Sample size calculation in metabolic phenotyping studies

Elise Billoir, Vincent Navratil and Benjamin J. Blaise

Abstract

The number of samples needed to identify significant effects is a key question in biomedical studies, with consequences on experimental designs, costs and potential discoveries. In metabolic phenotyping studies, sample size determination remains a complex step. This is due particularly to the multiple hypothesis-testing framework and the top-down hypothesis-free approach, with no a priori known metabolic target. Until now, there was no standard procedure available to address this purpose. In this review, we discuss sample size estimation procedures for metabolic phenotyping studies. We release an automated implementation of the Data-driven Sample size Determination (DSD) algorithm for MATLAB and GNU Octave. Original research concerning DSD was published elsewhere. DSD allows the determination of an optimized sample size in metabolic phenotyping studies. The procedure uses analytical data only from a small pilot cohort to generate an expanded data set. The statistical recoupling of variables procedure is used to identify metabolic variables, and their intensity distributions are estimated by Kernel smoothing or log-normal density fitting. Statistically significant metabolic variations are evaluated using the Benjamini–Yekutieli correction and processed for data sets of various sizes. Optimal sample size determination is achieved in a context of biomarker discovery (at least one statistically significant variation) or metabolic exploration (a maximum of statistically significant variations). DSD toolbox is encoded in MATLAB R2008A (Mathworks, Natick, MA) for Kernel and log-normal estimates, and in GNU Octave for log-normal estimates (Kernel density estimates are not robust enough in GNU octave). It is available at http://www.prabi.fr/redmine/projects/dsd/repository, with a tutorial at http://www.prabi.fr/redmine/projects/dsd/wiki.

Key words: sample size determination; chemometrics; metabolic phenotyping

Introduction

In statistical test theory, type I and type II errors refer, respectively, to the rejection of the null hypothesis when it is true (false positive) and the failure to reject the null hypothesis when it is false (false negative). Type I error, which probability of occurring is usually denoted $\alpha$, may seem more important because it corresponds to the incorrect identification of statistically significant features. Type II error, which probability of occurring is usually denoted $\beta$, is nevertheless essential, as it is related to the statistical power of the test ($1-\beta$). This is the ability of a test to detect an effect, if the effect exists. Statistical power analysis can be performed after data analysis (post hoc) to assess what the power was in the study. It can also be conducted before the research study (a priori). In this case, it is typically used in estimating sufficient sample sizes to achieve adequate power.

The sample size determination is based on several components: the minimum expected magnitude of difference, the estimated measurement variability, the desired statistical power, the significance threshold and the type of statistical test used [1]. Besides, in omics studies (mainly genomics, transcriptomics, proteomics and metabolic phenotyping), the high dimensionality of data sets and the context of multiple comparisons have to be accounted for. Different approaches have been developed for...
Sample size determination in other top-down omics approaches, such as genomics or proteomics [2]. These produce relatively less complex and ambiguous measurements of gene/transcript/protein abundance or variants than those estimated for less characterized small chemical molecules. However, there is currently no standard method for sample size estimation in metabolic phenotyping studies, although they cover a wide range of biomedical applications, as well as translational medicine [3, 4].

Traditional approaches are not easily transferable due to the top-down hypothesis-free characteristic (untargeted approach) of metabolic phenotyping studies, where we do not know in advance what the result variables will be. The purpose of metabolic phenotyping is to generate hypothesis, which can be further tested and evaluated. This prevents the use of statistical tools developed for other omics sciences. Some methods have been derived from micro-array experiments such as the concept of power analysis for high-dimensionality data sets [5]. However, it relies on an estimation of the effect size, which precludes the absence of a priori metabolic targets. The effect size refers to the quantitative measure of the strength of a phenomenon and is not accessible in holistic hypothesis-free approaches. Two approaches have been recently introduced to master this issue. Our Data-driven Sample size Determination (DSD) algorithm [6] was shortly followed by the MetSizeR algorithm developed by Nyamundanda and coworkers [7]. These two articles identify the difficulties for sample size estimation in metabolic phenotyping studies and offer original tools based on simulation to achieve this goal. In this review, we will discuss sample size determination in a general context, before focusing on metabolic phenotyping. We also release an automated version of the DSD algorithm.

