

ECHANTILLONNAGE DANS LES ENQUÊTES DESCRIPTIVES

M. Sanaa ^[1], G. Gerbier ^[2], M. Eloit ^[3]
et B. Toma ^[3]

Résumé

L'objectif de cette article est de décrire les différentes étapes d'un plan de sondage dans le contexte des études descriptives en épidémiologie animale : définition de la population à étudier, choix de la méthode de sondage, calcul de la taille de l'échantillon. Le choix du plan de sondage est lié aux objectifs de l'enquête, aux outils d'observation et au budget disponible. Les calculs de la taille de l'échantillon ont été développés pour chacune des méthodes de sondage.

On a également développé les principes d'échantillonnage dans le cadre de la qualification d'un cheptel ou d'une région. Le calcul de la taille de l'échantillon et l'interprétation des résultats nécessite, dans le cadre de la qualification, l'estimation de la sensibilité et la spécificité des procédures de qualification à l'échelle de l'élevage.

Summary

The aim of this paper is to describe the different steps of a sampling plan within the context of descriptive studies in animal epidemiology : definition of the population to study, choice of the sampling method, calculation of the size of the sample. The choice of the sampling plan is linked to the objectives of the study, to the observation tools and to the available funds.

The principles of sampling within the frame of the qualification of a herd or of a region are also developed. The calculation of the size of the sample and the interpretation of the results is only possible, within this qualification frame, with an estimation of the sensitivity and the specificity of the qualification process at the herd level.



Les enquêtes descriptives ont pour "objet de décrire les caractéristiques d'un phénomène de santé dans une population, son évolution dans le temps, sa répartition, son évolution dans l'espace [7].

[1] Ecole nationale vétérinaire d'Alfort, Laboratoire d'épidémiologie et de gestion de la santé animale, 94700 Maisons-Alfort, France

[2] C.N.E.V.A., Laboratoire central de recherches vétérinaires, 22 rue Pierre Curie, 94700 Maisons-Alfort, France

[3] Unité pédagogique Maladies contagieuses, Ecole nationale vétérinaire d'Alfort, 94704 Maisons-Alfort, France

Si l'on voulait connaître avec exactitude et précision le nombre d'élevages français atteints de salmonellose clinique bovine, il faudrait les visiter tous. Une enquête exhaustive de ce type s'appelle un recensement. Périodiquement, l'administration effectue un tel recensement à propos de certaines caractéristiques essentielles de la population : sexe, âge, formation, etc.. En santé animale, dans le cadre des prophylaxies obligatoires, des recensements sont effectués régulièrement sur le statut sanitaire des élevages vis-à-vis de certaines maladies comme la tuberculose et la brucellose bovines. Mais le plus souvent, lorsqu'on s'intéresse à des nouvelles maladies ou lorsqu'on souhaite suivre dans le temps l'évolution d'une maladie, le coût et les délais des enquêtes exhaustives seraient prohibitifs. On peut se contenter d'une connaissance approchée (ou estimation) des caractéristiques de la population considérée, à l'aide d'une enquête réalisée sur une partie de la population, appelée **échantillon**. La validité et la précision de l'estimation dépendront essentiellement des plans d'échantillonnage (ou de sondage*) choisis.

Un plan d'échantillonnage comporte trois étapes principales :

- Définition de la population à étudier,
- Sélection de l'échantillon avec définition de la méthode de sondage qui sera utilisée,
- Détermination de la taille de l'échantillon

Le plan d'échantillonnage n'est qu'une étape dans la réalisation d'une enquête descriptive (enquête par sondage) qui comprend d'autres étapes dont la nature et la complexité varient beaucoup d'une enquête à une autre. Des étapes essentielles

comme la construction du questionnaire, l'analyse et l'interprétation des résultats font l'objet d'autres articles dans ce numéro spécial.

L'objectif de cette présentation est de décrire les différentes étapes d'un plan d'échantillonnage dans le contexte des études descriptives en épidémiologie animale.

Deux situations peuvent être alors distinguées : les études visant à **estimer la prévalence** d'une maladie dans une région déterminée et celles dont l'objectif est de **qualifier un élevage ou une zone "indemne"** d'une certaine maladie.

Dans la première situation, on peut s'intéresser à l'estimation de la prévalence des **élevages** atteints ou à l'estimation de la prévalence des **animaux** atteints.

