Databases: Working tools for networks

Meschy F., Morand-Fehr P., Bas P., Sauvant D.

in

Ben Salem H. (ed.), Nefzaoui A. (ed.), Morand-Fehr P. (ed.).
Nutrition and feeding strategies of sheep and goats under harsh climates

Zaragoza : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 59

2004
pages 257-264

Databases: Working tools for networks

F. Meschy, P. Morand-Fehr, P. Bas and D. Sauvant
Physiologie de la Nutrition et Alimentation, Unité Mixte de Recherche 791, INRA-INA Paris-Grignon, 16 rue Claude Bernard, 75231 Paris Cedex 05, France

SUMMARY – The objectives of the paper are to inform the network's members of our laboratory's experience on the creation of databases and their utilization in meta-analysis in order to analyze bibliographic or experimental data in the field of sheep and goat nutrition. It would be possible to create new databases or to open the existing databases to interested colleagues. But to be successful in this field, it would be necessary to proceed with a strict strategy. In an initial stage, it is necessary to clarify the objectives and organization of the database particularly if it is already working, the kind and quality of data to be included in the database and the conditions to be appointed for database members who can enrich and use the databank. Moreover the rules of management must be well defined. Nevertheless, some difficulties can be encountered due to non orthogonal experimental design, interactions between variables, discriminating levels of the chosen filters or keys of acceptance and inadequate information in published papers, etc. The paper gives examples of the actual difficulties resulting from a meta-analysis using a database on the factors modifying the composition of adipose deposits in lambs. The meta-analysis from databases is a good strategy to create new information and general rules but the risks of an inadequate interpretation exist if the scientific approach used is not very rigorous.

Key words: Database, meta-analysis, nutrition, sheep, goat.

RESUME – *"Les banques de données : Des outils de travail pour les réseaux". L'objectif de cette communication est d'informer les membres du réseau, de l'expérience de notre laboratoire dans la création de base de données et leur utilisation pour des méta-analyses sur des données expérimentales ou bibliographiques dans le secteur de la nutrition des ovins et des caprins. Il serait possible de créer de nouvelles bases ou d'ouvrir les bases existantes aux collègues intéressés. Mais pour réussir de telles opérations, il serait nécessaire de suivre une stratégie rigoureuse. D'abord, il est nécessaire de clarifier les objectifs et la situation de la base, en particulier si elle est déjà en fonctionnement, le type et la qualité des données devant être stockées dans la base, et les conditions à réunir pour qu'un membre du réseau enrichisse ou utilise la base. En outre, les règles de gestion de la base de données doivent être bien définies. Néanmoins, quelques difficultés peuvent être rencontrées relatives à des dispositifs expérimentaux non orthogonaux, aux interactions entre variables, au niveau de sélection par les filtres choisis ou les clés d'acceptation et à une information insuffisante dans les articles publiés. Cette communication donne des exemples des difficultés réelles rencontrées au cours d'une méta-analyse utilisant une base de données sur la composition des tissus adipeux d'agneaux et sur les facteurs de variation de cette composition. La méta-analyse à partir de la base de données est une bonne stratégie à utiliser pour créer de l'information nouvelle et des lois générales entre variables. Mais les risques d'une mauvaise interprétation des résultats existent si l'approche scientifique suivie n'est pas très rigoureuse.*

Mots-clés : *Banque de données, méta-analyse, nutrition, mouton, chèvre.*

Recent evolution of scientific network objectives

For 20 years, researches on small ruminants industry have largely been developed. But frequently difficulties are encountered to draw clear conclusions from all these researches due apparent discrepancies between results. In such cases, it has been shown that methodologies using meta-analysis might be interesting to establish general principles often in equation form and to make the bibliographic results of the whole more consistent (Bas and Morand-Fehr, 2000; Sauvant, 2001; Sauvant and Bas, 2001).

On the other hand, the objectives of research networks such as the FAO-CIHEAM network on sheep and goats were greatly modified during the last years (Morand-Fehr, 2000). Previously, this kind of network had to provide experimental results and scientific reflections mainly to the sector experts. Now it has been requested more and more to be open to peoples not having a good scientific knowledge in the field and particularly to meet requirements of the social demand when it was requested to resolve the scientific discrepancies and to present scientific conclusions very clearly and simply.

