methods for establishing and validating reference intervals. Labmedicine. 2006;37(5):306-10.

37. Jalali N, Babanezhad M. Quantile Regression due to Skewness and Outliers. Applied Mathematical Sciences. 2011;5(39):1947-51.

38. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. Stat Med. 2006;25(8):1369-82.

39. Pan HQ, Goldstein H, Yang Q. Non-parametric estimation of age-related centiles over wide age ranges. Ann Hum Biol. 1990;17(6):475-81.

40. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. ApplStatist. 2005;54(3):507-54.

41. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. JSS. 2007;23(7):1-46.

42. Baraldi E, Carraro S, Alinovi R, Pesci A, Ghiro L, Bodini A, et al. Cysteinyl leukotrienes and 8-isoprostane in exhaled breath condensate of children with asthma exacerbations. Thorax. 2003;58(6):505-9.

43. Evans MD, Dizdaroglu M, Cooke MS. Oxidative DNA damage and disease: induction, repair and significance. Mutat Res. 2004;567(1):1-61.

44. Zanolin ME, Girardi P, Degan P, Rava M, Olivieri M, Di Gennaro G, et al. Measurement of a urinary marker (8-hydroxydeoxy-guanosine, 8-OHdG) of DNA oxidative stress in epidemiological surveys: a pilot study. Int J Biol Markers. 2015;30(3):e341-5.

45. Miwa M, Matsumaru H, Akimoto Y, Naito S, Ochi H. Quantitative determination of urinary 8-hydroxy-2'-deoxyguanosine level in healthy Japanese volunteers. Biofactors. 2004;22(1-4):249-53.

46. Ogino K, Takahashi N, Takigawa T, Obase Y, Wang DH. Association of serum arginase I with oxidative stress in a healthy population. Free Radic Res. 2011;45(2):147-55.

47. Topic A, Francuski D, Markovic B, Stankovic M, Dobrivojevic S, Drca S, et al. Gender-related reference intervals of urinary 8-oxo-7,8-dihydro-2'-deoxyguanosine determined by liquid chromatography-tandem mass spectrometry in Serbian population. Clin Biochem. 2013;46(4-5):321-6.

48. Tamura S, Tsukahara H, Ueno M, Maeda M, Kawakami H, Sekine K, et al. Evaluation of a urinary multi-parameter biomarker set for oxidative stress in children, adolescents and young adults. Free Radic Res. 2006;40(11):1198-205.

49. Rossner P, Jr., Svecova V, Milcova A, Lnenickova Z, Solansky I, Sram RJ. Seasonal variability of oxidative stress markers in city bus drivers. Part I. Oxidative damage to

DNA. Mutat Res. 2008;642(1-2):14-20.

50. Rossner P, Jr., Svecova V, Milcova A, Lnenickova Z, Solansky I, Sram RJ. Seasonal variability of oxidative stress markers in city bus drivers. Part II. Oxidative damage to lipids and proteins. Mutat Res. 2008;642(1-2):21-7.

51. de Marco R, Accordini S, Antonicelli L, Bellia V, Bettin MD, Bombieri C, et al. The Gene-Environment Interactions in Respiratory Diseases (GEIRD) Project. Int Arch Allergy Immunol. 2010;152(3):255-63.

52. European Community Respiratory Health Survey IISC. The European Community Respiratory Health Survey II. Eur Respir J. 2002;20(5):1071-9.

53. de Marco R, Cappa V, Accordini S, Rava M, Antonicelli L, Bortolami O, et al. Trends in the prevalence of asthma and allergic rhinitis in Italy between 1991 and 2010. Eur Respir J. 2012;39(4):883-92.

54. Marcon A, Girardi P, Ferrari M, Olivieri M, Accordini S, Bombieri C, et al. Mild asthma and chronic bronchitis seem to influence functional exercise capacity: a multi-case control study. Int Arch Allergy Immunol. 2013;161(2):181-8.

55. Standardization of Spirometry, 1994 Update. American Thoracic Society. Am J Respir Crit Care Med. 1995;152(3):1107-36.

56. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl. 1993;16:5-40.

57. Sarkar R. Establishment of Biological Reference Intervals and Reference Curve for Urea by Exploratory Parametric and Non-Parametric Quantile Regression Models. EJIFCC. 2013;24(2):61-7.

58. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM. Alternative regression models to assess increase in childhood BMI. BMC Med Res Methodol. 2008;8:59.

59. Hao L, Naiman DQ. Quantile Regression. Thousand Oaks. CA: Sage Publications; 2007.

60. Andriollo-Sanchez M, Hininger-Favier I, Meunier N, Venneria E, O'Connor JM, Maiani G, et al. Age-related oxidative stress and antioxidant parameters in middle-aged and older European subjects: the ZENITH study. Eur J Clin Nutr. 2005;59 Suppl 2:S58-62.

61. Yan W, Byrd GD, Ogden MW. Quantitation of isoprostane isomers in human urine from smokers and nonsmokers by LC-MS/MS. J Lipid Res. 2007;48(7):1607-17.

62. Mizuno K, Kataoka H. Analysis of urinary 8-isoprostane as an oxidative stress biomarker by stable isotope dilution using automated online in-tube solid-phase microextraction coupled with liquid chromatography-tandem mass spectrometry. J Pharm Biomed Anal. 2015;112:36-42.

63. Rigby RA, Stasinopoulos DM. A flexible regression approach using GAMLSS in R (2010). Available from: http://wwwgamlssorg/wp-content/uploads/2013/01/book-2010-Athens1pdf.

