*Article*

Innovative Approach for Human Semen Quality Assessment Based on Volatilomics

Simonetta Capone [1,*], Angiola Forleo [1], Antonio Vincenzo Radogna [1,2], Valentina Longo [1], Giulia My [1], Alessandra Genga [3], Alessandra Ferramosca [2], Giuseppe Grassi [4], Flavio Casino [1], Pietro Siciliano [1], Tiziana Notari [5], Sebastiana Pappalardo [6], Marina Piscopo [7] and Luigi Montano [8,9,10,*]

[1] National Research Council, Institute for Microelectronics and Microsystems (CNR-IMM), 73100 Lecce, Italy; angiola.forleo@cnr.it (A.F.); antonio.radogna@unisalento.it (A.V.R.); valentinalongo2110@gmail.com (V.L.); my.giulia@imm.cnr.it (G.M.); flavio.casino@cnr.it (F.C.); pietroaleardo.siciliano@cnr.it (P.S.)
[2] Department of Experimental Medicine, University of Salento, 73100 Lecce, Italy; alessandra.ferramosca@unisalento.it
[3] Department of Biological and Environmental Sciences and Technologies, University of Salento, 73100 Lecce, Italy; alessandra.genga@unisalento.it
[4] Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy; giuseppe.grassi@unisalento.it
[5] Reproductive Medicine Unit of Check Up Polydiagnostic Center, 84131 Salerno, Italy; tiziananotari7@gmail.com
[6] Reproduction and Fertility Center, 00128 Rome, Italy; pappalardo@riproduzionefertilita.it
[7] Department of Biology, University of Naples Federico II, 80138 Naples, Italy; marina.piscopo@unina.it
[8] Andrology Unit and Service of Lifestyle Medicine in UroAndrology, "S. Francesco d'Assisi" Hospital, Oliveto Citra, 84020 Salerno, Italy
[9] Coordination Unit of the Network for Environmental and Reproductive Health (EcoFoodFertility Project), "S. Francesco d'Assisi" Hospital, Oliveto Citra, 84020 Salerno, Italy
[10] Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy
* Correspondence: simonetta.capone@cnr.it (S.C.); l.montano@aslsalerno.it (L.M.); Tel.: +39-0832-422507 (S.C.)

Abstract: The volatilome profile of some biofluids (blood, urine, and human semen) identified by Solid-Phase Microextraction–Gas Chromatography/Mass Spectrometry (SPME-GC/MS) and collected from young men living in two high-pollution areas in Italy, i.e., Land of Fires and Valley of Sacco River, have been coupled to sperm parameters obtained by spermiogram analysis to build general multiple regression models. Panels of volatile organic compounds (VOCs) have been selected to optimize the models and used as predictive variables to estimate the different sperm quality parameters (sperm cell concentration, total and progressive motility/immotile cells, total/head/neck/tail morphology anomalies, semen round cell concentration). The results of the multiple linear regression models based on the different subgroups of data joining VOCs from one/two or three biofluids have been compared. Surprisingly, the models based on blood and urine VOCs have allowed an excellent estimate of spermiogram values, paving the way towards a new method of indirect evaluation of semen quality and preventive screening. The significance of VOCs in terms of toxicity and dangerousness was discussed with the support of chemical databases available online.

Keywords: semen quality; human exposome; volatilomics; GC/MS

1. Introduction

Over the last several decades, a hotly debated topic is the meaningful decline of male infertility worldwide [1]. In a meta-analysis, Sengupta P et al. concisely presented evidence of decreased sperm concentration in European males over the past 50 years [2]. Recently, the findings put forth by Levine H et al. [3,4] with a prominent upgraded systematic review and meta-regression analysis reflected that sperm count, as well as other sperm quality parameters, are declining in men not only from North America, Europe, and Australia but also from South/Central America, Africa, and Asia; the warning is that

human semen is globally deteriorating at an accelerate pace, constituting a significant public health problem. The contention is not universally accepted and is motivated by criticisms and limitations found in the reported studies. However, it must be emphasized that the moot point is not the deterioration of human semen quality but rather that this does not necessarily compromise the ability to conceive; hence, it does not automatically translate into a reduction in male fertility [5].

While it is difficult to determine the actual temporal trend in human semen quality in different geographical areas, it is crucial to understand the real impact of non-pathological factors, mainly environmental exposure, and lifestyle factors, on male fertility. The most correct approach to studying how multiple risk factors can undermine the quality of male semen is the human exposome. This is a new paradigm to encompass the totality of human environmental (meaning all non-genetic) exposures from conception onwards [6,7]. Numerous studies report different environmental (air, land, water pollution) and lifestyle factors (nutrients, tobacco, alcohol, and drug use; psychological stress; obesity; insufficient sleep; heat; etc.) that can affect semen quality [8]. However, some studies call for greater rigor in stating statistically significant associations between environmental pollutants and sperm parameters since the population recruited in this study and the methods of assessment of both pollution and semen quality parameters limit the generalizability of the findings [9]. Environment factors include pollution coming from several sources, such as motor vehicle exhaust, factories, fire, household, agriculture, waste treatment, oil refineries, and natural sources; pollution involves particulate matter, volatile organic compounds (VOCs), ozone, nitrogen oxides, sulfur dioxide, carbon monoxide, polycyclic aromatic hydrocarbons (PAHs), and radiations [10]. Agricultural and industrial activities have spread all over the world and are practiced more and more intensively, putting strong and constant pressure on the natural balance of ecosystems, which inevitably become polluted. In addition, the repeated disrespect for environmental protection measures and illegal waste management with the culpability of organized criminals greatly increase the health risk of the human population [11].

In ambient pollution, volatile organic compounds (VOCs) play a relevant role due to their potential hazard character. VOCs are chemicals with a high vapor pressure at room temperature and a standard atmospheric pressure, and they are generally classified by their volatility. VOCs can be divided between biological (endogenous or metabolite VOCs) and anthropogenic sources (exogenous VOCs). The latter are pervasive in daily life and are used in day-to-day activities, including industries, agriculture, office, home, transportation, etc. Though volatile compounds have an indispensable function in manufacturing and maintaining the stability of many products, the health impacts associated with their prolonged exposure cannot be overlooked [12,13]. VOCs are metabolized quickly and yield several toxic metabolites that are excreted in body biofluids. VOCs have, hence, aroused considerable scientific interest in human biomonitoring for studies of exposure to environmental contaminants [14]. Human volatilome, encompassing all the volatile organic compounds (VOCs) found in the human body (the volatilome), together with the related omics science (Volatilomics), are fundamental tools in evaluating the environmental and human health impacts of VOCs [15–18].

A particular interest is devoted to studying how exposure to low doses of VOCs and other contaminants (such as heavy metals, dioxins, plastic contaminants (bisphenols), pesticides and herbicides, phthalates) have adverse effects on reproductive health, mainly semen quality and male fertility [19–21]. Many of them accumulate in the organism, have negative synergic effects, and are endocrine-disrupting environmental pollutants, leading to adverse impacts on semen quality [22,23]. In both epidemiological and experimental studies, it was found that exposure to all these environmental contaminants led to alterations in sperm morphology, sperm motility, sperm count, protamine/histone ratios, sperm nuclear basic proteins (SNBPs), and DNA binding of these proteins [8,24,25]. While our understanding of chronic, low-dose exposures to chemicals is evolving, and the biological mechanisms (germ cell functions, somatic cell functions, reproductive hormone levels/production, hormone

receptors, DNA damage, epigenetic modifications, oxidative stress, and inflammation) by which environmental pollution impair spermatogenesis process and sperm functions are being investigated [26–30], it is evident that such exposures are prevalent risk factors for male infertility.

Recent studies suggest that human semen is an overlooked, early, and sensitive marker of both human health and environmental pollution, and the term "semen sentinel" was coined by the group of Montano L. [31,32], who promoted the EcoFoodfertility project, an Italian pilot biomonitoring initiative [33] to assess environmental impact on human health in risk areas.

This work joins the EcoFoodfertility project, and it is the extension of the previous results published elsewhere [22,34,35]. The volatilome of different human body fluid samples (blood, urine, and semen), collected from barely adult boys living in two contaminated areas in Italy, have been analyzed by two sensing technologies (GC/MS and gas sensor array) as novel tools of human biomonitoring to investigate health risk associated with contaminated sites. The two residential areas considered in this study are contaminated sites in Italy known at the European level, i.e., *Land of Fires* (LF) in Campania region (Italy), whose nickname is due to thousands of illegal wastes burned to eliminate traces, and *Valley of Sacco River* (VSR) in Lazio region (Italy), where toxic waste dumps of industrial origin have been poured into the river and used in agriculture and livestock breeding for decades, causing an unprecedented environmental and social disaster.

In this work, we performed an advanced data analysis step where volatilome profiles of blood, urine, and semen samples collected from the above-mentioned pilot study's male population and analyzed by Headspace–Solid-Phase Microextraction–Gas Chromatography/Mass Spectrometry (HS-SPME-GC/MS) were supported by seminogram analysis carried out on semen samples. Multiple regression models were applied to the data to evaluate the ability of the VOC patterns to predict some seminogram parameters related to sperm count, sperm motility, and sperm morphology.

2. Materials and Methods

2.1. Subjects

In this study, conducted within the EcoFoodfertility project, a small male population of students (n. 50 subjects in total) of about 18 years old residing in *Land of Fires* (LF) (n. 35 subjects) and *Valley of Sacco River* (VSR) (n. 15) were recruited as volunteers, and their biospecimens were collected (human semen, blood, and urine) after obtaining their informed consent. This study was carried out in accordance with the Code of Ethics of the World Medical Association [36] upon approval of the Ethical Committee of the Local Health Authority Campania Sud-Salerno (protocol number 43_r.p.s.o., 30 June 2015). To exclude special conditions of occupational exposure and consider only general conditions of environmental exposure of the population, only student boys were enrolled. Exclusion factors from this study were the habit of smoking or alcohol consumption, chronic diseases (diabetes or other systemic diseases), malfunctions of the reproductive system (varicocele, prostatitis, etc.), and a high body mass index (BMI > 33 kg/m^2). Recruitment was carried out at the "San Francesco d'Assisi" Hospital (Oliveto Citra, Salerno, Italy) and the Italian Association of Blood Volunteers (AVIS, Frosinone office, Frosinone, Italy) in 2018.

2.2. Sample Collection

Human semen was collected via morning masturbation after 3–4 days of sexual abstinence in sterile containers. Morning fasting blood samples were collected via venipuncture into sodium citrate tubes and shaken gently. The first morning urine samples, also called 8 h samples, were collected after emptying the bladder before going to sleep; each subject, therefore, provided such a urine sample in a sterile 50 mL PVC container. In a subsequent aliquoting phase, aliquots of the different biofluids (250 µL of ejaculated sperm, 1 mL of whole blood, and 1 mL of urine) were taken from standard collection containers and transferred into 5 mL headspace vials (code 27319-U, Shimadzu™, Kyoto, Japan) capped with a

screw cap assembled with a hole with a PTFE/silicone septum. The vials were immediately frozen and stored at −80 °C and were sent in batches to the Gas Sensor Lab of the CNR-IMM in Lecce (Italy) for joint analysis of VOCs using gas sensors and Headspace–Solid Phase Microextraction–Gas Chromatography/Mass Spectrometry (HS-SPME-GC/MS).

2.3. VOC Analysis

Following a standardized experimental protocol for the treatment of human semen, blood, and urine samples, VOCs were extracted from the headspace of the vials containing the different biological fluids. The frozen samples were thawed at room temperature, and then the vials were immersed in a water bath on a hot plate with a magnetic stirrer at 60 °C overnight. After this incubation, a Carboxen®/Polydimethylsiloxane (CAR/PDMS) fiber (57318, Supelco, Bellefonte, PA, USA) was exposed to each sample headspace for 15 min. After the SPME step, the GC/MS analysis of extracted volatiles was performed using GC (6890N series Agilent Technologies, Santa Clara, CA, USA) coupled to MS (5973 series Agilent Technologies, USA) equipped with a ZB-624 capillary column (Phenomenex, Torrance, CA, USA); the SPME fiber was injected into the GC injector and was kept at 250 °C to allow thermal desorption of VOCs. A flow rate of 1 mL/min for the helium carrier gas was used with the following GC temperature program: initial 34 °C held for 2 min and then ramped at 3 °C/min to 110 °C; after that, 5 °C/min to 220 °C held for 2 min. The MS analyses were carried out in full-scan mode with a scan range of 30–500 amu at 3.2 scans/s. The VOCs were searched by non-target analysis, identified by comparing mass spectra with those of the data system library (NIST14, P > 60%), and quantified by the internal standard (I.S.: 1,4-Dichlorobenzene-D4, Sigma Aldrich, St. Louis, MO, USA) method. Details of the recruitment of subjects, sample collection, and VOC analysis by HS-SPME-GC/MS on the different biospecimens have been just published [34,35].

2.4. Semen Analysis

Seminogram analysis was carried out on human semen samples collected in sterile containers following the last guidelines and procedures for the examination and processing of human semen given by World Health Organization criteria [37].

A semen analysis or sperm test, also known as a seminogram (or spermogram), is a non-invasive test whose purpose is to evaluate certain characteristics of a male's semen and the sperm contained therein, allowing for an initial, overall assessment of a man's fertility based on his semen quality. Thanks to it, the specialist can assess various macroscopic and microscopic parameters. For this study, we used the microscopic data of sperm concentration, motility (percentage of total and progressive motility and immotiles), morphological abnormalities (percentage of total, head, neck, and tail abnormalities), and round cell concentration. The most modern method of semen analysis is the computer-assisted method—the SCA system (Sperm Class Analyzer), which provides fast, accurate, modern, and objective analysis of the key parameters; a phase contrast microscope (Nikon Eclipse TE 300, Tokyo, Japan) was used. The SCA system meets the World Health Organization's criteria (WHO21) [37–39]. WHO21 has established reference values to determine what the normal results should be for a semen analysis report.

According to WHO21 criteria, sperm concentration is the number of sperm per milliliter, and it is expected to be 16 million or more per milliliter.

Sperm are motile cells, and how they move is important. Progressive motility is the spermatozoa's ability to progress in their advance and, therefore, traverse the fallopian tubes and reach the egg; the spermatozoa must show a progressive motility of 30% or more. Total motility refers to the percentage of sperm making any sort of movement, including non-progressive movements that occur when spermatozoa do not make a forward progression or swim in very tight circles. At least 42% of the sperm should be motile or moving; below 42% is considered low sperm motility. Immotile spermatozoa are unable to move in any way.

Sperm morphology, i.e., the size and shape of sperm, is another factor that is examined as part of a semen analysis; it is reported as the percentage of sperm that appear normal when semen is viewed under a microscope. Normal sperm have an oval head with a long tail. Abnormal sperm have head, neck, or tail defects, such as a large or misshapen head, a bent or irregular neck, or a crooked or double tail. These defects might affect the ability of the sperm to reach and penetrate an egg. However, having a large percentage of misshapen sperm is not uncommon. Typically, only around 4% to 14% of the sperm in a semen sample are normal, meaning that the vast majority do not look perfect under a microscope. The seminogram must show normal sperm forms greater than 4%, which means that the percentage of morphological anomalies must not exceed 96%.

Round cells in seminal fluid are defined as either white blood cells (leucocytes) or immature germ cells (precursors of sperm cells). The presence of leukocytes indicates the presence of an infection, which could lead to alterations in other parameters. Immature germ cells are normally found in the testicles, but sometimes they can be released into the semen. They can be harmless, but they can also indicate problems with sperm maturation or production. They can also affect sperm concentration and morphology [40]. The World Health Organization (WHO21) recommends estimating the concentration of round cells in semen analysis as part of the basic parameters. The normal values are less than 5 million round cells per milliliter of semen and less than 1 million leucocytes per milliliter of semen.

2.5. Statistical Data Analysis

Data obtained by both semen and VOC analysis are summarized with descriptive statistics. In an advanced data analysis step, spermiogram parameters and VOCs found in the considered body fluids were processed using statistical regression analysis techniques. The 9 spermiogram parameters (i.e., spermatozoa concentration in 10^6 cells/mL, progressive motility/total motility/immotiles in percentage, total/head/neck/tail morphological anomalies in percentage, and the round cell concentration in 10^6 cells/mL) determined by semen analysis were thus used as reference data in multivariate regression modeling, building a predictive model from the set of independent or predictor variables (here, the patterns of the VOC are expressed in ng/mL) to the set of continuous dependent variables (here, the spermiogram parameters). In particular, multiple linear regression (MLR) was used as a multiple regression technique to estimate the relationship between two or more independent variables and one dependent variable. We used it to determine (a) how strong the relationship is between the predictor variables (VOC concentration) and each dependent variable (spermiogram parameters) and (b) the predicted value of the dependent variable at certain values of the independent predictor variables.

All the statistical data analyses were carried out by Statistics for Data Analysis (SPSS v. 29.0.1) and OriginPro 8.6.