Interest of databases to meet these new requirements

Scientists have initiated the implementation of databases to reply to various questions from basic sciences, extension services or media. So, in the field of small ruminants nutrition, the researchers of our laboratory are implementing databases on the ruminal digestion (S. Giger-Reverdin), the duodenal flux (D. Sauvant), energy and protein nutrition of dairy goats (Ph. Schmidely), phosphorus in sheep and goats (F. Meschy), the composition of lamb fats and muscle lipids (P. Bas) and the composition of ewe and goat milks (Ph. Schmidely). The analysis of these data banks has allowed the creation of new and original information, which has been published in scientific journals such as Bas and Morand-Fehr (2000) and Schmidely and Sauvant (2001). They emphasized the interest of this method particularly when a lot of factors can influence the results by creating a complex situation and when clear information is difficult to find from bibliographic data. In this case, the research networks can be well adapted to enrich these databases, to enlarge their field of competency, to establish even more efficient principles. New structures or organizations are needed to establish successful databases.

Objectives of this paper

The objectives of this paper are to inform the network's members of the experience of our laboratory on the databases creating, and their utilization particularly in meta-analysis, and on the risks of wrong interpretation. As we are willing to open some of our databases if some network members might be interested to collaborate in the work of databases, or if they wish to create new ones in this field, this paper makes proposals to know how these databases would have to be managed particularly about the acceptance and analysis of data.

Typology of databases

In order to analyse the difficulties encountered in the implementation, management and utilization of a database, it is essential to analyse the parameters allowing to make a typology of existing databases. Table 1 describes 5 different approaches to type databases which correspond to 5 questions to solve for implementing and managing a database.

On a first stage, it is necessary to clarify the objectives of the database either to cumulate data only or to establish new scientific principles. For example, a database on ruminal digestion could foresee the risks of clinical or sub-clinical acidosis in ewes and goats.

The organization of a database, at a given time such as the period when it is opened to the network members must be well identified to solve the problems linked to this situation. The database either has to be created or it already exists.

It is necessary to know if the database accepts exclusively bibliographic data or also unpublished data. In the case of unpublished data, a process to validate them can be implemented or not.

The database access may be free or limited. It seems very important to establish clearly defined rules to access to or get out the database or to enrich the base.

Finally, the management rules must also be indicated and accepted by every base member by appointing a person or a group in charge of its management and by defining clearly every responsibility.

In practice, currently, the Table 1 is used to characterize a database when someone wants to become one of its members.

Some methodological aspects for meta-analysis approach

In the following lines we shall use a bibliographic database as example but the main principles are suitable for any kind of databases.

Table 1. Parameters allowing to identify the different kinds of databases

1. Database objectives
   1.1. Only pool basic data without statistic treatment
   1.2. Well identified but limited analyses
   1.3. Large field of data without identified objectives to reply to current questions for the future
   1.4. To make quantitative analysis to get the information more objective and synthetic from bank data
2. Position of the database at a well defined stage
   2.1. The databank does not exist and must be created because the subject is new and only emergent
   2.2. The databank exists but incomplete; it could be broadened by dependent variables or enriched with extra-observations
   2.3. The databank exists and works but its introduction into the network can make it more active again, and allows to reach a new level of information
3. Contents of the database
   3.1. Bibliographic untreated data (without calculated parameters)
   3.2. Bibliographic data with new calculated parameters
   3.3. Experimental data unpublished but of good quality to be published
   3.4. Experimental data unpublished but of variable quality
   3.5. Mixture of the cases 3.2, 3.3, 3.4 in various proportions
4. Bank access (type of data to include in the databank)
   4.1. Access decided by the first maker of the databank
   4.2. Access only for the members of a limited group accepting the bank constitution
   4.3. Access only for the network members
   4.4. Access only for specialists who are network members or not
   4.5. Access for everybody without constitution
5. Bank management
   5.1. Bank managed by one appointed person (who can be the pioneer)
   5.2. Bank managed by a selected members group
   5.3. Bank managed by the decisions of the member assembly
   5.4. Bank managed as 52 or 53 with an coordinator (or manager)

The quantitative analysis of the literature is necessary to obtain a consistent knowledge and information which can be used at a higher level of organization (e.g. modelling). It can be useful to explain some contradictions between data and to compare one experiment with the literature.