Dans la deuxième situation, on peut distinguer deux types de maladies : les maladies soumises réglementairement à un dépistage obligatoire et celles pour lesquelles le dépistage n'est pas obligatoire. La prévalence des maladies à dépistage obligatoire a beaucoup baissé depuis la mise en place des systèmes de prophylaxie. Ainsi, le coût de détection d'un foyer a fortement augmenté puisque l'on continue à tester l'ensemble des élevages et des animaux. En outre, de plus en plus de zones sont considérées indemnes et il faut donc adapter les systèmes de surveillance de ces maladies à dépistage obligatoire. Pour les autres maladies, il n'existe pas de structure permettant le recensement des cas ni de structure organisée de collecte de données.

Une approche par sondage est donc nécessaire pour ces deux types de maladies afin de qualifier les élevages ou des zones d'élevage.

I - DEFINITION DE LA POPULATION A ETUDIER

La première décision à prendre dans l'élaboration d'un plan d'échantillonnage concerne la définition de la population sur laquelle portera l'enquête. La population à étudier doit être définie avec le plus grand soin (par exemple, la population des élevages laitiers en France, la

population des élevages laitiers vendant du lait cru, la population canine de la région parisienne ...). Cette définition inclut la définition des unités à échantillonner (unités de sondage) qui sont l'objet de l'observation.

* Les termes « sondage » et « échantillonnage » sont employés indifféremment par certains auteurs. En fait, l'échantillonnage consiste simplement en la détermination d'un échantillon alors que le sondage comprend, en plus, la réalisation de l'étude de l'échantillon et l'analyse des résultats [7].

Dans certains cas, les unités de sondage sont les animaux : par exemple si l'on cherche à estimer la fréquence d'un motif de saisie à l'abattage des jeunes bovins de boucherie. Dans d'autres cas ce sera l'élevage, lorsqu'on cherche par exemple à déterminer la prévalence des cheptels infectés par le virus de la leucose enzootique bovine. Ce seront parfois aussi les communes, les départements, les régions, etc. Mais, dans la plupart des cas, compte tenu de la structure de la population animale, on a recours à deux ou plusieurs unités de sondage à la fois.

Pour des raisons pratiques, dans des cas où l'unité de sondage choisie est l'animal il arrive souvent que des groupes d'individus soient utilisés comme unité initiale de sondage. Par exemple, tout en ayant comme objectif d'estimer la prévalence de l'agalaxie contagieuse chez les brebis (unités choisies), l'unité initiale peut être le troupeau car une liste de toutes les brebis de la population étudiée serait très difficile à obtenir. Si, par contre, l'objectif de l'étude est d'estimer la prévalence des troupeaux atteints par l'agalaxie contagieuse, les troupeaux seront les unités de sondage alors que les animaux seront les unités d'observation (ou de mesure) permettant de qualifier l'élevage (atteint ou non atteint par l'événement pathologique étudié). Il n'est pas toujours nécessaire de tester l'ensemble des animaux pour qualifier l'élevage. Le cheptel sera considéré atteint si on l'observe au moins un animal réagissant parmi les animaux testés.

Dans d'autres situations, les unités d'observation constituent aussi les unités de sondage ; il en est ainsi, par exemple, lorsqu'on cherche à estimer la prévalence de contamination du lait par *Listeria monocytogenes* en choisissant l'élevage comme unité de sondage et le lait de tank comme unité d'observation.

Le choix de l'unité de sondage dépend de l'objectif de l'étude et des contraintes liées à la structure de la population. En santé animale, les décisions prophylactiques sont en général prises au niveau du troupeau. Les textes réglementaires s'appuient sur des seuils de taux de troupeaux infectés pour définir une région indemne ou pour décider du changement de la stratégie de lutte contre la maladie. Ainsi, l'objectif des études

descriptives est, le plus souvent, l'estimation de la prévalence des élevages atteints et, rarement, celle du taux des animaux atteints.

Lorsque l'objectif est d'estimer la prévalence des élevages atteints, il faut préciser la procédure permettant d'identifier les élevages infectés. Cette identification peut se fonder soit sur des observations effectuées sur l'ensemble ou une partie des animaux de l'élevage, soit sur une observation effectuée au niveau de l'élevage (test sur le lait de tank par exemple).