3. Results

3.1. Descriptive Statistics of Male Population Seminograms

The data obtained by semen analysis carried out on the sample population living in *Land of Fires* (LF) and *Valley of Sacco River* (VSR) were first explored by descriptive statistics. Box and whisker plots have been used for all the seminogram parameters as a graphical representation to describe data distribution and skewness in both subgroups (LF, VSR) and whole (all = LF + VSR) sample population through simple dispersion and position indices. Box and whisker plots displaying the dispersion of sperm count and motility are shown in Figure 1, whereas those displaying sperm anomalies and round cell concentrations are shown in Figure S1. The box indicating the interquartile range (IQR, middle 50% of our population) within the 25–75% quartiles, the central tendency (median and mean), the whiskers extending to 1.5 IQR, and the outliers have been computed for each group of cases. In the plots for each seminogram parameter, the reference lines (blue lines) indicate the threshold value above or below which the parameter is in the normal range.
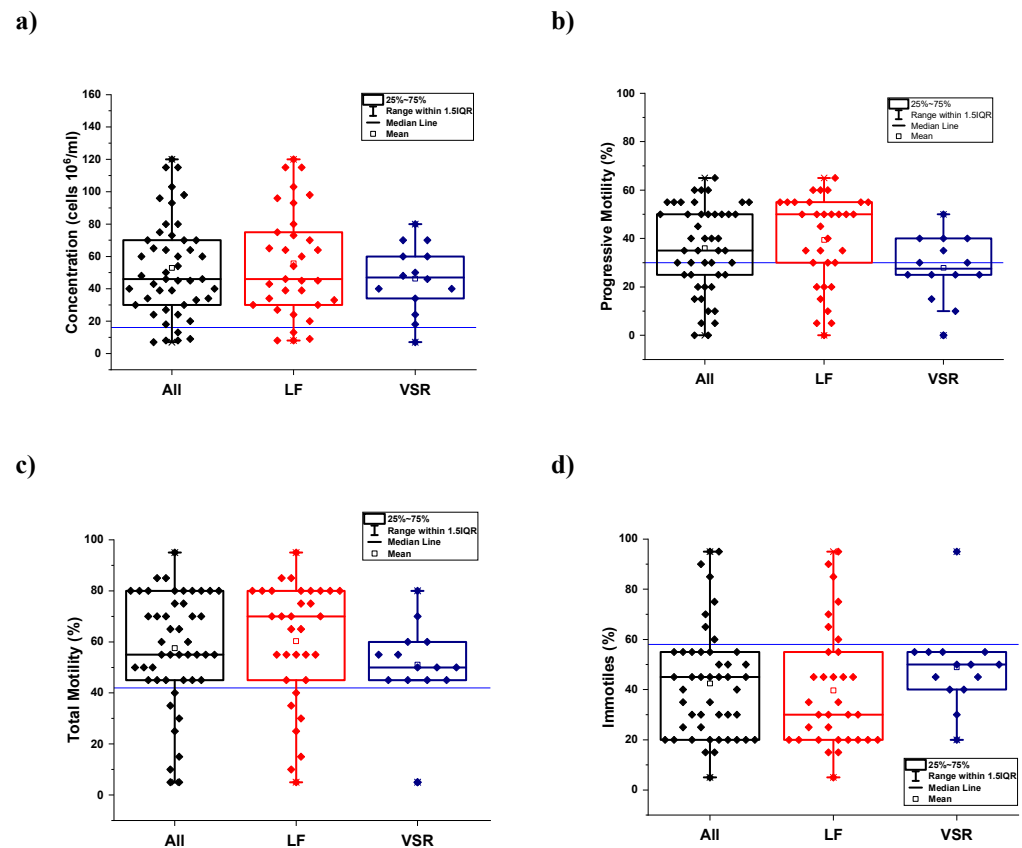
Figure 1. Box and whisker plots displaying the dispersion of the seminogram parameters related to sperm count and motility: (**a**) sperm cell concentration; (**b**) progressive motility; (**c**) total motility; (**d**) immotile cells in the subgroups of the population living in *Land of Fires* (LF) and in *Valley of Sacco river* (VSR) and the total population (all = LF + VSR).

In all box and whisker plots, it can be observed that although most of the data fall within normal ranges, there is a non-negligible part of data outside these ranges, which constitutes an alert factor.

Some cases (subjects) show a sperm count below the threshold value of 16 million cells per milliliter (Figure 1a). Regarding motility parameters, many observations are below the reference line; in the box and whisker plot of the sperm progressive motility that must be 30% or more, the reference line crosses the box plot, and a significant part of cases are below this threshold value (Figure 1b). Regarding sperm total motility, which also includes non-progressive movements and the complementary immotile spermatozoa, part of the sample population shows a sperm total motility below the cutoff value of 42% (Figure 1c) and, correspondingly, a percentage of immotile cells above 58% (Figure 1d).

Sperm morphological anomalies in the sample population have a high percentage but for the majority of cases, they are below the tolerable value of 96% (total sperm anomalies), and in a smaller but not insignificant part of cases they exceed this value. However, to evaluate semen sample quality, some laboratories still use Kruger's strict criteria (WHO99) [41], which is far stricter than the ones published by WHO10 [42] and WHO21 [37]. According to Kruger's strict criteria, teratozoospermia is present when more than 86% of spermatozoa have an abnormal shape. In other words, the borderline to consider whether a man has teratozoospermia is set at a minimum of 14% of spermatozoa with normal forms instead of 4%, which is the percent currently established by WHO21 [39,43]. It is significant that almost all our samples exceed the limit set by the Kruger criterion (Figure S1a).

It can be also observed that among the different types of morphological sperm anomalies, those present in the highest percentage are those of the head (Figure S1b), followed by those of the neck (Figure S1c) and tail (Figure S1d). Sperm pathologies primarily affecting

the head–neck and secondary tail impair sperm function; on a cellular basis, the hidden causes behind a head shape change play a significant role in fertilization.

Furthermore, a significant portion of the sample population was found to have a concentration of round cells equal to or greater than the limit of 5 million per milliliter set by the WHO21 criteria (Figure S1e). Although the size of the sample population is small, what emerges from this sperm characterization of boys residing in the two contaminated areas considered in this study is that there is a fraction of the sample population with poor sperm quality and not within the limits recommended by the World Health Organization. What is of concern in these data is the young age of the recruited subjects, as all boys are just over 18 at the peak of their reproductive capacity; one would not expect the first signs of deterioration of sperm parameters from them. In line with this experimental evidence, other studies on the same contaminated areas have recorded molecular alterations in the sperm of young people and metal contamination [25,44].

3.2. Data Analysis by Regression Methods

In this study, the joint use of GC/MS and seminograms gave the opportunity to relate the volatilome fingerprinting of the considered biosamples in terms of the most significant VOCs identified in the headspace of the considered biosamples (blood, urine, and semen) to the corresponding seminogram parameters. Specifically, 42 statically relevant VOCs identified and quantified by the HS-SPME-GC/MS analysis were combined with seminogram parameters. The rationale for this study is that the VOC profile of blood (B), urine (U), and sperm (HS), which essentially represents the health status of a subject based on his internal and external exposome, can be indicative not only of pathological conditions (applicable in diagnostics) but also of multiple environmental exposures that can contribute to compromising the general quality of sperm. Human exposome and human volatilome are strictly interconnected, and their hidden relationship could be exploited both to discriminate populations living in areas at high vs. low pollution impacts as the index of risk for health and to indirectly estimate clinical or physiological parameters. Looking at this last potential use, we pushed ourselves to evaluate the feasibility of the predictive potential of specific VOC patterns of human biospecimens towards variables of interest, such as spermiogram parameters, using multiple linear regression (MLR) analysis.

According to the VOC distribution in human semen, blood, and urine, some VOCs are common to all three biofluids (HS, B, and U) and others to two of them, whereas others are only present in one biofluid; thus, the forty-two statistically relevant VOCs selected by HS-SPME-GC/MS give rise to twenty-two VOC variables for HS, twenty-nine for B, and twenty-six for U. Multiple linear regression (MLR) analysis was hence applied to all nine spermiogram parameters used as dependent variables. Initially, all VOC variables were used as predictor variables in the MLR analysis, and different models were built, depending on which biofluids (1 or 2 or all 3) were included in the data analysis, i.e., human semen and blood and urine (HS + B + U), human semen and blood (HS + B), human semen and urine (HS + U), human semen only (HS), blood only (B), and urine only (U).

In model building, all effects (forward stepwise, backward stepwise, forward entry, backward removal, and best subsets) were tested. In each model and for each spermiogram parameter, a Pareto chart of the t-values for the regression coefficients was used as a guide for selecting the optimal predictors. The Pareto chart shows the parameter estimates (or t-values) sorted by their absolute size. On the Pareto chart, bars that cross the reference line (corresponding to the current criterion of statistical significance $p < 0.05$) are statistically significant; by working on retaining only the most significant predictive variables, a more performing regression model was refined based on the patterns of those VOCs that contribute more to the predictive model. We list all the VOCs that were found to be significant as predictive variables in the statistical regression analysis (VOC predictors) in a

supplementary table (Table S1) with the values relating to the level of concentration and distribution in the sample population.

It is worth noticing that in this way, starting from volatilome characterization of human biofluids, such as blood, urine, and semen, novel VOC-based models for the estimation of the spermiogram parameters were developed.

Figures 2a, 3a, 4a, 5a and S2a–S6a show the overall fit of the MLR model described by the test of the SS whole model vs. SS residual for all the nine spermiogram parameters estimated by the VOC-based model with a significant *p*-value ($p < 0.05$).

(a)

Concentration (10^6 cells/ml)	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,988824	0,977774	0,927764	12898,02	9	1433,113	293,194	4	73,2984	19,551760	0,005803
HS + B	0,970233	0,941352	0,847516	24610,99	16	1538,187	1533,303	10	153,3303	10,031850	0,000388
HS + U	0,969477	0,939886	0,862596	15282,99	9	1698,110	977,484	7	139,6405	12,160580	0,001679
B + U	0,989074	0,978268	0,907639	15121,03	13	1163,156	335,9117	4	83,97793	13,85074	0,010628
HS	0,495168	0,245191	0,194870	7428,854	2	3714,427	22869,39	30	762,3130	4,872575	0,014708
B	0,928148	0,861459	0,736772	29565,19	18	1642,510	4754,710	20	237,7355	6,908982	0,000039
U	0,880287	0,774905	0,669861	15263,00	7	2180,429	4433,608	15	295,5738	7,376934	0,000622

(b)

c) VOC predictors

1) 2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine_B
2) 2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-_U
3) Octane_B
4) Acetone_U
5) 4-(4-Chlorophenyl)-2,6-diphenylpyridine_B
6) 2-Pentanone_U
7) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro -5-[3-(methylamino)propyl]-_B
8) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
9) 3-Hexanone_U
10) 3-Aminopyrrolidine_B
11) Propanal, 2-methyl-_U
12) Auramine_B
13) 2-Ethyl-oxetane_U

Figure 2. (**a**) Overall fit of the MLR model described by the test of the SS whole model vs. SS residual for sperm concentration (in 10^6/mL); (**b**) observed vs. predicted values for sperm concentration as a result of the MLR analysis for the data group B + U based on selected VOC predictors; (**c**) pattern of VOCs used as predictor variables in the MLR model as selected by a Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X = B, blood, U, urine, or HS, human semen). Notation for numeric values: comma "," is the decimal separator (SI).

Each table reports the results for a predicted spermiogram parameter comparing the different groups of data (HS + B + U; HS + B; HS + U; B + U; HS; B; U); specifically, the multiple R, R^2, adjusted R^2, and overall F-test results were displayed for each group of data. The multiple R is the coefficient of multiple correlations, which is the positive square root of R-square (the coefficient of multiple determination); the R^2 value is an indicator of how well the model fits the data (e.g., an R^2 close to 1.0 indicates that we have accounted for almost all of the variability with the variables specified in the model); and the adjusted R^2 is interpreted similarly to the R^2 value, except the adjusted R^2 takes into consideration the number of degrees of freedom. The overall F-test determines whether the relationship between the dependent variable and the set of independent variables is statistically significant; if the *p*-value for the overall F-test is less than the significance level, you can conclude that the R^2 value is significantly different from zero.

(a)

Progressive motility %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998162	0,996328	0,988065	4319,79	9	479,977	15,921	4	3,9804	120,586400	0,000165
HS + B	0,994035	0,988107	0,948462	25833,35	20	1291,667	310,947	6	51,8245	24,923900	0,000342
HS + U	0,949293	0,901157	0,856229	4304,35	5	860,870	472,119	11	42,9199	20,057590	0,000034
B + U	0,833066	0,694000	0,599846	4148,58	4	1037,144	1829,2020	13	140,70780	7,37091	0,002503
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,900439	0,810791	0,687393	10912,41	15	727,494	2546,562	23	110,7201	6,570570	0,000035
U	0,914171	0,835708	0,741827	4957,929000	8,000000	619,741100	974,679900	14,000000	69,619990	8,901769	0,000244

(b)

c) VOC predictors
1) 1H-Indole, 5-methyl-2-phenyl-_U
2) 3-Hexanone_U
3) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
4) Aminopyrrolidine_U
5) 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-_U
6) Butanal, 3-methyl-_U
7) 2,4,5-Trioxoimidazolidine_U
8) Pyrrole_U

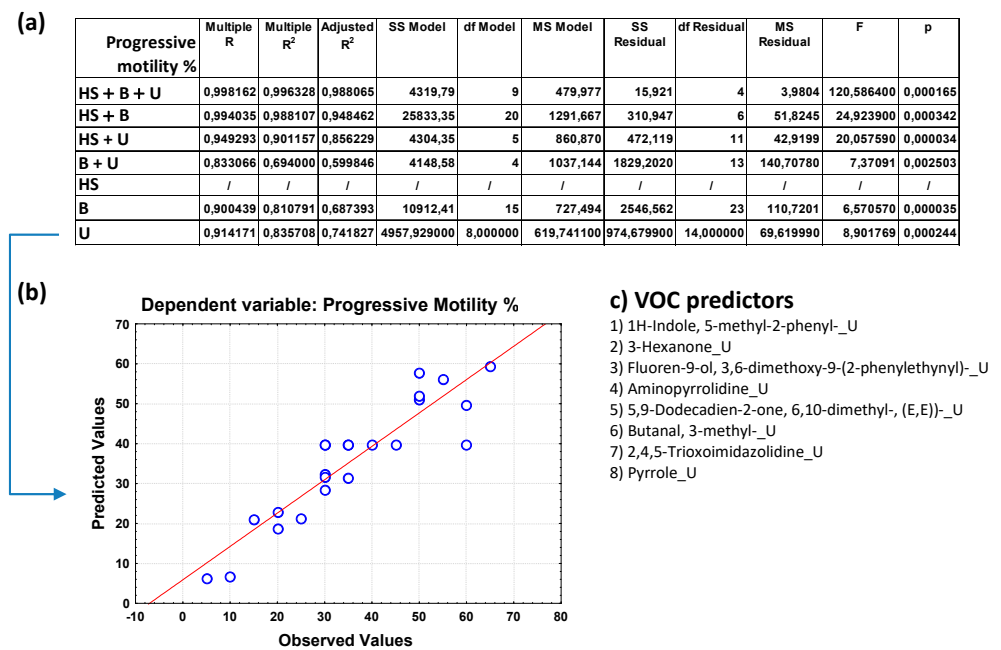**Figure 3.** (**a**) Overall fit of the MLR model described by the test of the SS whole model vs. SS residual for sperm progressive motility (in percentage); (**b**) observed vs. predicted values for sperm progressive motility as a result of the MLR analysis for the data group U based on selected VOC predictors; (**c**) pattern of VOCs used as predictor variables in the MLR model as selected by a Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X = B, blood, U, urine, or HS, human semen). Notation for numeric values: comma "," is the decimal separator (SI).