The meta-analysis of the literature is aimed at maximizing the valorization of published data by creating new knowledge. It allows the identification of areas where scientific information is missing and also to avoid redundant experimental work.

The main difficulties in quantitative analysis of the literature are related to the structure of the database itself presenting generally an incomplete table of data (data with gaps). Moreover some interesting results may come from non standard experimental design and it is always difficult to produce homogeneous data (assessments and recalculations from original papers). Almost all database show unbalanced and non orthogonal design. For these reasons, priority must be given to the graphical presentation of data and the study 2 by 2 of the variables keeping linearity when possible. Usually the data are studied under the general linear model of analysis of variance considering the relationship between variables under 3 aspects: global, "inter" and "intra". We must pay a special attention to the residual variations and interactions. Before building it, the scope of the database must be well defined so as to avoid too large data set.

In the second stage, filters or keys of acceptance of publication results in the database must be set. These filters are very important and must always be respected. Some of them are technical: for example variables the most interesting for the user must always be sufficiently discriminating. The confidence degrees of data have also to be defined: either only published papers or published papers and some unpublished data; in this case, the problem of the assessment of their validity must be solved. Some miscellaneous filters may also be used such as language, geographic area… Filtering data have important consequences on the power of the database: "open" filters lead to large

databases which can be reduced in sub-databases by encoding specific variables; quite narrow filters allow more specific databases which can be concatenated. In this case a careful attention must be paid to the data organization. Probably the setting and the respect of the filter rules are the main difficulties in sharing databases.

Building a database from the literature needs the encoding of the data from each reference: informative variables (experimental targets, methods, etc.), dependent variables and independent variables, as shown in Table 2 for a database on lamb adipose tissue and muscle composition.

So the database may allow to quantify and synthesize information pulled out from results of each publication considered separately and to generate general principles. As Fig. 1 shows, the pooling of a lot of results could allow to predict the $C_{18:2}$ percentage in fat deposits from the fatty acid composition of milk fat consumed by unweaned lambs. In the same way the melting point of their fat deposits in subcutaneous and perirenal adipose tissues of weaned lambs will be estimated with equations using the percentage of stearic acid which is the most efficient acid to predict the melting point (Fig. 2).

An other interest of the meta-analysis approach is to compare results from different experiments carried out in very various dietary conditions and managements. In order to compare the effects of one dietary treatment, it is important to consider the greatest number of independent variables. For example, to analyse the effects of nitrogen sources such as soybean meal, alfalfa meal or fish meal on the fatty acid composition of adipose tissues in lambs, we have to take into account the type of diets and particularly the proportions of roughages and concentrate in diets, and the type of the main energy and nitrogen sources as Fig. 3 shows in order to get unbiased ideas about the specific effect of the main nitrogen source independently of the other variables.

We have to keep in mind that the power and the consistency of a meta-analysis depend on the relevance of encoding (bias). Bias could result from the selection references, to erroneous interpretation, leading to overestimate or to underestimate the effect of one treatment or another. For a same type of diet, the lack of homogeneity for a factor could lead to an invalid conclusion, as illustrated in Fig. 4, where the sex effect (male, female, castrate) was studied on fatty acid composition of adipose tissues and muscles of weaned lambs. In this case, we expected that the fatty acid composition of castrates is on a middle position between females and males. But the results are different because the castration reduced lamb growth rate and consequently before fattening period, they are left on pasture or in sheep house and this pre-fattening period can strongly modify the composition of fat deposits in castrated lambs and interfere on the composition results at slaughtering.

Conclusions

On the basis of these examples, it appears that using databases for meta-analysis present a lot of risks if the approach of users is not scientifically accurate. Even in this case some risks might occur due to the subjective choice of used basic or calculated variables and of the references selected by filters.