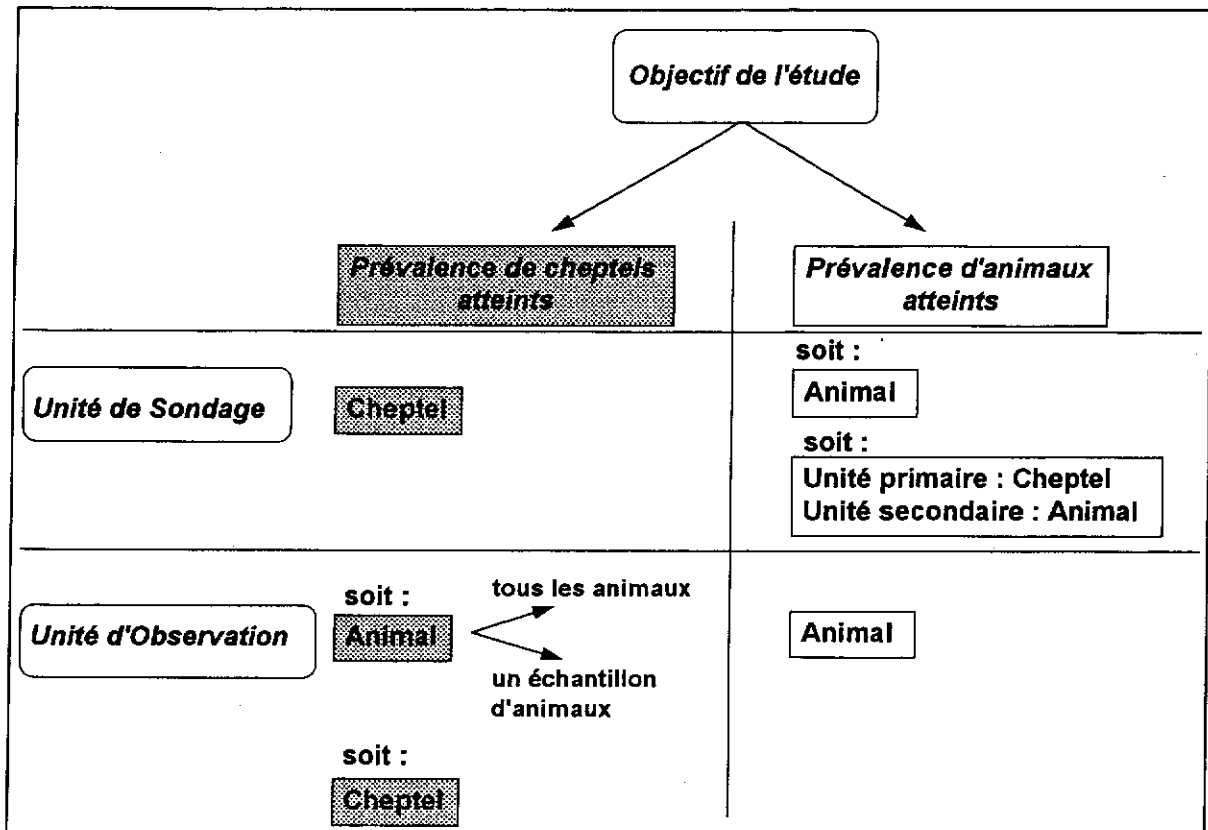
L'estimation des taux d'animaux atteints peut avoir un intérêt dans la planification des prophylaxies et dans le calcul des coûts économiques liés au dépistage et à la lutte contre la maladie. En raison du regroupement des animaux dans les élevages, la constitution de l'échantillon se fait en général en deux étapes : tirage au sort des élevages, puis tirage au sort des animaux à partir des élevages tirés. Mais on peut envisager dans certains cas de tirer au sort directement les animaux (par exemple : enquêtes réalisées dans les abattoirs).

Ces différentes situations peuvent être résumées dans la figure 1.

Après la définition de la population, intervient la phase de recherche d'une base de sondage c'est-à-dire, théoriquement, d'une liste exhaustive des individus de la population. Il arrive très souvent que cette liste soit incomplète, recouvrant mal la population à étudier. Il arrive aussi que cette liste ne soit pas disponible. L'existence de cette liste, comme on le verra plus tard, conditionne fortement le choix de la méthode d'échantillonnage.