(a)

Total Motility %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998866	0,997733	0,992633	5368,16	9	596,462	12,196	4	3,0491	195,619100	0,000063
HS + B	0,974892	0,950414	0,892564	10746,72	14	767,623	560,689	12	46,7241	16,428840	0,000011
HS + U	0,982579	0,965462	0,907900	6761,08	10	676,108	241,865	6	40,3108	16,772360	0,001299
B + U	0,970786	0,942426	0,891250	6184,67	8	773,084	377,8270	9	41,98078	18,41519	0,000102
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,944713	0,892483	0,795717	17751,25	18	986,181	2138,491	20	106,9246	9,223145	0,000004
U	0,985701	0,971606	0,921917	7870,01	14	562,144	229,990	8	28,7487	19,553710	0,000119

(b)

c) VOC Predictors
1) 1H-Indole, 5-methyl-2-phenyl-_U
2) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
3) Hexanal_U
4) 2,4,5-Trioxoimidazolidine_U
5) 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-_U
6) Propanal, 2-methyl-_U
7) n-Hexane_U
8) 4-Heptanone_U
9) 3-Aminopyrrolidine_U
10) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_U
11) 2-Anthracenamine_U
12) Disulfide, dimethyl_U
13) 4-(4-Chlorophenyl)-2,6-diphenylpyridine_U
14) Oxime-, methoxy-phenyl-__U

Figure 4. (**a**) Overall fit of the MLR model described by the test of the SS whole model vs. SS residual for sperm total motility (in percentage); (**b**) observed vs. predicted values for sperm total motility as a result of the MLR analysis for the data group U based on selected VOC predictors; (**c**) pattern of VOCs used as predictor variables in the MLR model as selected by a Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X = B, blood, U, urine, or HS, human semen). Notation for numeric values: comma "," is the decimal separator (SI).
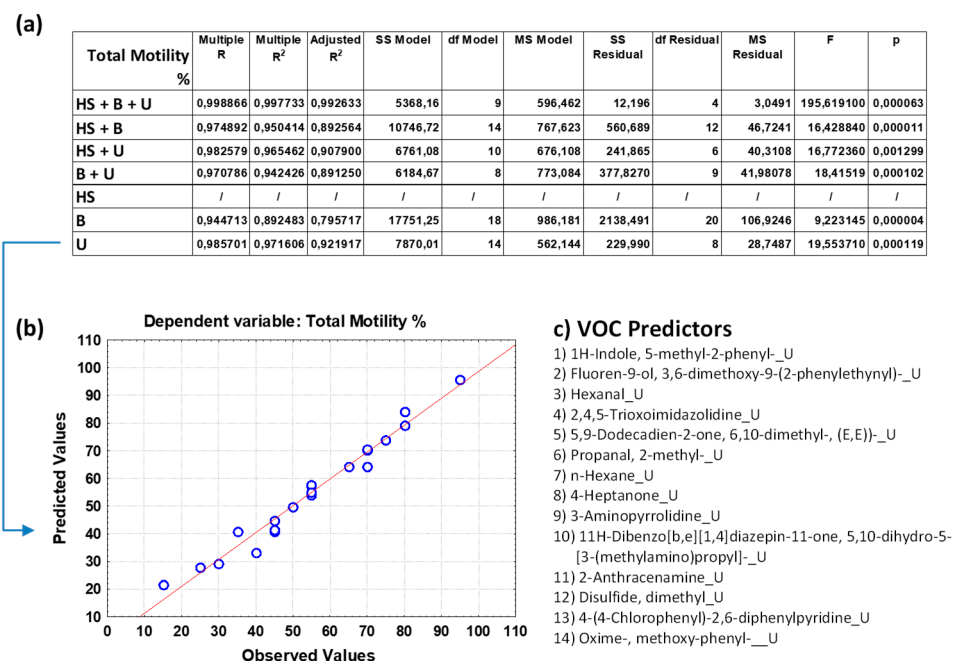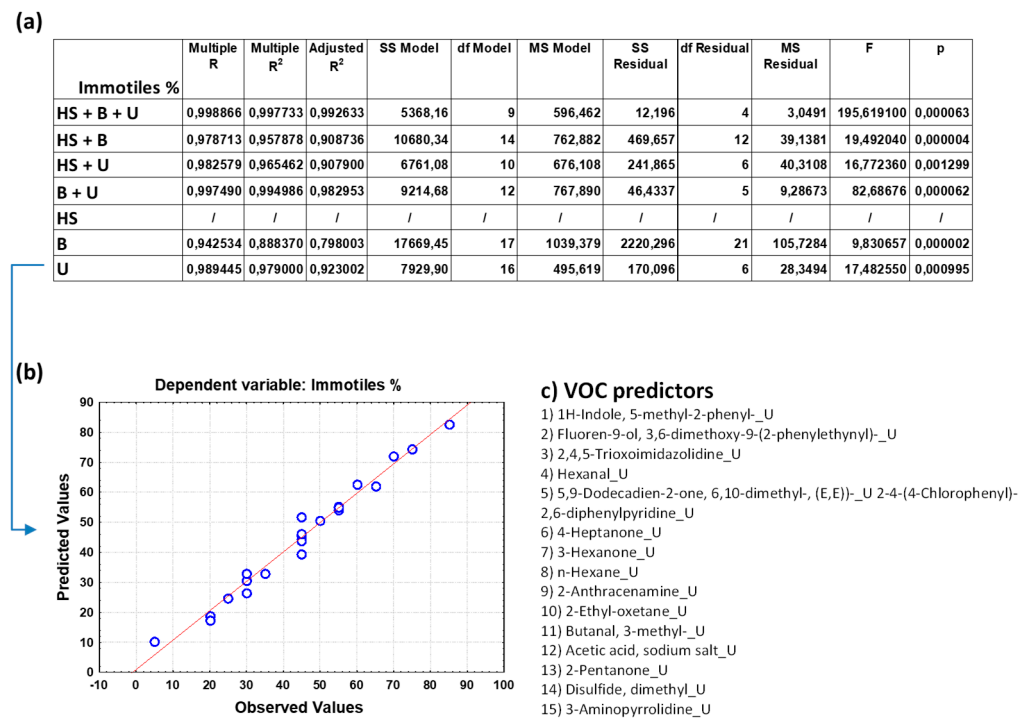
(a)

Immotiles %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998866	0,997733	0,992633	5368,16	9	596,462	12,196	4	3,0491	195,619100	0,000063
HS + B	0,978713	0,957878	0,908736	10680,34	14	762,882	469,657	12	39,1381	19,492040	0,000004
HS + U	0,982579	0,965462	0,907900	6761,08	10	676,108	241,865	6	40,3108	16,772360	0,001299
B + U	0,997490	0,994986	0,982953	9214,68	12	767,890	46,4337	5	9,28673	82,68676	0,000062
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,942534	0,888370	0,798003	17669,45	17	1039,379	2220,296	21	105,7284	9,830657	0,000002
U	0,989445	0,979000	0,923002	7929,90	16	495,619	170,096	6	28,3494	17,482550	0,000995

(b)

c) VOC predictors

1) 1H-Indole, 5-methyl-2-phenyl-_U
2) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
3) 2,4,5-Trioxoimidazolidine_U
4) Hexanal_U
5) 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-_U 2-4-(4-Chlorophenyl)-2,6-diphenylpyridine_U
6) 4-Heptanone_U
7) 3-Hexanone_U
8) n-Hexane_U
9) 2-Anthracenamine_U
10) 2-Ethyl-oxetane_U
11) Butanal, 3-methyl-_U
12) Acetic acid, sodium salt_U
13) 2-Pentanone_U
14) Disulfide, dimethyl_U
15) 3-Aminopyrrolidine_U

Figure 5. (**a**) Overall fit of the MLR model described by the test of the SS whole model vs. SS residual for sperm immotiles (in percentage); (**b**) observed vs. predicted values for sperm immotiles as a result of the MLR analysis for the data group U based on selected VOC predictors; (**c**) pattern of VOCs used as predictor variables in the MLR model as selected by a Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X = B, blood, U, urine, or HS, human semen). Notation for numeric values: comma "," is the decimal separator (SI).

The comparison of the overall fit of the multiple regression model between the different data groups (Figures 2a, 3a, 4a, 5a and S2a–S6a) allows us to highlight some surprising results and make some considerations on the great potential of the analysis of VOCs in human biofluids. Although the MLR model that considers the trio of biofluids (HS + B + U) generally has a better fit for most sperm parameters than regression models that consider only a duo of biofluids or those that consider only one, the predictive power of the latter remains very high. Unexpectedly it can be observed that the HS data group is not efficient in estimating sperm concentration, whereas when combined with B and/or U data, the predictive power also increases with the contribution of the VOCs found in HS. Attention is focused on the binary model that considers blood and urine (B + U) data and on the single models that consider blood (B) data and urine (U) data. Although these models exclude the VOC variables of the human semen from the multiple regression analysis, they unexpectedly allow us to estimate the value of the different sperm parameters with a good fitting. In particular, the B + U data group can be considered sufficient to evaluate sperm concentration with a high multiple R coefficient for most of the spermiogram parameters, as evident from good linear data distribution in the graph of observed vs. predicted values. However, some sperm parameters are better or equally estimated by the U data group compared to the B + U data group.

To give greater emphasis to the application potential of these results, we decided to show the predicted vs. observed graphs resulting from the MLR analysis based on the U data subgroup for the following six spermiogram parameters (total motility, progressive motility, immotile cells, head and neck anomalies, and round cells, respectively, in Figures 3b, 4b, 5b, S3b, S4b and S6b) and those obtained with the B + U data subgroup for the remaining three spermiogram parameters (cells concentration, morphology anomalies, and tail anomalies, respectively, in Figures 2b, S2b and S5b). Contextually, for the mentioned MLR analysis based on the specific U and B + U data subgroups for the different

sperm parameters, Figures 2c, 3c, 4c, 5c and S2c–S6c report the corresponding pattern of VOCs used as independent predictive variables in the regression model as selected by a Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X = B, blood, U, urine, or HS, human semen). The detailed MLR results related to the HS + B + U dataset are shown in Figures S7–S15; the predictive ability of the different spermiogram parameters is constructed with the contribution of VOCs found in the three biofluids.

4. Discussion

The role of the VOCs found in the biofluids considered in the work and used as predictive variables in the regression model was explored in depth on the basis of some databases that collect the VOC compounds present in the human volatilome: the Human Metabolomic Database (HMDB version 5.0) [45,46]; the CompTox Chemicals Dashboard v2.3.0 [47,48] by the United States Environmental Protection Agency (EPA); the Comparative Toxicogenomics Database [49]; and the PubChem Compound database [50].

It is worth arguing that assessing chemicals for their potential to cause male reproductive toxicity involves the evaluation of evidence obtained from experimental, epidemiological, and mechanistic studies. For male reproductive toxicants, eight key characteristics were identified based on a survey of established mechanisms and include alterations in (1) germ cell functions, (2) somatic cell functions, (3) reproductive hormone levels/production, (4) hormone receptors, (5) DNA damage, (6) epigenetic modifications, (7) oxidative stress, and (8) inflammation. All these pathways contribute to the pathogenesis of impaired male fertility potential [27–30,51,52].

Regarding sperm concentration, we can comment on the contribution to parameter estimation given by blood and urinary VOC predictors (Figure 2c). First, a main contribution is given by pyridine derivates and pyrimidines. Pyridine is a compound considered carcinogenic; it is often used as a denaturant for antifreeze mixtures, ethyl alcohol, and fungicides, as well as a dyeing aid for textiles. It is a harmful substance if inhaled, ingested, or absorbed through the skin, and it also reduces male fertility. Pyrimidine is a heterocyclic aromatic organic compound similar to benzene and pyridine; pyrimidines serve essential functions in human metabolism. Pyrrole itself is not naturally occurring, but many of its derivatives are found in a variety of cofactors and natural products; pyrrole and pyrrole-based compounds are largely as flavoring and pharmaceutical ingredients, as well as ingredients for pesticides and insecticides. Octane is a potentially toxic compound hydrocarbon. Ketones, such as acetone, 2-pentanone, and 3-hexanone, mainly come from diet and have been found to be associated with several diseases. Fluorenes are potentially toxic compounds containing a fluorene moiety, which consists of two benzene rings connected through either a cyclopentane, cyclopentene, or cyclopenta-1,3-diene. Auramine is not a naturally occurring metabolite and is only found in those individuals exposed to this compound or its derivatives; it is a known human carcinogen [53]. Propanal, 2-methyl- is an aldehyde with many toxicologies (chronic, subchronic, developmental, genotoxicity).

Looking at the urinary VOC predictors of sperm progressive motility, we can find some of the mentioned VOC predictors (3-hexanone, pyrrole-based, fluorenes) and other exogenous compounds, such as 1H-Indole,5-methyl-2-phenyl, 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-, 2,4,5-trioxoimidazolidine, butanal, and 3-methyl (Figure 3c). The latter is found in low concentrations in many types of food, and commercially, it is used as a reagent to produce pharmaceuticals, perfumes, and pesticides.

Regarding total sperm motility (Figure 4c), it is interesting to note among the urinary VOC predictors, apart from some predictive VOCs common to the previous sperm parameters, the presence of hexanal, which is an alkyl aldehyde found in human biofluids with genotoxic and cytotoxic properties, as well as of n-hexane, which is a gasoline tasting compound found in food; it is potentially toxic, causing degeneration of the peripheral nervous system, and exposure to it may also damage the lungs and reproductive system. 4-Heptanone, belonging to ketones, is considered to be an oxygenated hydrocarbon lipid

molecule that is practically insoluble in water. 2-Aminoanthracene, a genotoxic chemical belonging to anthracenes (organic compounds containing a system of three linearly fused benzene rings), was also registered; it is not a naturally occurring metabolite and is only found in individuals exposed to this compound or its derivatives. Dimethyl disulfide (DMDS) is widespread in nature, as it is emitted by bacteria, fungi, plants, and animals, and it is used as a food additive, industrial sulfiding agent, and effective soil fumigant in agriculture.

As expected, in the model for immotile sperm cell estimation based on urinary VOCs, we found many of the VOC predictors of the previous sperm motility parameters (Figure 5c).

These results suggest hidden mechanisms in which predictive VOCs are representative of sperm cell concentration and play a role in determining their motility characteristics. Since many of the VOCs found in biofluids belong to human exposome, their influence on semen quality is indicative of excessive exposure to multiple environmental stressors and could be significant in expanding the list of chemicals of emerging concern and biomarkers with health effects in human biomonitoring.

The role of VOCs, selected from the volatilome of the considered biofluids in the deterioration of sperm quality, also extends to morphological abnormalities and round cells. We also observe that curiously, the regressive model based on the blood dataset (B) is not suitable for estimating the total morphological anomalies of the sperm, those of the head and those of the neck, or round cells, while it fairly estimates the anomalies of the tail. For this reason, it was decided to indicate the results deriving from the model based on the B + U dataset for morphological anomalies (Figure S2) and tail anomalies (Figure S5) and the one based on the U dataset for head (Figure S3) and neck anomalies (Figure S4), as well as round cells (Figure S6). It should also be noted that the regression model based on the human semen dataset (HS) with the related panel of semen VOC predictors performs less well than the models based on the B + U and U datasets.

Looking at total morphological anomalies, we found compounds as VOC predictors belonging to pyrroles, pyridine and fluorenes, hexanal, 2-anthracenamine, and other chemicals, such as propane, 2-(ethenyloxy)-, oxime-,methoxy-phenyl-, 2-ethyl-oxetane, and 1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea (Figure S2c).

Many compounds found in the previous model are also VOC predictors of head morphological anomalies; as novel predictors, we mentioned the main role of 11H-Dibenzo[b,e] diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl], and secondary butanal (Figure S3c). Butanal is an aldehyde that exists in all living organisms, ranging from bacteria to humans, and outside of the human body, it is found in several different foods.

Regarding VOC predictors of neck morphological anomalies, we can observe, as not yet mentioned, the contribution of 2-Butanone (Figure S4c). It has been detected in several different foods, and it is also a secondary metabolite; with regard to humans, has been found to be associated with several diseases, such as Crohn's disease, pervasive developmental disorder not otherwise specified, asthma, and ulcerative colitis; butanone has also been linked to the inborn metabolic disorder celiac disease.

Regarding VOC predictors of tail morphological anomalies, we can observe, as not yet mentioned, the contribution of Benzaldehyde,2-nitro-,diaminomethylidenhydrazone and alpha-Pinene (Figure S5c). The latter is an organic compound of the terpene class and is found in the oils of many species of many coniferous trees; it is also found in cannabis plants and in finished, dried cannabis flower preparation, commonly known as marijuana.

Finally, looking at the VOC predictors of round cells, many of the compounds predictive of the other spermiogram parameters are present (Figure S6c).

It is also worth discussing the results of the regression models that consider the VOC variables identified in the human semen, in particular the model based on the HS + B + U data, to evaluate which sperm VOCs were selected by the model as predictive variables. Looking at the results related to the HS + B + U dataset and shown in Figures S7–S15, we can observe that some VOCs of HS with toxic properties played

a role in the prediction model, together with the VOCs of blood and urine. Among the VOC predictors for the HS + B + U model, we highlight the following VOCs of HS (label VOCname_HS): (a) 1-anthracenamine; propane, 2-(ethenyloxy); 3-aminopyrrolidine; and 2-ethyl-oxetane (for sperm cells concentrations, Figure S7); (b) acetic acid; sodium salt; hexanal; butanal; and 2-methyl-; 1-anthracenamine (for sperm progressive motility, Figure S8); (c) 1-anthracenamine; n-hexane; butanal; and 2-methyl- (for sperm total motility, Figure S9); (d) 1-anthracenamine; butanal and 2-methyl- (for sperm immotiles, Figure S10); (e) n-hexane (for sperm morphology anomalies, Figure S11); (f) auramine and hexanal (for sperm head anomalies, Figure S12); (g) acetic acid; sodium salt; butanal; and 2-methyl- (for sperm neck anomalies, Figure S13); (h) butanal; pentanal; acetic acid; sodium salt; and oxime,methoxy-phenyl- (for sperm tail anomalies, Figure S14); and (i) acetone; n-hexane; and pentanal (for round cell concentration, Figure S15).