Moreover, it was observed that when the objectives of data analysis are slightly modified, generally we have to come back to basic papers included in the base to find new variables necessary to solve questions requested in the new objectives. Consequently it is always necessary to keep a set of papers included in the database, on a paper support to be able to adapt quickly the database to new objectives.

The opening of a database to several persons must allow to improve the efficiency of the database but it increases the above mentioned risks. To limit them, everyone who wants to participate to a database must think over his objectives and the role that he can play in the working group of database. Before database opening, the working group must define the conditions for participating to the database and to be accepted in the working group, the filter parameters to apply when one reference or experimental results can be included in the database. When a new problem must be

Table 2. Variables used in the databank: lamb feeding, carcass and fats (Bas and Morand-Fehr, 2000)

I. Coded variable (CV)				
CV linked to research work	CV linked to animals	CV linked to diets	CV linked to diet distribution	CV linked to adipose tissues (AD)
Reference	Animal groups	Type of diet	Distribution	Type of AD
Research team or country	Number of animals	Type of milk	Individual	Location of AD
Experimental factors	Breed	Type of forages	In group	Types of lipids
	Sex	Type of concentrates	Level of distributed or ingested feeds	
		Main energy source		
		Main protein source		
		Main fat source		

II. Explicative variables		
Diets, roughage and concentrates composition	Diet feed	Animals
Crude fiber or NDF	Main energy source	Initial live weight (LW)
Fat or ether extract (EE)	Energy/total energy	Final live weight (LW)
Crude protein	Fat/total fat	Age at weaning
Energy density	Main protein source	Age at slaughter
	Energy/total energy	Growth rate
	Fat/total fat	Carcass rate
	Duration of experimental diets distribution	Fatness index

III. Explained variables (EV)		
Physical on chemical EV in adipose tissues or muscles	Individual fatty acids (FA) %	Group FA %
Lipids	C12:0,C14:0,C14:1,C15:0,C16:0	Saturated and unsaturated FA
Cholesterol mg/100g	C16:1,C17:0,C17:1,C18:0,ΣC18:1	Mono-saturated FA
Triacylglycerols (%)	C18:1,c9,C18:1tr,C18:2,C18:3,CLA,C20:0	PUFA n3 and PUFA n6
Melting point	C20:1,C20:2n6,C20:3n6,C20:4n6,C20:5n3,	PUFA n6/n3
Firm index	C22:4n6,C22:5n3,C22:6n3	Saturated and total odd FA
		Branched chain FA (BRFA)
		BR FA + odd FA
		Unidentified FA

solved, the working group must find an agreement and the coordinator must have the responsibility to apply the working group's decisions[1].