Enfin, il reste un point à soulever concernant la distinction entre la population étudiée (population cible) et la population réellement enquêtée (population parente). Ces deux populations diffèrent à cause des contraintes de réalisation de l'enquête, des imperfections de la base de sondage et des non-réponses ou refus de participation à l'enquête. En toute rigueur, les extrapolations ne pourront être faites qu'à la population parente

Figure 1 : Choix des unités de sondage et des unités d'observation en fonction de l'objectif de l'étude



II - PROCEDURE DE SELECTION : METHODES D'ECHANTILLONNAGE

L'échantillon observé dans l'étude doit être représentatif de la population-cible.

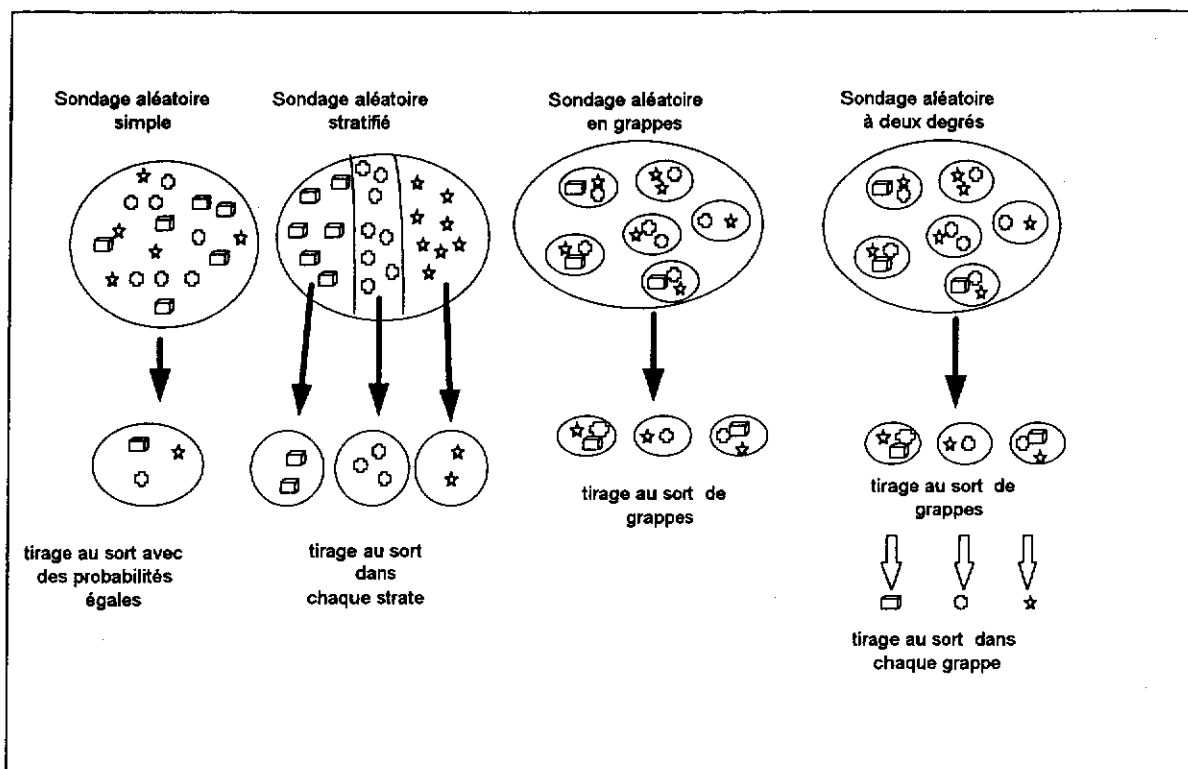
En théorie, les seules méthodes permettant d'obtenir un échantillon **représentatif** sont les méthodes dites **aléatoires**. Elles se fondent sur un tirage au sort donnant à chacune des unités de sondage de la population étudiée une chance égale ou, du moins, une chance connue non nulle. En toute rigueur, seules les méthodes aléatoires permettent d'extrapoler les

caractéristiques observées sur l'échantillon à celles de la population étudiée (ou cible).