To have an overview of all VOC variables that emerged as having a predictive role in panels based on the analyses carried out, Table S2 summarizes all the VOC predictors in the different regression models (based on datasets HS + B + U; HS + B; HS + U; B + U; HS; B; U). Since a compound was found in more than a biofluid, for each VOC predictor, the origin biofluid jointed to the related dataset of the regression models is reported, in which it was a predictive variable. Moreover, Table S2 shows the CAS number of the VOCs and the relevant chemical dashboards in the freely available electronic database used in this work, which offer rich information on the VOCs' properties and their harmful effects recorded so far.

5. Conclusions

This pilot biomonitoring study addresses the issue of health risk assessment in populations in contaminated areas. In this work, we contributed to demonstrating that the human volatilome contains fundamental and hidden information about the organism's reactions to environmental exposures, particularly toxic chemical exposures that negatively impact health.

The results of this work propose an original approach to exploit the correlations between the human volatilome and other measurable health effects, such as spermiogram parameters, which are indicative of sperm quality. The idea is to use the general regression model based on optimized panels of VOCs as independent variables in different biofluids (sperm, blood, urine) as an indirect measure of the spermiogram parameters. Such an innovative approach can be extended to other areas of the human exposome for the estimation of other clinical parameters as well as known or company biomarkers that describe the health status of a subject. In this way, the appropriate use of the data analysis technique contributes to validating the panels of VOCs, selected from the biofluids' volatilome, as new candidates for biomarkers of exposure and/or disease risk.

Furthermore, it should be underlined that it is certainly encouraging to observe the good performance of MLR models that do not use sperm data to estimate spermiogram parameters. The implications of this result are very interesting, as they have strong implications for preventive reproductive medicine together with a strong social impact. In fact, a simple urine or blood sample may be sufficient to obtain an indirect assessment of sperm quality, thus also overcoming the psychological stress and anxiety associated with sperm analysis in young men. VOC analysis of blood, urine, and (optionally) semen can hence be a new methodology for the analysis of semen quality. Sperm analysis is indeed usually prescribed to men only when seeking paternity within a couple's relationship. Postponing this test and an andrological visit to a more adult age effectively prevents the screening of boys who already show signs of sperm deterioration at a young age and on whom it is appropriate to carry out more complete medical investigations.

It is beyond the scope of this study to hypothesize possible biological pathways of toxicity involving the VOCs identified as predictive, for which we refer to the database platforms that collect all the data available for a deeper discussion. In this study, we limit reporting on the association between VOCs found in body fluids and the detrimental

impacts on sperm parameters. It is difficult to explain the precise molecular mechanisms of those significant VOCs in eliciting adverse health effects in populations, as they have not always been fully elucidated. Anyway, a survey of some key characteristics has been identified, involving (1) alterations in germ cell development, functions, or death; (2) alteration in somatic cell development, functions, or death; (3) alterations in production and levels of reproductive hormones; (4) alterations of hormone receptor levels/functions; (5) genotoxic DNA damage, chromosome fragmentation, altered sperm cell chromosome numbers; (6) induction of epigenetic alterations; (7) induction of oxidative stress; and (8) induction of inflammation [27–30,51,52].

However, it is necessary to point out that although the predictive capacity of the VOCs identified in biofluids regarding sperm concentration, motility, shape anomalies, and round cells leads to the hypothesis of more potent biochemical interactions within body fluids, which ultimately affect sperm parameters, it is not possible to draw causal conclusions now, but it is legitimate to consider the presence of predictive VOCs as an alarm signal for sperm quality.

This work also lays the foundations for exposome studies that include innovative volatilomics studies on human biofluids; the latter can offer a tool for investigating possible health risks and for measuring parameters of clinical interest that are subtly affected by chemical exposure factors. The point of strong debate in the scientific field relating to how to demonstrate the impact of chemical exposure resulting from external factors would, hence, make use of a new analysis method for human biomonitoring based on volatilomics, which would validate the approach both as an explorer of new biomarkers and as a measurement method of its effects on health through general regression models combined with other biomarkers or measurable parameters. The idea is that the human volatilome with all VOCs, both endogenous and exogenous, contains information on actual health, and this concept can, therefore, also be extended to volatilomics applied to early diagnostics.

Author Contributions: Conceptualization, S.C., A.F. (Angiola Forleo), A.V.R., V.L., A.G., A.F. (Alessandra Ferramosca), G.G., P.S., S.P., M.P. and L.M.; methodology, S.C., A.F. (Angiola Forleo), V.L., A.G., A.F. (Alessandra Ferramosca), T.N., S.P., M.P. and L.M.; software, A.V.R.; validation, S.C., A.F. (Angiola Forleo) and V.L.; formal analysis, S.C., A.F. (Angiola Forleo), V.L. and G.M.; investigation, A.F. (Angiola Forleo), V.L., F.C., T.N. and S.P.; writing—original draft preparation, S.C.; writing—review and editing, S.C., A.F. (Angiola Forleo), V.L., G.M., A.G., A.F. (Alessandra Ferramosca), M.P. and L.M.; supervision, S.C. and L.M.; funding acquisition, S.C., G.G., P.S. and L.M.; resources, S.P. and L.M. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Informed consent was obtained from the recruited subjects before sample collection.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical reasons.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

VSR	Valley Of Sacco River
LF	Land of Fires
HS	Headspace
SPME	Solid-Phase Microextraction
GC	Gas Chromatography
MS	Mass Spectrometry
VOCs	Volatile Organic Compounds
PAH	Polycyclic Aromatic Hydrocarbons
DNA	Deoxyribon Nucleic Acid
PVC	Polyvinyl Chloride
PTFE	Polytetrafluoroethylene
SCA	Sperm Class Analyzer
WHO	World Health Organization
CAS	Chemical Abstracts Service registry number
MLR	Multiple Linear Regression
IQR	Interquartile Range

References

1. Auger, J.; Eustache, F.; Chevrier, C.; Jégou, B. Spatiotemporal trends in human semen quality. *Nat. Rev. Urol.* **2022**, *19*, 597–626. [CrossRef] [PubMed]
2. Sengupta, P.; Borges, E., Jr.; Dutta, S.; Krajewska-Kulak, E. Decline in sperm count in European men during the past 50 years. *Hum. Exp. Toxicol.* **2018**, *37*, 247–255. [CrossRef] [PubMed]
3. Levine, H.; Jørgensen, N.; Martino-Andrade, A.; Mendiola, J.; Weksler-Derri, D.; Mindlis, I.; Pinotti, R.; Swan, S.H. Temporal trends in sperm count: A systematic review and meta-regression analysis. *Hum. Reprod. Update* **2017**, *23*, 646–659. [CrossRef] [PubMed]
4. Levine, H.; Jørgensen, N.; Martino-Andrade, A.; Mendiola, J.; Weksler-Derri, D.; Jolles, M.; Pinotti, R.; Swan, S.H. Temporal trends in sperm count: A systematic review and meta-regression analysis of samples collected globally in the 20th and 21st centuries. *Hum. Reprod. Update* **2023**, *29*, 157–176. [CrossRef] [PubMed]
5. Tong, N.; Witherspoon, L.; Dunne, C.; Flannigan, R. Global decline of male fertility: Fact or fiction? *BC Med. J.* **2022**, *64*, 126–130. Available online: https://bcmj.org/articles/global-decline-male-fertility-fact-or-fiction (accessed on 10 May 2024).
6. Vineis, P.; Robinson, O.; Chadeau-Hyam, M.; Dehghan, A.; Mudway, I.; Dagnino, S. What is new in the exposome? *Environ. Int.* **2020**, *143*, 105887. [CrossRef] [PubMed]
7. Miller, G.W. *The Exposome—A New Paradigm for the Environment and Health*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2020; ISBN 978-0-12-814079-6. [CrossRef]
8. Kumar, N.; Singh, A.K. Impact of environmental factors on human semen quality and male fertility: A narrative review. *Environ. Sci. Eur.* **2022**, *34*, 6. [CrossRef]
9. Nobles, C.J.; Schisterman, E.F.; Ha, S.; Kima, K.; Mumford, S.L.; Buck Louis, G.M.; Chen, Z.; Liu, D.; Sherman, S.; Mendola, P. Ambient air pollution and semen quality. *Environ. Res.* **2018**, *163*, 228–236. [CrossRef] [PubMed]
10. Jurewicz, J.; Dziewirska, E.; Radwan, M.; Hanke, W. Air pollution from natural and anthropic sources and male fertility. *Reprod. Biol. Endocrin.* **2018**, *16*, 109. [CrossRef]
11. Interpol—Environmental Crime. Available online: https://www.interpol.int/en/Crimes/Environmental-crime (accessed on 10 May 2024).
12. Mangotra, A.; Singh, S.K. Volatile organic compounds: A threat to the environment and health hazards to living organisms—A review. *J. Biotechnol.* **2024**, *382*, 51–69. [CrossRef]
13. Zhou, X.; Zhou, X.; Wang, C.; Zhou, H. Environmental and human health impacts of volatile organic compounds: A perspective review. *Chemosphere* **2023**, *313*, 137489. [CrossRef]
14. Li, A.J.; Pal, V.K.; Kannan, K. A review of environmental occurrence, toxicity, biotransformation and biomonitoring of volatile organic compounds. *Environ. Chem. Ecotoxicol.* **2021**, *3*, 91–116. [CrossRef]

15. de Lacy Costello, B.; Amann, A.; Al-Kateb, H.; Flynn, C.; Filipiak, W.; Khalid, T.; Osborne, D.; Ratcliffe, N.M.J. A review of the volatiles from the healthy human body. *J. Breath Res.* **2014**, *8*, 014001. [CrossRef] [PubMed]

16. Drabińska, N.; Flynn, C.; Ratcliffe, N.; Belluomo, I.; Myridakis, A.; Gould, O.; Fois, M.; Smart, A.; Devine, T.; De Lacy Costello, B. A literature survey of all volatiles from healthy human breath and bodily fluids: The human volatilome. *J. Breath Res.* **2021**, *15*, 034001. [CrossRef] [PubMed]

17. Ogbodo, J.O.; Arazu, A.V.; Iguh, T.C.; Onwodi, N.J.; Ezike, T.C. Volatile organic compounds: A proinflammatory activator in autoimmune diseases. *Front. Immunol.* **2022**, *13*, 928379. [CrossRef] [PubMed]

18. Calenic, B.; Miricescu, D.; Greabu, M.; Kuznetsov, A.V.; Troppmair, J.; Ruzsanyi, V.; Amann, A. Oxidative stress and volatile organic compounds: Interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* **2015**, *13*, 1020–1030. [CrossRef]

19. Dziewirska, E.; Radwan, M.; Wielgomas, B.; Klimowska, A.; Radwan, P.; Kałużny, P.; Hanke, W.; Słodki, M.; Jurewicz, J. Human Semen Quality, Sperm DNA Damage, and the Level of Urinary Concentrations of 1N and TCPY, the Biomarkers of Nonpersistent Insecticides. *Am. J. Men's Health* **2018**, *13*, 1557988318816598. [CrossRef]

20. Gabrielsen, J.S.; Tanrikut, C. Chronic exposures and male fertility: The impacts of environment, diet, and drug use on spermatogenesis. *Andrology* **2016**, *4*, 648–661. [CrossRef] [PubMed]

21. Longo, V.; Forleo, A.; Giampetruzzi, L.; Siciliano, P.; Capone, S. Human biomonitoring of environmental and occupational exposures by GC-MS and gas sensor systems: A systematic review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10236. [CrossRef] [PubMed]

22. Pironti, C.; Ricciardi, M.; Proto, A.; Bianco, P.M.; Montano, L.; Motta, O. Endocrine-Disrupting Compounds: An Overview on Their Occurrence in the Aquatic Environment and Human Exposure. *Water* **2021**, *13*, 1347. [CrossRef]

23. Caporossi, L.; Alteri, A.; Campo, G.; Paci, E.; Tranfo, G.; Capanna, S.; Papaleo, E.; Pigini, D.; Viganò, P.; Papaleo, B. Cross Sectional Study on Exposure to BPA and Phthalates and Semen Parameters in Men Attending a Fertility Center. *Int. J. Environ. Res. Public Health* **2020**, *17*, 489. [CrossRef]

24. Kleshchev, M.; Osadchuk, A.; Osadchuk, L. Impaired semen quality, an increase of sperm morphological defects and DNA fragmentation associated with environmental pollution in urban population of young men from Western Siberia, Russia. *PLoS ONE* **2021**, *16*, e0258900. [CrossRef]

25. Perrone, P.; Lettieri, G.; Marinaro, C.; Longo, V.; Capone, S.; Forleo, A.; Pappalardo, S.; Montano, L.; Piscopo, M. Molecular Alterations and Severe Abnormalities in Spermatozoa of Young Men Living in the "Valley of Sacco River" (Latium, Italy): A Preliminary Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11023. [CrossRef]

26. Selvaraju, V.; Baskaran, S.; Agarwal, A.; Henkel, R. Environmental contaminants and male infertility: Effects and mechanisms. *Andrologia* **2021**, *53*, e13646. [CrossRef] [PubMed]

27. Arzuaga, X.; Smith, M.T.; Gibbons, C.F.; Skakkebæk, N.E.; Yost, E.E.; Beverly, B.E.J.; Hotchkiss, A.K.; Hauser, R.; Pagani, R.L.; Schrader, S.M.; et al. Proposed Key Characteristics of Male Reproductive Toxicants as an Approach for Organizing and Evaluating Mechanistic Evidence in Human Health Hazard Assessments. *Environ. Health Perspect.* **2019**, *127*, 065001. [CrossRef] [PubMed]

28. Roychoudhury, S.; Kesari, K.K. A Comprehensive Update on Male Infertility Volume II. In *Oxidative Stress and Toxicity in Reproductive Biology and Medicine*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; ISBN 9783031129650.

29. Omolaoye, T.S.; Skosana, B.T.; Ferguson, L.M.; Ramsunder, Y.; Ayad, B.M.; Du Plessis, S.S. Implications of Exposure to Air Pollution on Male Reproduction: The Role of Oxidative Stress. *Antioxidants* **2024**, *13*, 64. [CrossRef] [PubMed]

30. Carré, J.; Gatimel, N.; Moreau, J.; Parinaud, J.; Léandri, L. Does air pollution play a role in infertility?: A systematic review. *Environ. Health* **2017**, *16*, 82. [CrossRef] [PubMed]

31. Montano, L.; Bergamo, P.; Andreassi, M.G.; Lorenzetti, S. *The Role of Human Semen as an Early and Reliable Tool of Environmental Impact Assessment on Human Health*; Intechopen: London, UK, 2018; pp. 173–202. [CrossRef]

32. Montano, L.; Ceretti, E.; Donato, F.; Bergamo, P.; Zani, C.; Viola, G.C.V.; Notari, T.; Pappalardo, S.; Zani, D.; Ubaldi, S.; et al. FASt study group. Effects of a Lifestyle Change Intervention on Semen Quality in Healthy Young Men Living in Highly Polluted Areas in Italy: The FASt Randomized Controlled Trial. *EU Focus* **2022**, *8*, 351–359. [CrossRef]

33. Ecofoodfertility. Available online: https://ecofoodfertility.it/ (accessed on 11 April 2024).

34. Longo, V.; Forleo, A.; Ferramosca, A.; Notari, T.; Pappalardo, S.; Siciliano, P.; Capone, S.; Montano, L. Blood, urine and semen Volatile Organic Compound (VOC) pattern analysis for assessing health environmental impact in highly polluted areas in Italy. *Environ. Pollut.* **2021**, *286*, 117410. [CrossRef]

35. Longo, V.; Forleo, A.; Radogna, A.V.; Siciliano, P.; Notari, T.; Pappalardo, S.; Piscopo, M.; Montano, L.; Capone, S. A novel human biomonitoring study by semiconductor gas sensors in Exposomics: Investigation of health risk in contaminated sites. *Environ. Pollut.* **2022**, *304*, 119119. [CrossRef]

36. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subject. *JAMA* **2013**, *310*, 2191–2194. [CrossRef]

37. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 6th ed.; World Health Organization: Geneva, Switzerland, 2021; ISBN 978-92-4-003078-7.

38. Boitrelle, F.; Shah, R.; Saleh, R.; Henkel, R.; Kandil, H.; Chung, E.; Vogiatzi, P.; Zini, A.; Arafa, M.; Agarwal, A. The Sixth Edition of the WHO Manual for Human Semen Analysis: A Critical Review and SWOT Analysis. *Life* **2021**, *11*, 1368. [CrossRef]

39. Boeri, L.; Fallara, G.; Pozzi, E.; Belladelli, F.; Corsini, C.; Raffo, M.; Schifano, N.; Capogrosso, P.; d'Arma, A.; Montorsi, F.; et al. The impact of different WHO reference criteria for semen analysis in clinical practice: Who will benefit from the new 2021 thresholds for normal semen parameters? *Andrology* **2022**, *10*, 1134–1142. [CrossRef]

40. Long, S.; Kenworthy, S. Round Cells in Diagnostic Semen Analysis: A Guide for Laboratories and Clinicians. *Br. J. Biomed. Sci.* **2022**, *79*, 10129. [CrossRef] [PubMed]

41. World Health Organization. *WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction*, 4th ed.; Cambridge University Press: Cambridge, UK, 1999; ISBN 0-521-64599-9.

42. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th ed.; World Health Organization: Geneva, Switzerland, 2010; ISBN 978-92-4-154778-9.

43. Wald, G.; Punjani, N.; Hayden, R.; Feliciano, M.; Dudley, V.; Goldstein, M. Assessing the clinical value of the Kruger strict morphology criteria over the World Health Organization fourth edition criteria. *F S Rep.* **2021**, *19*, 176–180. [CrossRef] [PubMed]

44. Nunzio, A.D.; Giarra, A.; Toscanesi, M.; Amoresano, A.; Piscopo, M.; Ceretti, E.; Zani, C.; Lorenzetti, S.; Trifuoggi, M.; Montano, L. Comparison between Macro and Trace Element Concentrations in Human Semen and Blood Serum in Highly Polluted Areas in Italy. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11635. [CrossRef] [PubMed]

45. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, 34986597. [CrossRef] [PubMed]

46. HMDB the Human Metabolome Database. Available online: https://hmdb.ca/ (accessed on 31 May 2024).

47. Williams, A.J.; Grulke, C.M.; Edwards, J.; McEachran, A.D.; Mansouri, K.; Baker, N.C.; Patlewicz, G.; Shah, I.; Wambaugh, J.F.; Judson, R.S.; et al. The CompTox Chemistry Dashboard: A community data resource for environmental chemistry. *J. Cheminform.* **2017**, *9*, 61. [CrossRef] [PubMed]

48. CompTox Chemicals Dashboard (EPA). Available online: https://comptox.epa.gov/dashboard/ (accessed on 31 May 2024).

49. CTDbase Comparative Toxicogenomics Database. Available online: https://ctdbase.org/ (accessed on 10 April 2024).

50. PubChem Compound. Available online: https://pubchem.ncbi.nlm.nih.gov/ (accessed on 10 April 2024).

51. Wei, C.; Cao, L.; Zhou, Y.; Zhang, W.; Zhang, P.; Wang, M.; Xiong, X.; Deng, C.; Xiong, Q.; Liu, W.; et al. Multiple statistical models reveal specific volatile organic compounds affect sex hormones in American adult male: NHANES 2013–2016. *Front. Endocrinol.* **2023**, *13*, 1076664. [CrossRef]

52. Webb, E.; Bushkin-Bedient, S.; Cheng, A.; Kassotis, C.D.; Balise, V.; Nagel, S.C. Developmental and reproductive effects of chemicals associated with unconventional oil and natural gas operations. *Rev. Environ. Health* **2014**, *29*, 307–318. [CrossRef]

53. OEHHA—California Office of Environmental Health Hazard Assessment. Available online: https://oehha.ca.gov/proposition-65/proposition-65-list/ (accessed on 10 April 2024).

Innovative Approach for Human Semen Quality Assessment Based on Volatilomics

Simonetta Capone [1,*], Angiola Forleo [1], Antonio Vincenzo Radogna [1,2], Valentina Longo [1], Giulia My [1], Alessandra Genga [3], Alessandra Ferramosca [2], Giuseppe Grassi [4], Flavio Casino [1], Pietro Siciliano [1], Tiziana Notari [5], Sebastiana Pappalardo [6], Marina Piscopo [7] and Luigi Montano [8,9,10]*

[1] National Research Council, Institute for Microelectronics and Microsystems (CNR-IMM), 73100 Lecce, Italy; angiola.forleo@cnr.it (A.F.); antonio.radogna@unisalento.it (A.V.R.); valentinalongo2110@gmail.com (V.L.); my.giulia@imm.cnr.it (G.M.); flavio.casino@cnr.it (F.C.); pietroaleardo.siciliano@cnr.it (P.S.)

[2] Department of Experimental Medicine, University of Salento, 73100 Lecce, Italy; alessandra.ferramosca@unisalento.it

[3] Department of Biological and Environmental Sciences and Technologies, University of Salento, 73100 Lecce, Italy; alessandra.genga@unisalento.it

[4] Department of Engineering for Innovation, University of Salento, 73100 Lecce, Italy; giuseppe.grassi@unisalento.it

[5] Reproductive Medicine Unit of Check Up Polydiagnostic Center, 84131 Salerno, Italy; tiziananotari7@gmail.com

[6] Reproduction and Fertility Center, 00128 Rome, Italy; pappalardo@riproduzionefertilita.it

[7] Department of Biology, University of Naples Federico II, 80138 Naples, Italy; marina.piscopo@unina.it

[8] Andrology Unit and Service of Lifestyle Medicine in UroAndrology, "S. Francesco d'Assisi" Hospital, Oliveto Citra, 84020 Salerno, Italy

[9] Coordination Unit of the Network for Environmental and Reproductive Health (EcoFoodFertility Project), "S. Francesco d'Assisi" Hospital, Oliveto Citra, 84020 Salerno, Italy

[10] Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy

[*] Correspondence: simonetta.capone@cnr.it (S.C.); l.montano@aslsalerno.it (L.M.); Tel.: +39-0832-422507 (S.C.)

Table S1 List of VOC predictors. For each VOC it's reported the concentration in ng/ml in the biofluid where it was found (HS: Human Semen; B: Blood; U: Urine), the mean, standard deviation (SD), median, upper and lower limit, and the 95th percentile in the sample population. Notation for numeric values: comma "," is the decimal separator (SI).

n.	VOC predictor	BIOFLUIDS (HS, B, U)	MEAN (ng/ml)	SD (ng/ml)	MEDIAN (ng/ml)	CI 95%		
						LOWER LIMIT (ng/ml)	UPPER LIMIT (ng/ml)	PERCENTILE 95th (ng/ml)
1	(3-Methoxy-phenyl)-(6-methyl-4-phenyl-quinazolin-2-yl)-amine	HS	4,84	19,38	0	0	84,84	77,78
		B	6,62	24,74	0	0	132,11	79,77
2	1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea	HS	7,54	24,69	0	0	100,28	90,59
		B	4,77	20,62	0	0	127,7	29,36
3	11H-Dibenzo[b,e][1,4] diazepin-11-one, 5,10-dihydro-5-[3-(methylamino) propyl]-	B	1,53	4,54	0	0	22,22	12,61
		U	0,60	1,79	0	0	7,73	6,84
4	1-Anthracenamine	HS	9,98	27,44	0	0	87,88	87,77
5	1H-Indole, 5-methyl-2-phenyl-	U	0,24	0,80	0	0	2,84	2,81
6	2,4,5-Trioxoimidazolidine	U	1,01	2,27	0	0	6,64	6,49
7	2-Anthracenamine	HS	16,54	39,64	0	0	177,16	118,86
		B	3,17	9,12	0	0	33,91	33,26
		U	1,96	4,65	0	0	17,35	15,96
8	2-Butanone	U	6,97	12,64	0	0	38,22	37,19
9	2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine	B	3,69	11,42	0	0	58,57	32,85
10	2-Ethyl-oxetane	HS	31,87	102,01	0	0	523,56	324,84
		B	241,25	1368,04	0	0	8767,93	463,77
		U	1,48	2,83	0	0	9,89	9,39

11	2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-	B	2,60	9,47	0	0	46,83	35,16
		U	0,58	2,00	0	0	8,57	7,79
12	2-Pentanone	B	6,48	22,18	0	0	135,45	29,99
		U	33,98	51,12	12,4	0	165,26	163,75
13	3-Aminopyrrolidine	HS	6,23	20,24	0	0	80,76	72,11
		B	0,63	2,82	0	0	14,02	10,54
		U	1,84	4,73	0	0	19,53	17,67
14	3-Hexanone	U	1,31	3,06	0	0	10,07	9,83
15	4-(4-Chlorophenyl)-2,6-diphenylpyridine	HS	10,44	35,22	0	0	142,52	138,54
		B	6,63	21,46	0	0	111,15	66,19
		U	4,30	8,65	0	0	25,17	24,69
16	4-Heptanone	U	70,11	163,62	23,01	0	773,91	663,82
17	5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-	U	1,72	5,91	0	0	25,06	22,93
18	Acetic acid, sodium salt	HS	7,97	27,69	0	0	135,8	90,47
		B	6,74	29,35	0	0	182,61	42,50
		U	6,24	26,35	0	0	126,71	103,19
19	Acetone	HS	17,39	218,93	128,92	0	1109,2	695,16
		B	3432,27	2898,01	3803,49	0	8767,93	7802,82
		U	115,87	190,99	52,99	0	897,39	769,88
20	alpha-Pinene	B	2,64	7,30	0	0	34,12	24,03
21	Auramine	HS	11,94	25,78	0	0	73,08	69,13
		B	0,98	3,63	0	0	17,52	12,56
22	Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone	B	1,65	4,64	0	0	19,26	15,20
		U	0,59	1,70	0	0	7,14	6,38
23	Butanal	HS	8,28	27,45	0	0	117,2	100,32
		U	0,84	2,05	0	0	7,97	7,28

24	Butanal, 2-methyl-	HS	26,32	41,50	0	0	111,12	106,92
25	Butanal, 3-methyl-	HS	73,72	54,67	74,12	0	173,76	173,34
		B	2,57	7,58	0	0	34,89	22,29
		U	1,67	3,81	0	0	15,1	13,91
26	Butane, 2-methyl-	HS	21,20	60,18	0	0	246,12	230,33
		B	0,91	4,21	0	0	23,43	12,58
27	Cyclohexane	B	23,37	43,60	0	0	223,37	124,36
28	Cyclopentane, methyl-	B	109,80	122,30	77,41	0	418,62	398,52
29	Disulfide, dimethyl	U	31,74	58,48	11,74	0	268,57	324,86
30	D-Limonene	HS	80,89	296,67	0	0	1353,56	1163,05
		B	1,13	5,13	0	0	26,89	17,59
31	Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-	HS	34,74	49,75	0	0	208,44	138,78
		U	1,30	3,67	0	0	14,76	13,57
32	Heptanal	B	0,60	2,67	0	0	12,87	10,46
33	Heptane	B	10,19	27,39	0	0	150,9	71,29
34	Hexanal	HS	29,88	52,34	0	0	220,64	154,28
		U	0,62	1,76	0	0	7,1	6,55
35	N-Benzyl-N-ethyl-p-isopropylbenzamide	B	24,78	95,50	0	0	504,85	326,23
36	n-Hexane	HS	158,27	371,24	0	0	1467,4	1304,08
		B	115,33	190,06	28,01	0	850,54	729,95
		U	4,01	5,86	2,81	0	25,38	22,68
37	Octane	B	13,79	39,44	0	0	221,4	107,77
38	Oxime-, methoxy-phenyl-	HS	14,31	31,61	0	0	110	91,49

		B	6,13	15,54	0	0	91,55	29,06
		U	11,96	46,75	0	0	225,33	184,06
39	Pentanal	HS	17,17	36,06	0	0	139,16	116,12
		B	6,29	16,05	0	0	90,5	34,41
40	Pentane	B	113,08	244,68	0	0	1086,09	853,25
41	Propanal, 2-methyl-	U	1,99	6,46	0	0	29,62	25,44
42	Propane, 2-(ethenyloxy)-	HS	6,06	19,62	0	0	78,12	66,81
		B	2,72	7,70	0	0	35,09	26,51
		U	2,65	5,21	0	0	16,05	15,67
43	Pyrrole	HS	9,71	36,95	0	0	198,28	102,86
		U	2,44	6,28	0	0	24,29	21,87

Table S2 Characteristics of VOC predictors. For each VOC it's reported its CAS number, CID in PubChem [50], code in CompTox Dashboard (EPA) [48], ID in Human Metabolome Database HMDB [46] and its presence in Comparative Toxicogenomics Database [49]. The column Biofluids (HS, B, U) reports in which biofluids the VOC was found. The Spermiogram Parameter Prediction reports for all the spermiogram parameters the label: Biofluid/Datasets that refers to the origin biofluid/dataset of the regression model where the specific VOC contributed as a predictive variable.

								Spermiogram Parameter Prediction — Label: Biofluid/Datasets								
n.	VOC predictor	CAS	PubChem CID [a]	CompTox Dashboard (EPA) [b]	Human Metabolome Database HMDB ID [c]	Comparative Toxicogenomics Database [d]	Biofluids (HS, B, U)	Cells conc.	Progressive motility	Total motility	Immotiles	Morph. Anomalies	Head Anomalies	Neck Anomalies	Tail Anomalies	Round cells
1	(3-Methoxy-phenyl)-(6-methyl-4-phenyl-quinazolin-2-yl)-amine	1000317-62-7	/	/	/	/	HS; B	B/B and HS/HS+B and B/HS+B	B/B and HS/HS+B and B/HS+B	HS/HS+B and B/HS+B	HS/HS+B and B/HS+B		B/B and HS/HS and HS/HS+B	HS/HS	B/B and B/HS+B	B/B
2	1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea	131120-14-4	/	/	/	/	HS; B	B/B and HS/HS+B and B/HS+B				B/B+U	B/B and B/HS+B	B/B and HS/HS+B	B/B and HS/HS+B and B/HS+B	B/B+U and HS/HS and B/HS+B+U
3	11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]	13450-70-9	45221	DTXSID20158754	/	X	B; U	B/B+U and B/HS+B and B/HS+B+U	B/HS+B and B/HS+B+U	U/U and B/HS+B and B/HS+B+U	B/HS+B and B/HS+B+U	B/B and U/U	U/U and U/HS+U	B/B	B/B and B/HS+B and U/B+U and U/HS+B+U	U/U and B/HS+B
4	1-Anthracenamine	610-49-1	11885	DTXSID00209859	HMDB0243820	X	HS	HS/HS+B and HS/HS+U and HS/HS+B+U	HS/HS+B and HS/HS+U and HS/HS+B+U	HS/HS+B and HS/HS+U and HS/HS+B+U	HS/HS+B and HS/HS+B+U		HS/HS and HS/HS+B		HS/HS+B	HS/HS+B
5	1H-Indole, 5-methyl-2-phenyl-	13228-36-9	83247	DTXSID9074692	HMDB0000466	/	U	U/U and U/B+U and U/HS+U and U/HS+B+U	U/U and U/B+U and U/HS+U and U/HS+B+U	U/U and U/B+U and U/HS+B+U	U/U and U/HS+B+U	U/U				U/U and U/HS+U
6	2,4,5-Trioxoimidazolidine	120-89-8	67126	DTXSID2059516	HMDB0062802	X	U		U/U	U/U and HS/HS+U	U/U	U/U	U/U	U/U and U/HS+B+U	U/B+U	U/U
7	2-Anthracenamine	613-13-8	11937	DTXSID2024458	HMDB0245007	X	HS; B; U			U/U and B/B and U/HS+U	U/U and U/B+U	U/U and U/B+U and B/HS+B	U/U and B/B and B/HS+B	U/U and B/B and B/HS+B	HS/HS+B and B/HS+B	U/U and B/HS+B
8	2-Butanone	78-93-3	6569	DTXSID3021516	HMDB0000474	X	U	U/U and U/HS+B+U			U/U		U/U	U/U		
9	2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine	63673-76-7	631100	/	/	/	B	B/B+U and B/B	B/B and B/HS+B and B/HS+B+U	B/B and B/B+U and B/HS+B and B/HS+B+U	B/B and B/HS+B and B/HS+B+U					B/B+U
10	2-Ethyl-oxetane	1000386-40-2	521218	/	/	/	HS; B; U	U/B+U and B/B and HS/HS+B+U	B/B	B/B	U/U and B/B and B/B+U	U/U and B/B and B/B	U/U and B/B+U and HS/HS	B/B and B/HS+B+U	B/B+U and B/B and HS/HS+B and B/HS+B and B/HS+B+U	B/B and U/HS+U and U/B+U
11	2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-	53774-23-5	631062	/	/	/	B; U	U/B+U	B/B	B/B	B/B	U/U and B/B+U and B/B	U/U		B/HS+B	U/U and U/B+U and U/HS+B+U