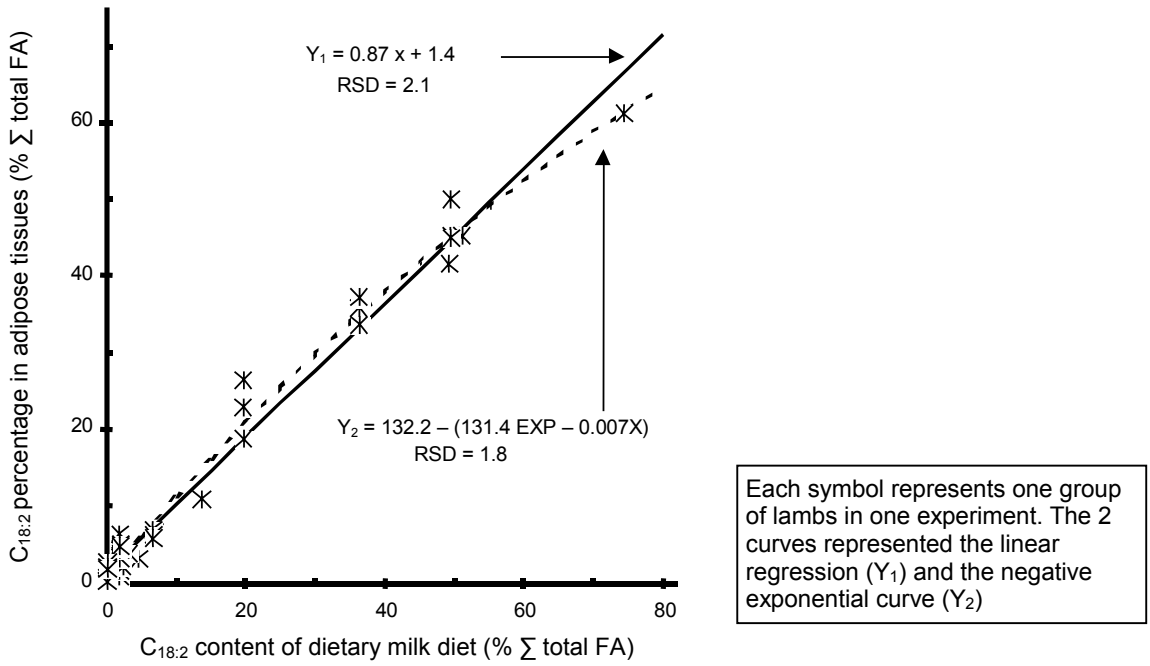


Fig. 1. Relationship between the $C_{18:2}$ percentage in fat deposits and the $C_{18:2}$ percentage of milk fat ingested by unweaned lambs (Bas *et al.*, 2001, unpublished data).

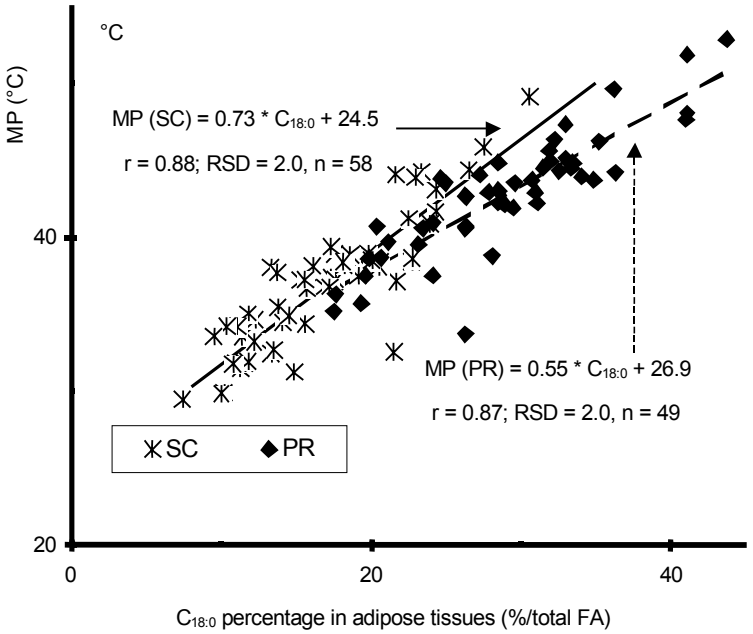
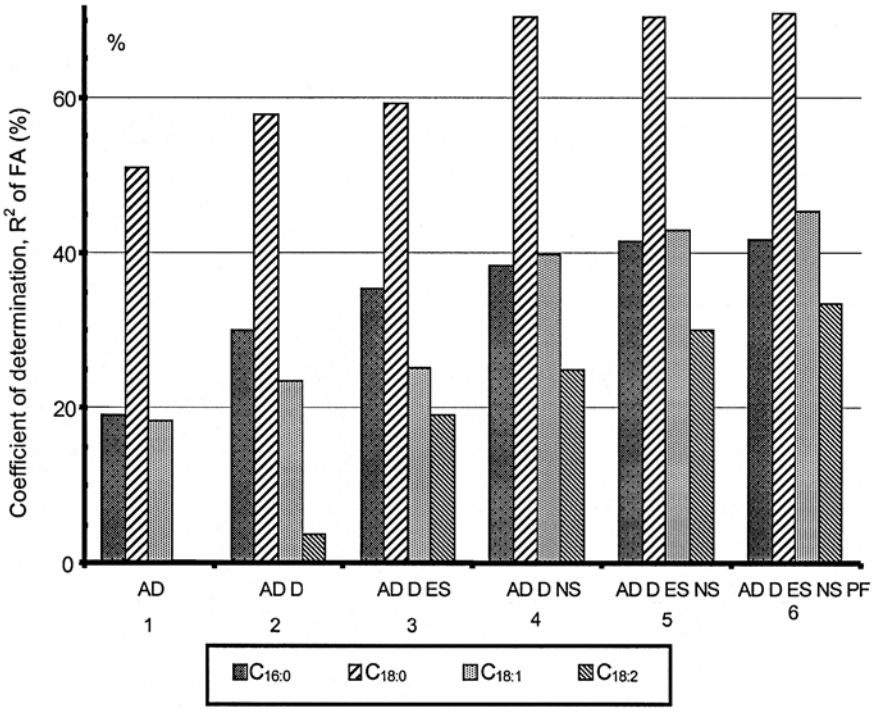