Il existe plusieurs méthodes d'échantillonnage ou de sondage aléatoire : (figure 2)

- Sondage aléatoire simple
- Sondage aléatoire stratifié
- Sondage aléatoire en grappes
- Sondage aléatoire à plusieurs degrés

Figure 2 : Les principales méthodes de sondage aléatoire



A - LES METHODES D'ECHANTILLONNAGE

□ SONDAGE ALEATOIRE SIMPLE

Il consiste à dresser une liste exhaustive de toutes les unités de sondage de la population étudiée (base de sondage) et à procéder à un tirage au sort à l'aide par exemple d'une table de nombres au hasard. Avec cette procédure, toutes les unités de sondage ont la même probabilité d'appartenir à l'échantillon. L'échantillon obtenu est appelé **échantillon aléatoire simple**. Cependant, cette méthode devient très difficile à appliquer dans certaines conditions de terrain. Par exemple, si l'on est dans un élevage de 500 brebis et que l'on désire un échantillon de 50, il est plus pratique d'utiliser la procédure d'**échantillonnage aléatoire systématique**. Elle consiste à choisir au hasard un nombre entre 1 et 10 (par exemple 8). Les 8^{ème}, 18^{ème}, 28^{ème} ... , 498^{ème} brebis passant dans le couloir de contention formeront l'échantillon.

Dans les procédures de sélection précédentes, le choix de l'échantillon est laissé à l'aléas du tirage au sort. On ignore ainsi toute une série d'informations que la base de sondage peut

fournir : les unités sont identifiables et peuvent être donc classées et regroupées...

Le sondage aléatoire simple est, en fait, **rarement utilisé** :

- Si la population est très hétérogène par rapport au problème étudié, les résultats obtenus par sondage aléatoire simple peuvent être peu précis. Dans la mesure où l'on dispose d'informations permettant de répartir les individus en classes ou en groupes plus homogènes, on a intérêt à utiliser le sondage aléatoire stratifié.
- Si l'on ne dispose pas de base de sondage ou si l'on cherche à réduire le coût de l'enquête, on utilise le sondage en grappes ou le sondage à plusieurs degrés.

□ SONDAGE ALEATOIRE STRATIFIE

Si avant l'étude, on sait qu'une ou plusieurs parties de la population ont des propriétés différentes, on peut essayer de tenir compte de ces informations *a priori* dans un objectif d'amélioration de la qualité de l'estimation. Par exemple, si l'on sait avant l'étude que les élevages porcins engraisseurs sont plus touchés que les élevages reproducteurs par le virus de la maladie

d'Aujeszyk, on peut planifier autrement la sélection en stratifiant la population d'élevages porcins étudiée. On peut donc décider de faire deux strates, la première regroupant les élevages engraisseurs et la deuxième les élevages de reproducteurs. A l'intérieur de chaque strate, un échantillon aléatoire simple est obtenu. La réunion des échantillons simples forme l'échantillon global désiré, appelé **échantillon aléatoire stratifié**. D'une façon générale, la stratification consiste à considérer chaque sous-population (strate) comme une population particulière et à réaliser à l'intérieur de chaque strate un échantillon aléatoire simple indépendant des sondages réalisés dans les autres strates.

□ SONDAGE EN GRAPPES ET SONDAGE A PLUSIEURS DEGRES

Dans de nombreux sondages, la population peut être considérée comme un ensemble de groupes d'individus appelés **grappes** (dans le sondage en grappes) et **unités primaires** (dans le cas du sondage à deux degrés) : on se base en général sur des regroupements naturels des unités de sondage. Par exemple, les élevages sont des grappes d'animaux, les communes correspondent à des grappes d'élevages ...

On considère que la population est constituée de plusieurs agrégats (grappes) des unités concernées et on réalise un échantillon aléatoire simple parmi ces grappes. Si, à l'intérieur des grappes sélectionnées on retient tous les individus, la méthode est appelée **sondage en grappes**. Si, par contre, on ne retient qu'un échantillon des individus, la méthode est appelée **sondage à deux degrés**.

Le choix des critères de regroupement ou de classement des unités de sondage est primordial. Dans le cas de la stratification, la subdivision de la population doit fournir des strates **les plus homogènes possibles** car l'efficacité de la stratification est fortement liée à l'homogénéité à l'intérieur des strates. Tandis que dans le cas des sondages en grappes, le regroupement des unités doit aboutir à des grappes **les plus hétérogènes possibles**, car l'efficacité d'un sondage en grappes dépend de cette hétérogénéité.