No.	Compound	CAS	PubChem	DTXSID	HMDB												
12	2-Pentanone	107-87-9	7895	DTXSID0021888	HMDB0034235	X		B; U	U/U and U/B+U and B/B and U/HS+U and U/HS+B+U	U/B+U and B/HS+B	B/B and B/HS+B	U/U and B/B and U/B+U and B/HS+B	B/HS+B and U/HS+U and U/HS+B+U	B/B+U and B/HS+B	B/B and U/B+U	B/B	U/U and B/B and U/HS+U and U/B+U
13	3-Aminopyrrolidine	79286-79-6	164401	DTXSID70276538		/	/	HS; B; U	U/U and B/B+U and HS/HS and HS/HS+B and B/HS+B and HS/HS+U and U/HS+U and U/HS+B+U	U/U and B/B and U/B+U and U/HS+U	U/U and B/B and HS/HS+B and B/HS+B and U/HS+U	U/U and B/B and HS/HS+B and B/HS+B	U/U and B/B and HS/HS+B	U/U	U/U and B/B and HS/HS+B	B/B	B/B and HS/HS and B/HS+B and B/HS+B+U
14	3-Hexanone	589-38-8	11509	DTXSID2021608	HMDB0000753	X		U	U/B+U	U/U		U/U	U/HS+B+U and U/HS+B+U	U/U and U/B+U and U/HS+B+U	U/U		U/U
15	4-(4-Chlorophenyl)-2,6-diphenylpyridine	1498-82-4	631072	DTXSID20348153		/	/	HS; B;U	B/B+U and B/B			U/U and B/B and B/B+U and HS/HS+U	B/B+U	B/B and HS/HS and U/B+U and HS/HS+B and B/HS+B and	B/B+U and HS/HS and HS/HS+B and B/HS+B	B/B and HS/HS and HS/HS+B	B/B and HS/HS+B
16	4-Heptanone	123-19-3	31246	DTXSID6047650	HMDB0004814	X		U				U/U	U/U and U/B+U		U/U and U/HS+U and U/HS+B+U	U/HS+U	
17	5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E)-	13125-74-1	114621	DTXSID5065358		/	/	U	U/U	U/U	U/U	U/U		U/B+U		U/B+U	U/U
18	Acetic acid, sodium salt	127-09-3	517045	DTXSID2027044		/	/	HS; B; U	B/B	B/B and B/HS+B and HS/HS+B+U	B/B and B/B+U and B/HS+B and B/HS+B+U	U/U and B/B and B/B+U and B/HS+B and B/HS+B+U	B/HS+B	HS/HS and U/B+U	U/U and U/B+U and B/HS+B and HS/HS+U and HS+B+U	HS/HS and HS/HS+B and HS/HS+U and HS/HS+B+U	
19	Acetone	67-64-1	180	DTXSID8021482	HMDB01659	X		HS; B; U	U/B+U and B/B and HS/HS and U/HS+U and U/HS+B+U		B/B	B/B	HS/HS+B	U/U and B/HS+B	U/U and HS/HS and HS/HS+B	B/B and HS/HS and B/HS+B	U/U and HS/HS and HS/HS+B+U
20	alpha-Pinene	80-56-8	6654	DTXSID4026501	HMDB0006525	X		B			B/B		B/B+U		B/B and B/HS+B+U	B/B+U and B/HS+B+U	
21	Auramine	492-80-8	10298	DTXSID7043821	HMDB0248724	X		HS; B	B/B+U and B/B and B/HS+B	B/B and B/HS+B	B/B	B/B		HS/HS and HS/HS+B and B/HS+B and HS/HS+U and HS/HS+B+U		B/B and HS/HS and HS/HS+B	HS/HS and HS/HS+B and B/HS+B
22	Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone	102632-31-5	1E+07	DTXSID20879432		/	/	B; U			B/B+U	B/B+U		B/HS+B		U/B+U	U/U and B/B

23	Butanal	123-72-8	261	DTXSID8021513	HMDB0003543	X	HS; U	U/U and U/HS+U	U/HS+U and U/HS+B+U	U/B+U and U/HS+U and U/HS+B+U	U/HS/HS+U	U/U	U/U and HS/HS and HS/HS+B	U/B+U and U/HS+B+U	HS/HS and HS/HS+B and HS/HS+B+U	
24	Butanal, 2-methyl	96-17-3	7284	DTXSID2021818	HMDB0031526	/	HS	HS/HS+U	HS/HS+U and HS/HS+B+U	HS/HS+B+U	HS/HS+B+U	HS/HS	HS/HS and HS/HS+U	HS/HS+U and HS/HS+B+U	HS/HS	HS/HS and HS/HS+B
25	Butanal, 3-methyl-	590-86-3	11552	DTXSID1021619	HMDB0006478	X	HS; B; U	B/B and B/HS+B	U/U and B/B and B/HS+B and HS/HS+U	B/B	U/U and B/B and U/B+U	HS/HS	U/U and HS/HS+U	U/U and B/B and B/HS+B and HS/HS+B	B/B	U/U and B/B and HS/HS
26	Butane, 2-methyl-	78-78-4	6556	DTXSID8025468	HMDB0253668	X	HS; B	HS/HS+U	B/B and HS/HS+B and B/HS+B	B/HS+B and HS/HS+U	B/HS+B	HS/HS and HS/HS+B	HS/HS and HS/HS+B	HS/HS+B	HS/HS and HS/HS+B	HS/HS
27	Cyclohexane	110-82-7	8078	DTXSID4021923	HMDB0029597	/	B	B/B		B/B	B/B	B/B+U and B/B	B/B and B/HS+B and B/HS+B+U	B/B and B/HS+B		B/B
28	Cyclopentane, methyl-	96-37-7	7296	DTXSID3025590	HMDB0031542	X	B	B/B	B/HS+B	B/HS+B	B/B and B/HS+B	B/B	B/B and B/HS+B	B/B	B/HS+B	
29	Disulfide, dimethyl	624-92-0	12232	DTXSID4025117	HMDB0005879	X	U			U/U and U/HS+U	U/U			U/B+U and U/HS+B+U	U/B+U	U/U and U/HS+U
30	D-Limonene	5989-27-5	440917	DTXSID1020778	HMDB0003375	X	HS; B	B/B and HS/HS+B	B/B	B/B	B/B	B/HS+B and HS/HS+U	HS/HS and HS/HS+U		HS/HS	HS/HS
31	Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-	1000217-31-2	631096	/	/	/	HS; U	U/U and U/B+U and U/HS+U	U/U and U/HS+B+U	U/U and U/B+U and U/HS+B+U	U/U and U/HS+B+U	U/B+U	U/U and HS/HS	U/U and HS/HS+B	HS/HS	U/U
32	Heptanal	111-71-7	8130	DTXSID0021597	HMDB0031475	X	B	B/B and B/HS+B		B/B	B/B	B/B	B/B and B/B+U and B/HS+B	B/B	B/B+U and B/B and B/HS+B	B/B
33	Heptane	142-82-5	8900	DTXSID6024127	HMDB0031447	X	B	B/B and B/HS+B		B/B and B/HS+B	B/B and B/HS+B	B/HS+B	B/HS+B			B/B+U
34	Hexanal	66-25-1	6184	DTXSID2021604	HMDB0005994	X	HS; U	HS/HS+B	HS/HS+B+U	U/U and U/HS+U	U/U	U/B+U	U/U and U/B+U and HS/HS and HS/HS+U and HS/HS+B+U	U/U and HS/HS		U/U and HS/HS and HS/HS+U
35	N-Benzyl-N-ethyl-p-isopropylbenzamide	015089-22-2	590384	/	/	/	B	B/B				B/B		B/B	B/B and B/HS+B	
36	n-Hexane	110-54-3	8058	DTXSID0021917	HMDB0029600	X	HS; B; U	HS/HS+B and B/HS+B	HS/HS+B and B/HS+B	U/U and B/B and HS/HS+B and B/HS+B and HS/HS+B+U		HS/HS and HS/HS+B and HS/HS+B+U	U/U and B/B and HS/HS+B and B/HS+B	U/U	B/B and HS/HS and HS/HS+B and B/HS+B and U/B+U and U/HS+B+U	B/B and B/B+U and HS/HS+B and HS/HS+U and HS/HS+B+U

#	Name	CAS	[a]	[b]	[c]	[d]										
37	Octane	111-65-9	356	DTXSID0026882	HMDB0001485	X	B	B/B+U and B/B and B/HS+B	B/B and B/HS+B							B+B+U and B/HS+B+U
38	Oxime-, methoxy-phenyl-	1000222-86-6	1E+07	/	/	X	HS; U		U/U		U/B+U	U/U		HS/HS+B	HS/HS and HS/HS+B+U	U/U and HS/HS+U
39	Pentanal	110-62-3	8063	DTXSID7021653	HMDB0031206	X	HS; B	B/B	B/B	B/B	B/B	HS/HS and B/HS+B			HS/HS and HS/HS+B+U	B/B and HS/HS and HS/HS+B and B/HS+B and HS/HS+U and
40	Pentane	109-66-0	8003	DTXSID2025846	HMDB0029603	X	B	B/HS+B	B/B	B/B	B/B		B/HS+B and B/HS+B+U			
41	Propanal, 2-methyl-	78-84-2	6561	DTXSID9021635	HMDB0031243	X	U	U/B+U		U/U and U/B+U	U/B+U			U/U		U/U
42	Propane, 2-(ethenyloxy)-	926-65-8	13557	DTXSID0061292	/	/	HS; B; U	U/U and B/B and HS/HS+B+U			U/B+U and U/HS+U and B/HS+B+U and U/HS+B+U		U/U and HS/HS+B and B/HS+B	B/B+U and HS/HS and HS/HS+B	B/B+U and B/HS+B and HS/HS+B	U/U and B/B and U/B+U and HS/HS and U/HS+U and U/HS+B+U
43	Pyrrole	109-97-7	8027	DTXSID5021910	HMDB35924	/	HS; U		U/U and U/B+U		U/B+U	U/U and HS/HS and HS/HS+B and HS/HS+U	U/U		U/B+U	U/U and HS/HS

[a] PubChem Compound, website https://pubchem.ncbi.nlm.nih.gov/ ([50])

[b] CompTox Chemicals Dashboard (EPA), website https://comptox.epa.gov/dashboard/ ([48])

[c] HMDB The Human Metabolome Database, website https://hmdb.ca/ ([46])

[d] CTDbase Comparative Toxicogenomics Database, website https://ctdbase.org/ ([49])
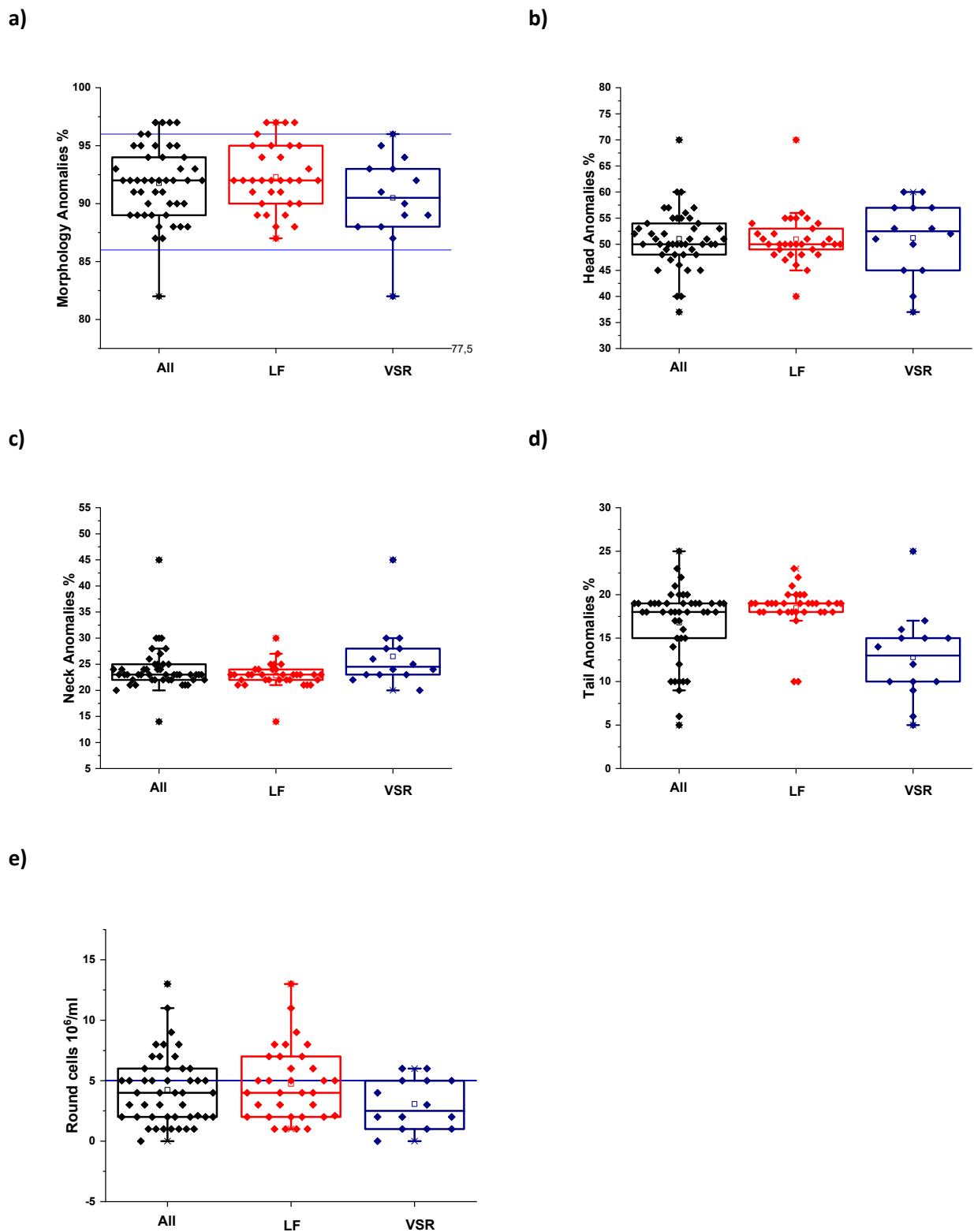
Figure S1 Box & whiskers plot displaying the dispersion of the seminogram parameters related to sperm morphology and round cells: a) total morphology anomalies; b) head anomalies; c) neck anomalies; d) tail anomalies; e) round cells concentration, in the subgroups of population living in *Land of Fire*s (LF) and in *Valley of Sacco river* (VSR) and in the total population (all=LF+VSR).
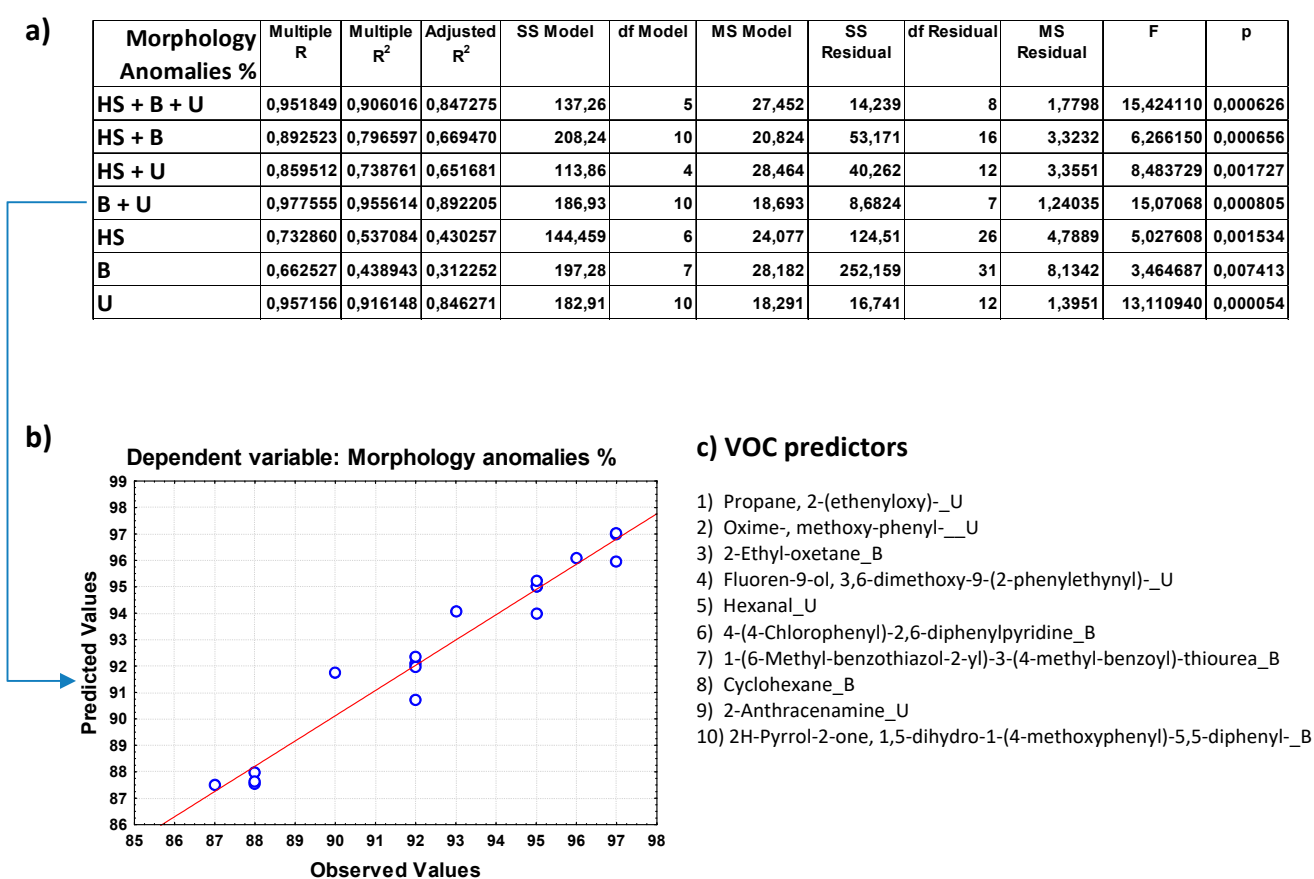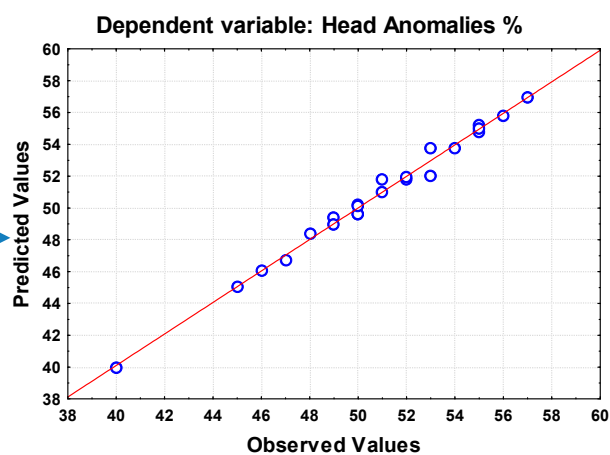
Morphology Anomalies %	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,951849	0,906016	0,847275	137,26	5	27,452	14,239	8	1,7798	15,424110	0,000626
HS + B	0,892523	0,796597	0,669470	208,24	10	20,824	53,171	16	3,3232	6,266150	0,000656
HS + U	0,859512	0,738761	0,651681	113,86	4	28,464	40,262	12	3,3551	8,483729	0,001727
B + U	0,977555	0,955614	0,892205	186,93	10	18,693	8,6824	7	1,24035	15,07068	0,000805
HS	0,732860	0,537084	0,430257	144,459	6	24,077	124,51	26	4,7889	5,027608	0,001534
B	0,662527	0,438943	0,312252	197,28	7	28,182	252,159	31	8,1342	3,464687	0,007413
U	0,957156	0,916148	0,846271	182,91	10	18,291	16,741	12	1,3951	13,110940	0,000054

b)