Fig. 2. Relation between the melting point (MP) and the $C_{18:0}$ percentage of subcutaneous (SC) and perirenal (PR) adipose tissues in weaned and unweaned lambs (Bas and Morand-Fehr, 2000).

[1] NB To have more information concerning the databases on sheep and goat nutrition and products, please contact one of the authors.



AD: Adipose tissues; D: Type of diets (roughage alone, roughage + concentrate, complete diet, concentrate alone); ES: Type of the main energy source (barley, maize, wheat); NS: Type of the main nitrogen source (alfalfa meal soybean meal, fish meal); PF: Percentage of forage in the diets.

1: Analysis with one dependent variable (AD)
2: Analysis with two dependent variables (AD + D)
3: Analysis with three dependent variables (AD + D + ES)
4: Analysis with three dependent variables (AD + D + NS)
5: Analysis with four dependent variables (AD + D + ES + NS)
6: Analysis with four dependent variables (AD + D + ES + NS) and a covariate (PF)

Fig. 3. Variations of the coefficients of determination of fatty acid (FA) percentage in adipose tissues of lambs with including the main types of constituents of the diets in the statistical analysis-of-variance and analysis-of-covariance. Source (Bas *et al.*, 2001, unpublished data).

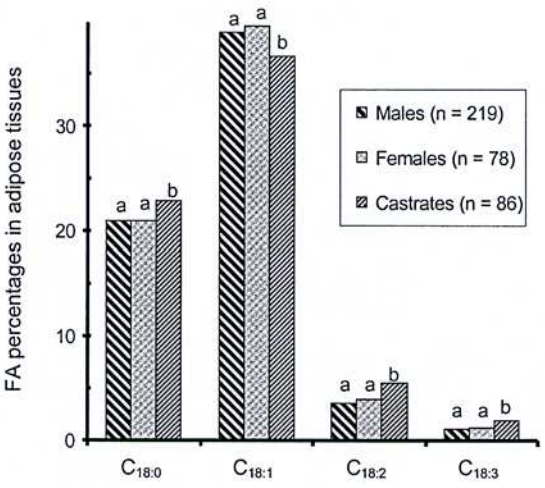


Fig. 4. Biais in comparison of effect of sex upon fatty acid (FA) composition in adipose tissue and muscle of weaned lambs. Source Bas and Morand-Fehr, 2000, unpublished data.

References

Bas, P. and Morand-Fehr, P. (2000). Effect of nutritional factors on fatty acid composition of lamb fat deposits. *Livestock. Prod. Sci*., 64: 61-79.

Morand-Fehr, P. (2000). Interest of network research: The case of the FAO-CIHEAM Cooperative Research Network in sheep and goats. In: *Proc. 7th International Conference on Goats*, Vol. 2, Tours-Poitiers (France), 15-20 May 2000, pp. 907.

Sauvant, D. (2001). Quelles limites ingestives et digestives pour les hautes productrices de lait. *Bull. Acad. Vét. de France*, 154: 217-223.

Sauvant, D. and Bas, P. (2001). La digestion des lipides chez le ruminant. *INRA Prod. Animales*, 14(5): 303-310.

Schmidely, P. and Sauvant, D. (2001). Taux butyreux et composition de la matière grasse du lait de chèvre et de brebis selon l'alimentation. *INRA Prod. Animales*, 14(5): 337-354.