**Sample size determination**

Sample size determination corresponds to the statistical calculation of the number of samples that are necessary to include in a study, to potentially identify differences among groups. In medical studies, it represents the balance between a sufficient statistical power of test and costs connected to patient inclusion, data collection, analysis and interpretation. If the study is undersized, the risk is that one misses interesting features. This results in a waste of time and money. Alternatively, if the study is oversized, too many patients or animals are submitted to potentially harmful or invasive tests. It also increases the overall cost of the study. Moreover, this might deny access to interesting treatments or diagnostic methods, while delaying the analysis of results, owing to an oversized sample size. This may raise ethical questions. Sample size determination is thus a keystone in experimental design, as it sums up time and financial requirements as well as investigation capacities. It is closely related to the power of test. Power of test represents the risk that one is ready to accept, not to identify a difference, when there is truly one.

Sample size determination is far from being a trivial question. Even in the simplest univariate cases, such as the comparison of population means, sample size determination requires a priori assumptions on data, such as the minimum expected difference (the smallest difference that one expects to observe between the groups under study) and the estimated measurement variability. These two are combined to define a so-called effect size (relative magnitude of the effect). For example, (i) for t-tests comparing the means of two groups with a same variance, the effect size is assessed as $d = \frac{\mu_1 - \mu_0}{\sigma}$ where $\mu_1$ is the mean of group 1, $\mu_0$ is the mean of group 2 and $\sigma^2$ is the common variance; (ii) for t-tests comparing the mean of one group to a given value $\mu_0$, the effect size is assessed as $d = \frac{\mu_1 - \mu_0}{\sigma}$ where $\mu_1$ is the mean of the group and $\sigma^2$ is the variance. The expected difference is hard to evaluate. It can be based on previous measurement or experience, such as pilot studies, or senior researcher expertise. Likewise, the measurement variability should be estimated by expertise or, ideally, extracted from initial pilot data.

Sample size determination also depends on the desired statistical power, the significance threshold and the type of statistical test used, which the investigators choose with respect to the effects under study. The statistical power of test is usually set to 0.8, the significance threshold to 0.05. However, these values can be modified according to the expected results. A common approach for sample size determination known as power approach relies on these parameters [1, 8].

To summarize, the following four quantities have an intimate relationship: sample size, effect size, significance level $\alpha$ and power $1 - \beta$. Given any three, we can determine the fourth.

In the case of several independent observations extracted from a normal distribution of unknown mean $\mu$, one may be interested in testing the fact that $\mu$ is null (null hypothesis $H_0$: $\mu > 0$) or not (alternative hypothesis $H_1$: $\mu > 0$). To solve this issue, different approaches have been developed [9, 10]. To reject the null hypothesis with a probability $P$ superior to $1 - \beta$ when the alternative hypothesis is true, we require Equation (1) to be validated:

$$P(\mu > \frac{Z_{\alpha} \times \sigma}{\sqrt{n}}, H_{\text{true}}) > 1 - \beta$$

where $\sigma$ is the standard deviation of the distribution, $Z_\alpha$ is the upper $\alpha$ percentage point of the standard normal distribution and $n$ the number of samples. If we consider the normal cumulative distribution function $\Phi$ and the effect size $d = \frac{\mu - \mu_0}{\sigma}$, Equation (1) can be rewritten as Equation (2):

$$n > \frac{\left(Z_{\alpha} - \Phi^{-1}(1 - \beta)\right)^2}{d^2}$$

Sample size determination thus requires unknown input parameters that in turn require to be estimated. Although pilot data are the best source to obtain these parameter estimates, they might be hard to collect due to sample availabilities or experimental costs. In these situations, simulation approaches can help in evaluating sample size. However, simulation can also mislead the effect size evaluation when considering effects that have not been previously investigated. For metabolic phenotyping by nuclear magnetic resonance (NMR) or mass spectrometry (MS), spectral databases could also be useful. However, it might be hazardous to exploit them to predict effects other than the original outcomes. It is indeed obvious that one cannot transfer previous knowledge of a certain disease in adults and expect the same results in newborns. The physiological changes observed, especially the organ maturation, lead to strong variations that can profoundly affect effect sizes. It is even tricky to evaluate the metabolic consequences of high-blood pressure from data collected to investigate strokes for instance. On top of that, experimental conditions should be strictly identical. Pilot studies are clearly a gold standard for sample size determination.