Dependent variable: Morphology anomalies %

c) VOC predictors

1) Propane, 2-(ethenyloxy)-_U
2) Oxime-, methoxy-phenyl-__U
3) 2-Ethyl-oxetane_B
4) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
5) Hexanal_U
6) 4-(4-Chlorophenyl)-2,6-diphenylpyridine_B
7) 1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea_B
8) Cyclohexane_B
9) 2-Anthracenamine_U
10) 2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-_B

Figure S2 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **morphology anomalies** (in percentage); b) Observed vs. Predicted values for sperm morphology anomalies as result of the MLR analysis for the data group U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

Head Anomalies%	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,986646	0,973470	0,950730	230,92	6	38,487	6,293	7	0,8990	42,808650	0,000036
HS + B	0,999603	0,999206	0,994842	743,71	22	33,805	0,591	4	0,1477	228,939800	0,000041
HS + U	0,954244	0,910582	0,841034	218,43	7	31,205	21,450	9	2,3833	13,092920	0,000458
B + U	0,950425	0,903307	0,794527	298,94	9	33,216	32,0000	8	4,00000	8,30401	0,003320
HS	0,869429	0,755907	0,609451	622,501	12	51,875	201,01	20	10,0507	5,161331	0,000648
B	0,730072	0,533006	0,388076	726,80	9	80,756	636,789	29	21,9582	3,677693	0,003573
U	0,995314	0,990650	0,965718	342,68	16	21,417	3,234	6	0,5390	39,733730	0,000093

b)

Dependent variable: Head Anomalies %

c) VOC predictors
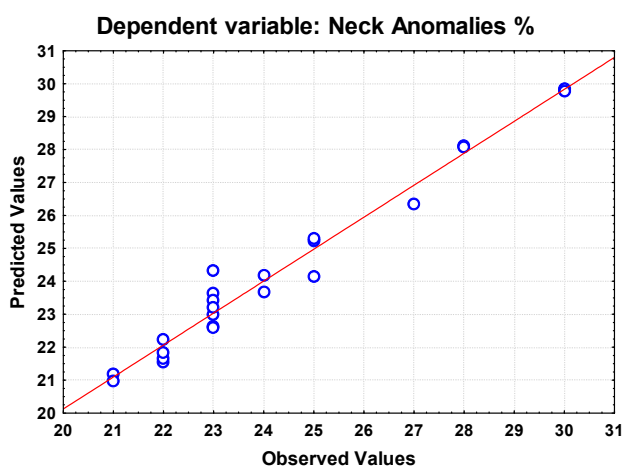
1) Oxime-, methoxy-phenyl-__U
2) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro -5-[3-(methylamino)propyl]-_U
3) 3-Hexanone_U
4) 2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-_U
5) Propane, 2-(ethenyloxy)-_U
6) Acetone_U
7) Hexanal_U
8) 2-Ethyl-oxetane_U
9) n-Hexane_U
10) 3-Aminopyrrolidine_U
11) Butanal_U
12) 2-Anthracenamine_U
13) Butanal, 3-methyl-_U
14) Pyrrole_U
15) 2,4,5-Trioxoimidazolidine_U
16) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U

Figure S3 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **head anomalies** (in percentage); b) Observed vs. Predicted values for sperm head anomalies as result of the MLR analysis for the data group U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

Neck Anomalies %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,997846	0,995696	0,988809	130,65	8	16,331	0,565	5	0,1130	144,578500	0,000017
HS + B	0,907557	0,823660	0,617931	208,36	14	14,883	44,607	12	3,7173	4,003611	0,010501
HS + U	0,902950	0,815320	0,769149	110,68	3	36,893	25,070	12	2,0892	17,659040	0,000107
B + U	0,813958	0,662528	0,478452	98,35	6	16,391	50,0959	11	4,55417	3,59921	0,031759
HS	0,761019	0,579150	0,461312	147,525	7	21,075	107,20	25	4,2881	4,914799	0,001346
B	0,721634	0,520755	0,299565	414,47	12	34,539	381,430	26	14,6704	2,354334	0,032890
U	0,985085	0,970392	0,918578	151,30	14	10,807	4,616	8	0,5770	18,728430	0,000140

b)

Dependent variable: Neck Anomalies %

c) VOC predictors
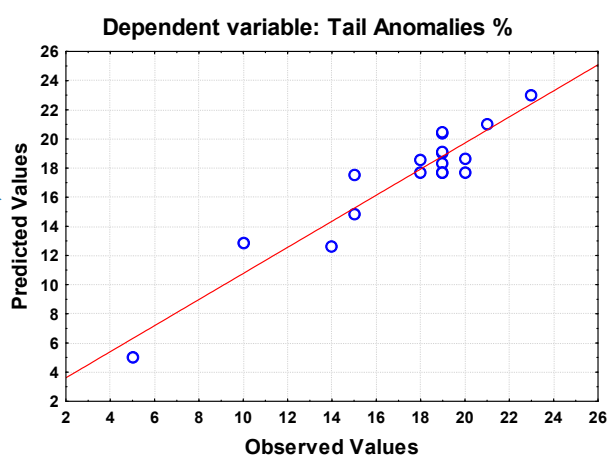
1) Butanal_U
2) 3-Aminopyrrolidine_U
3) Acetic acid, sodium salt_U
4) 2,4,5-Trioxoimidazolidine_U
5) n-Hexane_U
6) 2-Butanone_U
7) 4-Heptanone_U
8) Acetone_U
9) Hexanal_U
10) 2-Anthracenamine_U
11) Propanal, 2-methyl-_U
12) 3-Hexanone_U
13) Butanal, 3-methyl-_U
14) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U

Figure S4 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **neck anomalies** (in percentage); b) Observed vs. Predicted values for sperm neck anomalies as result of the MLR analysis for the data group U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

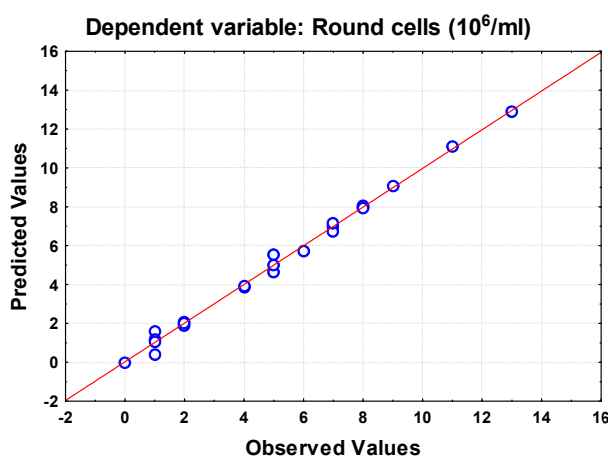| Tail Anomalies %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,999462	0,998925	0,993012	4331,05	11	393,732	4,661	2	2,3307	168,935000	0,005899
HS + B	0,999772	0,999545	0,996056	496,44	23	21,584	0,226	3	0,0753	286,495900	0,000293
HS + U	0,608807	0,370646	0,280738	106,70	2	53,351	181,180	14	12,9414	4,122518	0,039108
B + U	0,946014	0,894943	0,821404	275,64	7	39,378	32,3574	10	3,23574	12,16954	0,000358
HS	0,862206	0,743399	0,567830	411,077	13	31,621	141,89	19	7,4680	4,234225	0,002315
B	0,910540	0,829082	0,740205	682,82	13	52,525	140,766	25	5,6306	9,328398	0,000001
U	0,629402	0,396146	0,261957	126,46	4	31,614	192,761	18	10,7089	2,952139	0,048737

b)

c) VOC predictors

1) Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone_U
2) Disulfide, dimethyl_U
3) 2-Ethyl-oxetane_B
4) Heptanal_B
5) Propane, 2-(ethenyloxy)-_B
6) .alpha.-Pinene_B
7) 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-_U

Figure S5 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **sperm tail anomalies** (in percentage); b) Observed vs. Predicted values for sperm tail anomalies as result of the MLR analysis for the data group U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

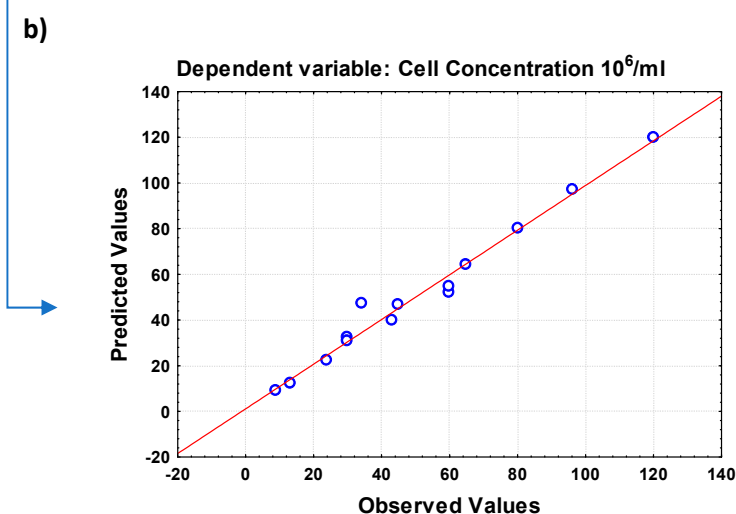| Round cells 10^6/ml	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,975114	0,950847	0,872202	142,15	8	17,769	7,348	5	1,4697	12,090350	0,006950
HS + B	0,918359	0,843384	0,728532	162,86	11	14,805	30,242	15	2,0162	7,343238	0,000306
HS + U	0,988538	0,977206	0,947901	192,57	9	21,396	4,492	7	0,6417	33,345000	0,000063
B + U	0,960418	0,922403	0,835106	192,12	9	21,346	16,1618	8	2,02022	10,56628	0,001458
HS	0,895012	0,801047	0,681675	202,624	12	16,885	50,33	20	2,5163	6,710513	0,000108
B	0,796776	0,634852	0,466323	174,97	12	14,581	100,638	26	3,8707	3,767008	0,002249
U	0,997482	0,994971	0,972338	272,45	18	15,136	1,377	4	0,3443	43,961700	0,001108

b)

c) VOC predictors

1) 1H-Indole, 5-methyl-2-phenyl-_U
2) Propane, 2-(ethenyloxy)-_U
3) 2-Anthracenamine_U
4) Hexanal_U
5) Disulfide, dimethyl_U
6) 2-Pentanone_U
7) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro -5-[3-(methylamino)propyl]-_U
8) 3-Hexanone_U
9) Pyrrole_U
10) Benzaldehyde, 2-nitro-, diaminomethylidenhydrazone_U
11) 5,9-Dodecadien-2-one, 6,10-dimethyl-, (E,E))-_U
12) Propanal, 2-methyl-_U
13) Acetone_U
14) Oxime-, methoxy-phenyl-__U
15) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
16) 2,4,5-Trioxoimidazolidine_U
17) 2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-_U

Figure S6 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **round cells** (in percentage); b) Observed vs. Predicted values for round cells as result of the MLR analysis for the data group U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

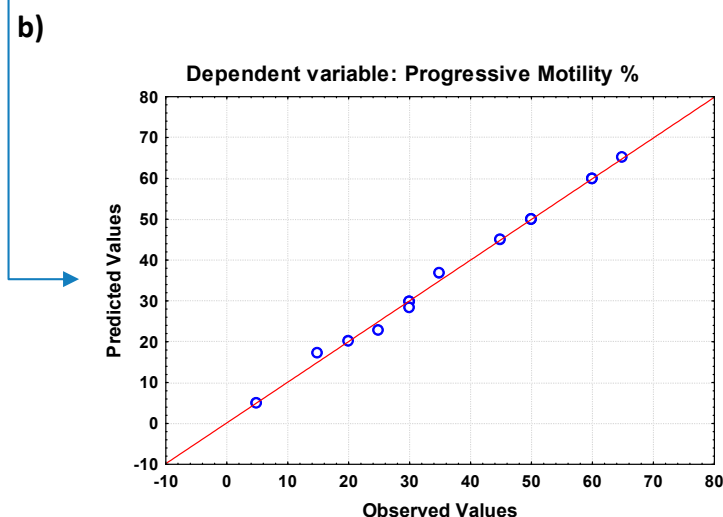| Concentration (10^6 cells/ml)	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,988824	0,977774	0,927764	12898,02	9	1433,113	293,194	4	73,2984	19,551760	0,005803
HS + B	0,970233	0,941352	0,847516	24610,99	16	1538,187	1533,303	10	153,3303	10,031850	0,000388
HS + U	0,969477	0,939886	0,862596	15282,99	9	1698,110	977,484	7	139,6405	12,160580	0,001679
B + U	0,989074	0,978268	0,907639	15121,03	13	1163,156	335,9117	4	83,97793	13,85074	0,010628
HS	0,495168	0,245191	0,194870	7428,854	2	3714,427	22869,39	30	762,3130	4,872575	0,014708
B	0,928148	0,861459	0,736772	29565,19	18	1642,510	4754,710	20	237,7355	6,908982	0,000039
U	0,880287	0,774905	0,669861	15263,00	7	2180,429	4433,608	15	295,5738	7,376934	0,000622

b)

c) VOC predictors

1) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_B
2) 1-Anthracenamine_HS
3) 2-Pentanone_U
4) Propane, 2-(ethenyloxy)_HS
5) 2-Butanone_U
6) 3-Aminopyrrolidine_HS
7) 2-Ethyl-oxetane_HS
8) Acetone_U

Figure S7 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **sperm concentration** (in 10^6/ml); b) Observed vs. Predicted values for sperm concentration as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