64. Quanjer PH, Stanojevic S, Stocks J, Cole TJ. GAMLSS in action: annotated examples of working with R and GAMLSS. Global Lung Function Initiative, 2012. Available from: wwwlungfunctionorg/gamlsshtml.

65. Cole TJ, Stanojevic S, Stocks J, Coates AL, Hankinson JL, Wade AM. Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. Stat Med. 2009;28(5):880-98.

66. Stanojevic S, Wade A, Stocks J, Hankinson J, Coates AL, Pan H, et al. Reference ranges for spirometry across all ages: a new approach. Am J Respir Crit Care Med. 2008;177(3):253-60.

67. Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med. 2006;25(2):247-65.

68. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.

69. Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. WIREs Comput Stat 2012;4:199–203.

70. Rigby RA, Stasinopoulos DM. Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. Statistical Modelling. 2006;6:209-29.

71. Fox J. Applied Regression Analysis and Generalized Linear Models, Second Edition. Sage. . 2008.

72. Fox J, Weisberg S. An R Companion to Applied Regression, Second Edition, Sage. . 2011.

73. Shigenaga MK, Gimeno CJ, Ames BN. Urinary 8-hydroxy-2'-deoxyguanosine as a biological marker of in vivo oxidative DNA damage. Proc Natl Acad Sci U S A. 1989;86(24):9697-701.

74. Pilger A, Rudiger HW. 8-Hydroxy-2'-deoxyguanosine as a marker of oxidative DNA damage related to occupational and environmental exposures. Int Arch Occup Environ Health. 2006;80(1):1-15.

75. Rigby RA, Stasinopoulos DM, Voudouris V. Discussion: A comparison of GAMLSS with quantile regression. Statistical Modelling. 2013;13(4):335–48.

76. Matsumoto Y, Ogawa Y, Yoshida R, Shimamori A, Kasai H, Ohta H. The stability of the oxidative stress marker, urinary 8-hydroxy-2'- deoxyguanosine (8-OHdG), when stored at room temperature. J Occup Health. 2008;50(4):366-72.

77. Loft S, Svoboda P, Kasai H, Tjonneland A, Vogel U, Moller P, et al. Prospective study of 8-oxo-7,8-dihydro-2'-deoxyguanosine excretion and the risk of lung cancer. Carcinogenesis. 2006;27(6):1245-50.

78. Soffler C, Campbell VL, Hassel DM. Measurement of urinary F2-isoprostanes as markers of in vivo lipid peroxidation: a comparison of enzyme immunoassays with gas chromatography-mass spectrometry in domestic animal species. J Vet Diagn Invest. 2010;22(2):200-9.

79. Il'yasova D, Scarbrough P, Spasojevic I. Urinary biomarkers of oxidative status. Clin Chim Acta. 2012;413(19-20):1446-53.

6 APPENDIX I.

Table 6-1. Combination of linear and quadratic polynomials for μ and σ link functions in the LOGNO models for 8-oxodG during the cold and warm seasons and in the GA model for 8-isoprostane during the warm season.

		μ link function (median)		σ link function (variance)	
Number of models per combination	**Total number of models**	**linear**	**quadratic**	**linear**	**quadratic**
		A	**B**	**C**	**D**
		Combination of 4 functions			
1	**1**	A	B	C	D
		Combination of 3 functions			
1	**2**	A	B	C	
2	**3**	A	B		D
3	**4**	A		C	D
4	**5**		B	C	D
		Combination of 2 functions			
1	**6**	A	B		
2	**7**	A		C	
3	**8**	A			D
4	**9**		B	C	
5	**10**		B		D
		1 function			
1	**11**	A			
2	**12**		B		

where empty cells mean an absence of this parameter in a current model

Table 6-2. Combination of linear and quadratic polynomials for μ, σ and ν link functions in the BCCGo model for 8-isoprostane during the cold season.

Number of models per combination	Total number of models	μ link function (median)		σ link function (variance)		ν link function (skewness)	
		linear	quadratic	linear	quadratic	linear	quadratic
		A	**B**	**C**	**D**	**E**	**F**
		Combinations of 6 functions					
1	**1**	A	B	C	D	E	F
		Combinations of 5 functions					
1	**2**	A	B	C	D	E	
2	**3**	A	B	C	D		F
3	**4**	A	B	C		E	F
4	**5**	A	B		D	E	F
5	**6**	A		C	D	E	F
		Combinations of 4 functions					
1	**8**	A	B	C	D		
2	**9**	A	B	C		E	
3	**10**	A	B	C			F
4	**11**	A	B		D	E	
5	**12**	A	B		D		F
6	**13**	A	B			E	F
7	**14**	A		C	D	E	
8	**15**	A		C	D		F
9	**16**	A		C		E	F
10	**17**	A			D	E	F
11	**18**		B	C	D	E	
12	**19**		B	C	D		F
13	**20**		B	C		E	F
14	**21**		B		D	E	F
		Combinations of 3 functions					
1	**22**	A	B	C			
2	**23**	A	B		D		
3	**24**	A	B			E	
4	**25**	A	B				F
5	**26**	A		C	D		
6	**27**	A		C		E	
7	**28**	A		C			F
8	**29**	A			D	E	
9	**30**	A			D		F
10	**31**	A				E	F
11	**32**		B	C	D		
12	**33**		B	C		E	
13	**34**		B	C			F
14	**35**		B		D	E	
15	**36**		B		D		F
16	**37**		B			E	F

		Combinations of 2 functions					
1	**38**	A	B				
2	**39**	A		C			
3	**40**	A			D		
4	**41**	A				E	
5	**42**	A					F
6	**43**		B	C			
7	**44**		B		D		
8	**45**		B			E	
9	**46**		B				F
		1 function					
1	**47**	A					
2	**48**		B				