A non-optimal sample size could drive increased costs and sample waste that overpass the requirements of a pilot study. Moreover, it would raise serious ethical issues, in terms of
submitting patients to potentially harmful procedures or delaying their access to new strategies. Some authors estimate that a pilot study including 20 samples is generally sufficient to evaluate sample size, in a rigorous manner [8].

Power analysis and sample size determination have been discussed in various contexts (e.g. comparison of means, of proportions, linear regression, logistic regression, survival analysis) [11], especially in the past decades for applications related to medicine [12–20]. In omics sciences, data analysis is complex, owing to the large-scale hypothesis-testing context. Traditionally, data sets are composed of hundreds to thousands of variables, describing tens to hundreds of samples. This is particularly true in genome-wide association studies, but also in transcriptomics and proteomics. One of the most difficult issues is to model the correlations within the data set, which results from the high dimensionality and dependency between variables. These elements are found in all omics sciences. Numerous methods and tools have been developed to master the issue of sample size determination in genomics or proteomics [21–24]. Some software are even available on line for immediate sample size determination [25]. For metabolic phenotyping, before the two algorithms that will be discussed in the next section (DSD and MetSizeR), only one approach had been transferred from proteomics [5, 25]. However, this approach requires input parameters that are usually unknown, such as effect sizes.

Sample size determination in metabolic phenotyping

Comparison with other omics technologies

In metabolic phenotyping studies, many factors complicate the sample size determination, which explains the necessity of new and adapted tools to achieve experimental design. First of all, the metabolic targets are initially unknown. As other omics sciences, metabolic phenotyping is indeed a hypothesis-free approach. It aims at the identification of statistically significant metabolic variations. Parallel quantification is achieved by analytical chemistry procedures, and statistical analyses thus isolate interesting features. Metabolic phenotyping aims at the generation of plausible biochemical hypotheses, which should be further validated or rejected. Interesting variables cannot be identified before analysis. It thus seems impossible to use data other than spectroscopic to achieve sample size determination. We need to know the identifiable metabolites and their concentrations in the sample set under consideration. Once again, the use of previous biological data would focus the analysis on certain compounds and hinder the holistic approach used in metabolic phenotyping. Each sample leads to a spectrum (one statistical individual) that is divided into variables (0.001 ppm wide high-resolution buckets). The acquired data are then processed to identify buckets that correspond to the same metabolic signal. Multivariate statistical analyses are performed to combine metabolic variations that together allow the discrimination of the groups under study. This combination is named the metabolic phenotype. Additional univariate statistics can finally isolate candidate biomarkers. It is thus obvious that potential metabolic targets and associated effect sizes are unknown before the complete analysis.

Second, the metabolome is considered to be more diverse and thus complex than the other ensembles [26]. Biochemical compounds indeed show different biological, chemical and physical properties compared, for instance, with the 4 nucleotides that are combined in DNA sequences. If genomics and metabolic phenotyping deal with an approximately equivalent set of variables (around 35,000 genes and 40,000 metabolites), the observable metabolome is reduced by the detection capacities of analytical methods. For instance, a typical NMR spectrum displays 200 metabolites [27]. In this sense, the metabolome might seem easier to apprehend than the genome, transcriptome or proteome.

Third, metabolic phenotyping data sets are characterized by strong correlations between variables. Multiple signals can actually belong to the same metabolites, and metabolic connections can exist along physiological pathways [28, 29]. Furthermore, these variables exhibit different variances that are used for classification purposes [30]. Most of the methods developed for sample size determination hypothesize a variable independence, which is obviously wrong in metabolic phenotyping studies. These statistical approaches are thus inadequate for sample size determination in metabolic phenotyping studies.

Fourth, researchers have to deal with data sets of high dimensionality (the number of variables greatly exceeds the number of samples). This limitation could easily be mastered by approaches developed for other omics sciences. In particular, Ferreira and co-worker [5] adapted the concept of power analysis for microarray data sets. This approach was transferred to MS-based metabolic phenotyping studies [25]. However, it requires an estimation of effect size, which precludes the hypothesis-free concept and brings us back to the first limitation.