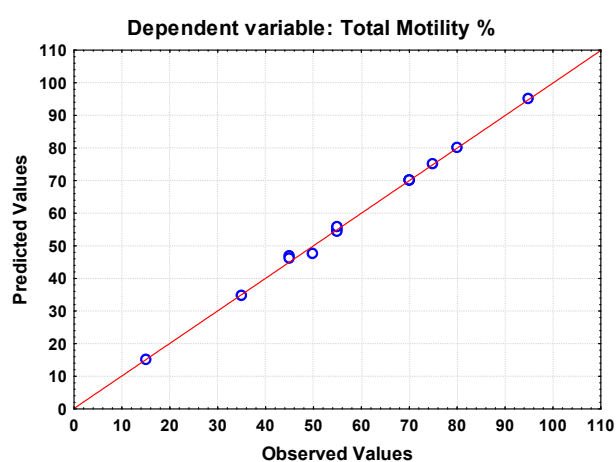| Progressive motility %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998162	0,996328	0,988065	4319,79	9	479,977	15,921	4	3,9804	120,586400	0,000165
HS + B	0,994035	0,988107	0,948462	25833,35	20	1291,667	310,947	6	51,8245	24,923900	0,000342
HS + U	0,949293	0,901157	0,856229	4304,35	5	860,870	472,119	11	42,9199	20,057590	0,000034
B + U	0,833066	0,694000	0,599846	4148,58	4	1037,144	1829,2020	13	140,70780	7,37091	0,002503
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,900439	0,810791	0,687393	10912,41	15	727,494	2546,562	23	110,7201	6,570570	0,000035
U	0,914171	0,835708	0,741827	4957,929000	8,000000	619,741100	974,679900	14,000000	69,619990	8,901769	0,000244

b)

c) VOC predictors

1) 1H-Indole, 5-methyl-2-phenyl-_U
2) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_B
3) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
4) Butanal_U
5) Acetic acid, sodium salt_HS
6) Hexanal_HS
7) 2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine_B
8) Butanal, 2-methyl-_HS
9) 1-Anthracenamine_HS

Figure S8 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **progressive motility** (in percentage); b) Observed vs. Predicted values for sperm progressive motility as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

Total Motility %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998866	0,997733	0,992633	5368,16	9	596,462	12,196	4	3,0491	195,619100	0,000063
HS + B	0,974892	0,950414	0,892564	10746,72	14	767,623	560,689	12	46,7241	16,428840	0,000011
HS + U	0,982579	0,965462	0,907900	6761,08	10	676,108	241,865	6	40,3108	16,772360	0,001299
B + U	0,970786	0,942426	0,891250	6184,67	8	773,084	377,8270	9	41,98078	18,41519	0,000102
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,944713	0,892483	0,795717	17751,25	18	986,181	2138,491	20	106,9246	9,223145	0,000004
U	0,985701	0,971606	0,921917	7870,01	14	562,144	229,990	8	28,7487	19,553710	0,000119

b)

Dependent variable: Total Motility %

c) VOC Predictors
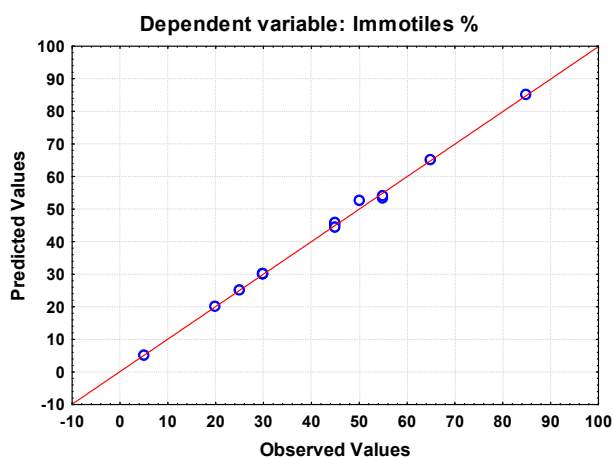
1) 1-Anthracenamine_HS
2) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_B
3) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
4) 1H-Indole, 5-methyl-2-phenyl-_U
5) Acetic acid, sodium salt_B
6) 2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine_B
7) n-Hexane_HS
8) Butanal, 2-methyl-_HS
9) Butanal_U

Figure S9 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **sperm total motility** (in percentage); b) Observed vs. Predicted values for sperm total motility as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

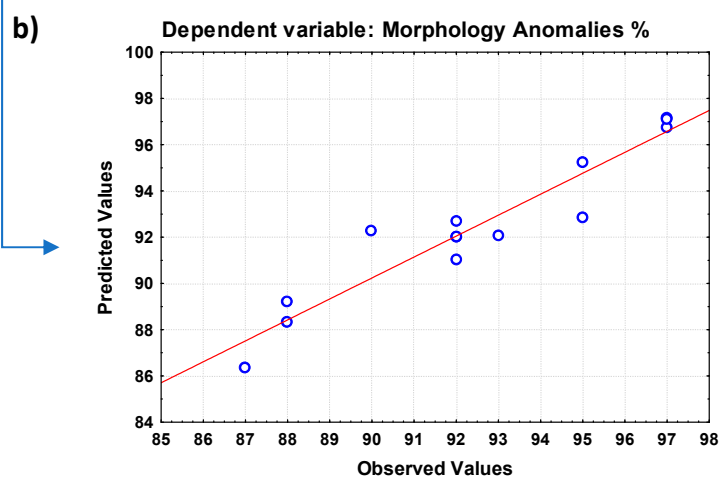| Immotiles %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,998866	0,997733	0,992633	5368,16	9	596,462	12,196	4	3,0491	195,619100	0,000063
HS + B	0,978713	0,957878	0,908736	10680,34	14	762,882	469,657	12	39,1381	19,492040	0,000004
HS + U	0,982579	0,965462	0,907900	6761,08	10	676,108	241,865	6	40,3108	16,772360	0,001299
B + U	0,997490	0,994986	0,982953	9214,68	12	767,890	46,4337	5	9,28673	82,68676	0,000062
HS	/	/	/	/	/	/	/	/	/	/	/
B	0,942534	0,888370	0,798003	17669,45	17	1039,379	2220,296	21	105,7284	9,830657	0,000002
U	0,989445	0,979000	0,923002	7929,90	16	495,619	170,096	6	28,3494	17,482550	0,000995

b)

c) VOC predictors

1) 1-Anthracenamine_HS
2) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_B
3) Fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-_U
4) 1H-Indole, 5-methyl-2-phenyl-_U
5) Acetic acid, sodium salt_B
6) 2-Chloro-4-(4-methoxyphenyl)-6-(4-nitrophenyl)pyrimidine_B
7) n-Hexane_HS
8) Butanal, 2-methyl-_HS
9) Butanal_U

Figure S10 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **sperm immotiles** (in percentage); b) Observed vs. Predicted values for sperm immotiles as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

Morphology Anomalies %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,951849	0,906016	0,847275	137,26	5	27,452	14,239	8	1,7798	15,424110	0,000626
HS + B	0,892523	0,796597	0,669470	208,24	10	20,824	53,171	16	3,3232	6,266150	0,000656
HS + U	0,859512	0,738761	0,651681	113,86	4	28,464	40,262	12	3,3551	8,483729	0,001727
B + U	0,977555	0,955614	0,892205	186,93	10	18,693	8,6824	7	1,24035	15,07068	0,000805
HS	0,732860	0,537084	0,430257	144,459	6	24,077	124,51	26	4,7889	5,027608	0,001534
B	0,662527	0,438943	0,312252	197,28	7	28,182	252,159	31	8,1342	3,464687	0,007413
U	0,957156	0,916148	0,846271	182,91	10	18,291	16,741	12	1,3951	13,110940	0,000054

b)

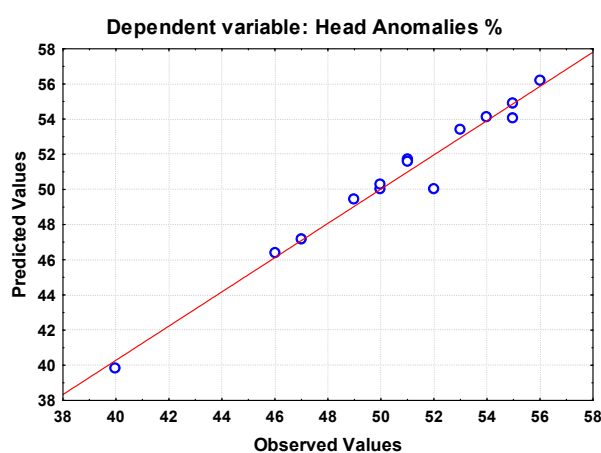Dependent variable: Morphology Anomalies %

c) VOC predictors

1) Propane, 2-(ethenyloxy)-_U
2) n-Hexane_HS
3) Propane, 2-(ethenyloxy)-_B
4) 2-Pentanone_U
5) 3-Hexanone_U

Figure S11 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **morphology anomalies** (in percentage); b) Observed vs. Predicted values for sperm morphology anomalies as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

Head Anomalies%	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,986646	0,973470	0,950730	230,92	6	38,487	6,293	7	0,8990	42,808650	0,000036
HS + B	0,999603	0,999206	0,994842	743,71	22	33,805	0,591	4	0,1477	228,939800	0,000041
HS + U	0,954244	0,910582	0,841034	218,43	7	31,205	21,450	9	2,3833	13,092920	0,000458
B + U	0,950425	0,903307	0,794527	298,94	9	33,216	32,0000	8	4,00000	8,30401	0,003320
HS	0,869429	0,755907	0,609451	622,501	12	51,875	201,01	20	10,0507	5,161331	0,000648
B	0,730072	0,533006	0,388076	726,80	9	80,756	636,789	29	21,9582	3,677693	0,003573
U	0,995314	0,990650	0,965718	342,68	16	21,417	3,234	6	0,5390	39,733730	0,000093

b)

Dependent variable: Head Anomalies %

c) VOC predictors
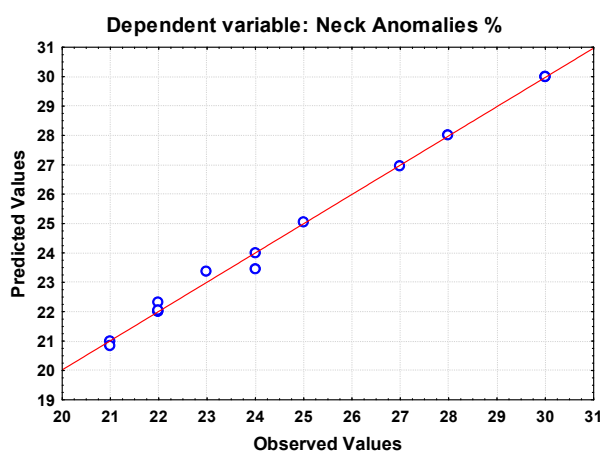
1) 3-Hexanone_U
2) 4-(4-Chlorophenyl)-2,6-diphenylpyridine_U
3) Cyclohexane_B
4) Auramine_HS
5) Pentane_B
6) Hexanal_HS

Figure S12 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **head anomalies** (in percentage); b) Observed vs. Predicted values for sperm head anomalies as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

Neck Anomalies %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,997846	0,995696	0,988809	130,65	8	16,331	0,565	5	0,1130	144,578500	0,000017
HS + B	0,907557	0,823660	0,617931	208,36	14	14,883	44,607	12	3,7173	4,003611	0,010501
HS + U	0,902950	0,815320	0,769149	110,68	3	36,893	25,070	12	2,0892	17,659040	0,000107
B + U	0,813958	0,662528	0,478452	98,35	6	16,391	50,0959	11	4,55417	3,59921	0,031759
HS	0,761019	0,579150	0,461312	147,525	7	21,075	107,20	25	4,2881	4,914799	0,001346
B	0,721634	0,520755	0,299565	414,47	12	34,539	381,430	26	14,6704	2,354334	0,032890
U	0,985085	0,970392	0,918578	151,30	14	10,807	4,616	8	0,5770	18,728430	0,000140

b)

Dependent variable: Neck Anomalies %

c) VOC predictors
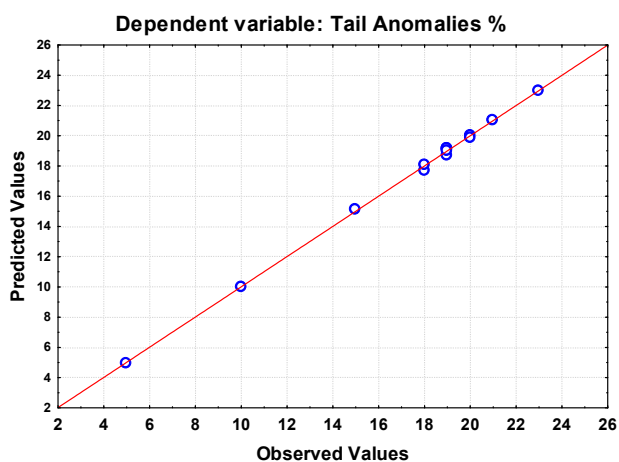
1) Acetic acid, sodium salt_HS
2) Butanal, 2-methyl-_HS
3) 4-Heptanone_U
4) Butanal_U
5) .alpha.-Pinene_B
6) Disulfide, dimethyl_U
7) 2,4,5-Trioxoimidazolidine_U
8) 2-Ethyl-oxetane_B

Figure S13 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for sperm **neck anomalies** (in percentage); b) Observed vs. Predicted values for sperm neck anomalies as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

Tail Anomalies %	Multiple R	Multiple R²	Adjusted R²	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,999462	0,998925	0,993012	4331,05	11	393,732	4,661	2	2,3307	168,935000	0,005899
HS + B	0,999772	0,999545	0,996056	496,44	23	21,584	0,226	3	0,0753	286,495900	0,000293
HS + U	0,608807	0,370646	0,280738	106,70	2	53,351	181,180	14	12,9414	4,122518	0,039108
B + U	0,946014	0,894943	0,821404	275,64	7	39,378	32,3574	10	3,23574	12,16954	0,000358
HS	0,862206	0,743399	0,567830	411,077	13	31,621	141,89	19	7,4680	4,234225	0,002315
B	0,910540	0,829082	0,740205	682,82	13	52,525	140,766	25	5,6306	9,328398	0,000001
U	0,629402	0,396146	0,261957	126,46	4	31,614	192,761	18	10,7089	2,952139	0,048737

b)

Dependent variable: Tail Anomalies %

Predicted Values (y-axis) vs. Observed Values (x-axis)

c) VOC predictors
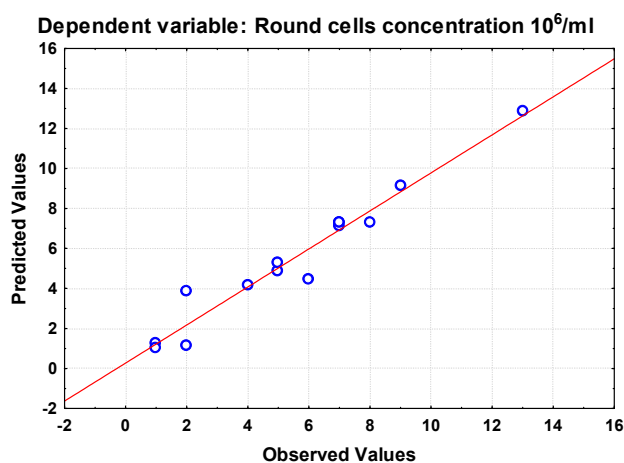
1) n-Hexane_U
2) Butanal_HS
3) .alpha.-Pinene_B
4) Pentanal_HS
5) Acetic acid, sodium salt_HS
6) Propane,2-(ethenyloxy)-_B
7) 2-ethyl-oxetane_B
8) 11H-Dibenzo[b,e][1,4]diazepin-11-one, 5,10-dihydro-5-[3-(methylamino)propyl]-_U
9) Oxime,methoxy-phenyl-_HS

Figure S14 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **sperm tail anomalies** (in percentage); b) Observed vs. Predicted values for sperm tail anomalies as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).

a)

Round cells 10^6/ml	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
HS + B + U	0,975114	0,950847	0,872202	142,15	8	17,769	7,348	5	1,4697	12,090350	0,006950
HS + B	0,918359	0,843384	0,728532	162,86	11	14,805	30,242	15	2,0162	7,343238	0,000306
HS + U	0,988538	0,977206	0,947901	192,57	9	21,396	4,492	7	0,6417	33,345000	0,000063
B + U	0,960418	0,922403	0,835106	192,12	9	21,346	16,1618	8	2,02022	10,56628	0,001458
HS	0,895012	0,801047	0,681675	202,624	12	16,885	50,33	20	2,5163	6,710513	0,000108
B	0,796776	0,634852	0,466323	174,97	12	14,581	100,638	26	3,8707	3,767008	0,002249
U	0,997482	0,994971	0,972338	272,45	18	15,136	1,377	4	0,3443	43,961700	0,001108

b)

c) VOC predictors

1) 2H-Pyrrol-2-one, 1,5-dihydro-1-(4-methoxyphenyl)-5,5-diphenyl-_U
2) Octane_B
3) Acetone_HS
4) n-Hexane_HS
4) Pentanal_HS
5) Propane, 2-(ethenyloxy)-_U
6) 1-(6-Methyl-benzothiazol-2-yl)-3-(4-methyl-benzoyl)-thiourea_B
7) 3-Aminopyrrolidine_B

Figure S15 a) Overall fit of the MLR model described by the test of SS Whole Model vs. SS Residual for **round cells** (in percentage); b) Observed vs. Predicted values for round cells as result of the MLR analysis for the data group HS+B+U based on selected VOC predictors; c) pattern of VOCs used as predictor variables in MLR model as selected by Pareto chart; the order in the list reflects the greatest predictive contribution to the model (VOC name_X; X= B blood or U urine or HS human semen). Notation for numeric values: comma "," is the decimal separator (SI).