Finally, all omics sciences, owing to high throughput, have to face similar constraints connected to multiple hypothesis-testing contexts. These experiments are usually associated with a high number of tests, increasing the potential risk of false discovery. Numerous methods have been developed to control the type I error risk, especially the family-wise error risk (FWER) controls (such as the Bonferroni correction) and false discovery rate (FDR) measurements (such as the Benjamini-type corrections). The FWER estimates the number of variables associated with a true null hypothesis and that are proposed as significant by any given statistical test. The FDR controls the rate at which variables identified as significant by a given test are in fact associated with a true null hypothesis. The optimal method is still debated. However, it is often recognized that the FWER corrections are too conservative in omics sciences, and FDR measurements are more suited. Different corrections can be used according to the structure of data sets or study aims [31].

A comparison between the different omics regarding sample size determination issues is given in Table 1.

Current approaches: DSD and MetSizeR

Two algorithms, DSD [6] and MetSizeR [7], were developed in the past year to master the different points mentioned above, regarding metabolic phenotyping studies. These two approaches identify similar limitations as the previously developed methods. The two conceptual responses are close, albeit distinct.

MetSizeR and DSD consider that sample size determination in metabolic phenotyping studies depends strongly upon the spectroscopic data and the type of statistical analysis that researchers plan to use. MetSizeR requires a preprocessing step of the raw data, as it can deal with a maximum of 375 variables. In this sense, it is a targeted approach, forcing its user to select variables before analysis. DSD, for its part, includes an automated identification process of biological variables, the
Statistical Recoupling of Variables (SRV) algorithm. SRV allows data reduction, before analysis, to focus on metabolic signals. DSD is thus an untargeted approach, capable of dealing with complete raw data. MetSizeR is based on probabilistic principal component analysis (PPCA), probabilistic principal components and covariates analysis (PPCCA) or dynamic PPCA (DPPCA), whereas DSD uses orthogonal partial least square regression (O-PLS). MetSizeR has been designed to take into account the existence or absence of pilot data. DSD requires the use of pilot data, which, from our point of view, is an essential element in estimating a correct effect size.

Significance testing in MetSizeR is based on two-sample t-statistics, whereas DSD uses one-way analysis of variance. Both tests are intensively used in metabolic phenotyping studies and both algorithms can easily be modified to manage other statistical tests. The number of statistically significant variables is measured by controlling the FDR. MetSizeR uses the Benjamini–Hochberg correction [32], which controls the FDR with an assumption of independence between the tests. In DSD, we support the use of the Benjamini–Yekutieli correction [33], which controls the FDR under negative correlation. It is more conservative than the Benjamini–Hochberg correction, or the Benjamini–Hochberg–Yekutieli correction [33] adapted to positively correlated variables. Metabolic signals can, indeed, show both positive and negative correlations. Negative correlations can be observed when they are, for instance, involved in the same obstructed pathway. Metabolites above the blockade will exhibit increased concentrations, and thus positive correlations. Metabolites beneath it will show decreased concentrations, and thus positive correlations. A correlation between metabolites located in the upstream and downstream of the blockade will, however, be negative. We recommend the use of the most conservative correction (the Benjamini–Yekutieli correction) to prevent the misidentification of nonsignificant features.

MetSizeR uses a permutation approach to estimate the null distribution of t-statistic discriminating variables. DSD uses receiver operating characteristics as well as resampling under the null hypothesis for cross and independent model validations. Finally, MetSizeR offers a larger adaptive panel of statistical analyses (PPCA, PPCCA, and DPPCA), whereas DSD has been developed to deal with the most commonly used approach (O-PLS). Both algorithms deliver graphic results on which the estimated optimal sample size can be read. These elements are summarized in Table 2.

The two algorithms allow the use of pilot data to evaluate the effect size. MetSizeR also enables the use of simulated data. These, once again, do not take into account the specificity of the planned study. It is obviously possible to simulate human data, for instance, using available databases. However, it is not possible to simulate efficiently spectroscopic data corresponding to a disease not previously investigated, or a pathophysiological effect. Moreover, experimental conditions (including sample preparation, storage, data acquisition) should be identical in the simulated or already acquired data, and in the planned experiments. Pilot studies thus seem inevitable despite their costs and sample availabilities. Inadequate sample size estimations would raise more ethical issues and costs than a properly designed pilot study. The importance of pilot studies is emphasized in all fields of medical science [34, 35].

### Table 1: Comparison between the different omics regarding sample size determination.

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Genomics (genome wide association study)</th>
<th>Transcriptomics (Array/RNA-sequencing)</th>
<th>Proteomics</th>
<th>Metabolic phenotyping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensionality</td>
<td>3 (10^5) base pairs (\sim 35,000) genes</td>
<td>&gt;100,000 transcripts</td>
<td>&gt;100,000 proteins</td>
<td>~40,000 metabolites</td>
</tr>
<tr>
<td>Complexity</td>
<td>Low (4 nucleotides)</td>
<td>Medium/high (splicing)</td>
<td>Medium (post-traduction)</td>
<td>High</td>
</tr>
<tr>
<td>Detection capacity limitation</td>
<td>Low</td>
<td>Yes (positive and negative)</td>
<td>Yes (positive and negative)</td>
<td>No</td>
</tr>
<tr>
<td>Variables known before analysis</td>
<td>Low</td>
<td>Yes (positive and negative)</td>
<td>Yes (positive and negative)</td>
<td>Yes (positive and negative)</td>
</tr>
<tr>
<td>Dependency between variables</td>
<td>Yes</td>
<td>Yes (positive and negative)</td>
<td>Yes (positive and negative)</td>
<td>Yes (positive and negative)</td>
</tr>
<tr>
<td>Usual false positive/discovery rate measurement</td>
<td>Bonferroni</td>
<td>Bonferroni</td>
<td>Bonferroni</td>
<td>Absence of standard procedure</td>
</tr>
<tr>
<td></td>
<td>Benjamini–Hochberg correction</td>
<td>Benjamini–Hochberg correction</td>
<td>Benjamini–Hochberg correction</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Comparison between the MetSizeR and DSD algorithms

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>MetSizeR</th>
<th>DSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted/untargeted</td>
<td>Targeted (375 bins maximum)</td>
<td>Targeted/Untargeted</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>PPCA, PPCCA, DPPCA</td>
<td>O-PLS on SRV clusters</td>
</tr>
<tr>
<td>Pilot data requirement</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Significance measurement</td>
<td>Two-sample t-statistics</td>
<td>One-way ANOVA</td>
</tr>
<tr>
<td>FDR measurement</td>
<td>Benjamini–Hochberg</td>
<td>Benjamini–Yekutieli</td>
</tr>
<tr>
<td>Validation</td>
<td>Permutation approach</td>
<td>ROC and Resampling under the null hypothesis</td>
</tr>
<tr>
<td>Output</td>
<td>Figure</td>
<td>Figure</td>
</tr>
<tr>
<td>Statistical environment</td>
<td>R</td>
<td>Matlab and Octave</td>
</tr>
<tr>
<td>Running mode</td>
<td>GUI/GTK</td>
<td>Command line</td>
</tr>
<tr>
<td>Operating system</td>
<td>Mainly Windows</td>
<td>Linux, Mac Os X, Windows</td>
</tr>
</tbody>
</table>

PPCA, probabilistic principal component analysis; PPCCA, probabilistic principal component and covariates analysis; DPPCA, dynamic PPCA; O-PLS, orthogonal partial least square regression; SRV, statistical recoupling of variables; ANOVA, analysis of variance; FDR, false discovery rate; ROC, receiver operating characteristics; GUI, graphical user interface; GTK, GIMP ToolKit.
Sample size calculation

Description of the dsd algorithm and implementation

Description

Here, we will exemplify the DSD algorithm on a data set published elsewhere, discriminating Caenorhabditis elegans nematodes according to their age (L4 larvae versus gravid adults) [36, 37]. High-resolution magic angle spinning NMR is used on entire worms to determine their metabolic composition. Statistical analyses are then performed to discriminate groups, according to biochemical variations and finally identify candidate biomarkers discriminating samples, with respect to the age effect.

DSD is an algorithm designed to determine an appropriate sample size in NMR-based metabolic phenotyping studies. The main idea of the DSD procedure is to focus on the spectroscopic data from a small pilot cohort, and then to expand the data set to a larger size without additional information. The sample size is derived from the number of samples needed to identify at least one statistically significant metabolic variation (biomarker discovery) or a maximum of statistically significant variations (metabolic exploration) [6]. A training cohort of at least 20 spectra (10 in each group) should be considered. Spectroscopic metabolic variables are initially identified by the SRV procedure [38, 39], available free of charge. SRV is an automated variable size bucketing procedure aiming at the identification of biological variables of interest. It is based on statistical relationship between consecutive variables inherited from high-resolution bucketing (0.001 ppm wide buckets). This aspect is fundamental. There is a loss of power, owing to standard bucketing procedures and the thousands of resulting variables. The identification of NMR metabolic signals and the removal of analytical noise compensate this loss, and allow an efficient sample size determination compared with other approaches.

SRV should be run before the use of DSD by selecting efficient recoupling parameters (typical singlet peak base width, bucketing resolution, correlation threshold). Once the metabolic variables identified, the distributions across the data set are considered and their density probabilities are computed, using Kernel smoothing density or log-normal estimates. From the estimated distributions, new samples are simulated and this simulation step is long, and may require more than 1 h using a standard laptop for Kernel estimates. The process is shortened by the use of log-normal estimates.

Step 1: Identification of SRV metabolic variables

Recoupling parameters (singlet peak base width, bucketing resolution, correlation threshold) should be evaluated before using the SRV algorithm. See Navratil et al. for details on this step [39]. A reasonable number of metabolic variables should be identified (between 150 and 200 SRV variables). The data set must contain a water exclusion area filled with zeros.

Step 2: Estimation of and simulation from the density probability of metabolic variables

Inverse cumulative density probabilities are computed for each SRV variable using Kernel density estimate function or log-normal distribution fitting. Random numbers are drawn in the [0, 1] interval and used as input quantiles of the inverse cumulative density probabilities. This generates an expanded data set of 400 samples. A noise-removing filter cancels out the generation of signals in noise areas. This simulation step is long, and may require more than 1 h using a standard laptop for kernel estimates.

Step 3: Sample size determination

One thousand equally sized data sets of various sizes (20, 50, 100, 150, 200, 300, 400 samples) are then randomly generated. The number of statistically significant variations is evaluated by one-way analysis of variance, and the FDR is measured according to the Benjamini–Yekutieli correction. The DSD procedure results in a 2 × 7 table (Data) including the mean numbers of variables highlighted as statistically significant for the 7 data set sizes (across 1000 equally sized data sets) in the first row, and the corresponding standard deviations in the second row. A plot is automatically displayed (Figure 1), on which the optimal sample size can be determined depending on the purpose of the planned study: biomarker discovery or metabolic exploration, with respect to the expected number of statistically significant variations.

Finally, we evaluated MetSizeR on our data set. Unfortunately, MetSizeR does not include any data reduction package, despite its limitation to 375 variables. Hence, raw data were reduced to 178 variables using the SRV algorithm before running MetSizeR. The following standard parameters, as described in the MetSizeR guide book, were used: n1 = 4, n2 = 4, p = 178, prop = 0.2, covars = FALSE, ncovar = 0, model = PPCA, plot.prop = FALSE, target.fdr = 0.05, Targeted = FALSE. The result of the procedure is given in Figure 2. MetSizeR determined a sample size of 30 samples, which is in good agreement with the
Sample size determination is not straightforward in metabolic phenotyping. Two algorithms are now available online (Data-driven Sample size Determination and MetSizeR). Pilot studies seem necessary to avoid inappropriate sample size determination.

**Key Points**

- Sample size determination is not straightforward in metabolic phenotyping.
- Two algorithms are now available online (Data-driven Sample size Determination and MetSizeR).
- Pilot studies seem necessary to avoid inappropriate sample size determination.

**Acknowledgements**

The authors thank Dr Colin Stern for his careful proofreading.

**Funding**

Benjamin J. Blaise is supported by the Fédération pour la Recherche Médicale, the Société Française d’Anesthésie et Réanimation, the Académie Nationale de Médecine, the Association de Néonatologie de Port-Royal and the City of Lyon.

**References**