En pratique, les sondages en grappes ou à plusieurs degrés sont plus rapides à réaliser et moins coûteux que les sondages aléatoires simple ou stratifié. En effet, en concentrant les prélèvements ou les interviews dans des lieux géographiques limités on arrive à réduire le temps de réalisation de l'étude et les frais de déplacement des enquêteurs. Cependant, pour

une taille d'échantillon donnée, le sondage en grappes est en général moins précis que le sondage aléatoire simple. Mais le sondage en grappes se justifie par son coût qui est plus faible que celui des sondages aléatoires simple et stratifié.

De ce fait, et pour des raisons pratiques évidentes, le sondage en grappes est très utilisé dans les enquêtes d'épidémiologie animale, et notamment son cas particulier le sondage aléatoire à plusieurs degrés. Généralement, les grappes tirées au premier degré sont les élevages, au second degré l'unité secondaire est l'animal, c'est celle que l'on aurait pris directement comme unité de sondage dans la méthode de sondage aléatoire simple.

La présentation des différents types de sondage peut laisser penser que dans une enquête par sondage, on choisit un seul de ces types. Or, il arrive souvent que ces différentes méthodes soient **combinées** dans un même plan de sondage. On peut par exemple suivre le schéma suivant dans une enquête par sondage dans la population bovine :

- Stratifier les élevages selon la double classification type de production et taille de l'élevage (1-10, 11-20, 21-30, 31-40 ... animaux par élevage),
- Tirer dans chaque strate un échantillon d'unités primaires (élevages),
- Tirer dans chaque unité primaire un nombre fixe ou une proportion d'animaux.

B - CHOIX D'UNE METHODE D'ECHANTILLONNAGE

Le choix de la méthode d'échantillonnage est directement influencé par :

- les objectifs de l'enquête
- le budget disponible
- les outils d'observation.

□ OBJECTIFS DE L'ENQUETE

Le choix du type de sondage est le résultat d'un compromis permettant de satisfaire les objectifs de l'enquête au meilleur coût et dans les meilleurs délais. Ce choix dépend donc des enjeux économiques et sanitaires associés aux résultats de l'enquête.

Dans le choix de la méthode de sondage, il est important de savoir si les résultats de l'enquête ont pour objectif de décrire la population dans sa globalité ou, également, de décrire des sous-populations, auquel cas il pourrait être indispensable de sur-représenter certaines de ces sous-populations dans l'échantillon. Par exemple, si l'on souhaite connaître la prévalence d'une maladie en France mais également la répartition géographique de la maladie, il faudra des tailles d'échantillon suffisamment importantes dans chaque région.

Le choix de la méthode de sondage dépend aussi de la nature de la population étudiée. Si la population est dispersée géographiquement (population canine) et si l'enquête doit se faire à domicile chez les propriétaires, on comprend bien que pour des raisons de coût lié au déplacement des enquêteurs il est préférable de concentrer les enquêtes dans un certain nombre de communes. La méthode choisie sera donc un sondage à deux degrés après stratification sur les communes.

Enfin, la disponibilité d'une base de sondage ou, à défaut, d'une liste d'unités intermédiaires est indispensable pour avoir recours à une méthode de sondage aléatoire.

□ BUDGET DISPONIBLE

Dans la pratique, la taille de l'échantillon est souvent déterminée par une simple opération qui divise le budget disponible pour l'enquête par le

coût unitaire de l'enquête. Cette approche peut être suffisante si l'on ne souhaite pas des résultats avec une grande précision. Cependant, ce calcul peut entraîner des résultats trop imprécis pour l'usage que l'on voudrait en faire.

Dans le cas d'enquêtes successives destinées par exemple à suivre l'évolution d'une maladie au cours du temps, si les tailles de l'échantillon choisies sont trop faibles, on ne pourra mettre en évidence que des évolutions importantes.

Le choix de la méthode de sondage et de la taille de l'échantillon devraient donc tenir compte à la fois des objectifs de l'enquête et des contraintes matérielles de l'enquête. La combinaison entre différentes méthodes de sondage (stratification, sondage à plusieurs degrés) permet en général de diminuer le coût de l'enquête tout en respectant le souhait d'obtenir des résultats d'assez bonne qualité.

□ OUTILS D'OBSERVATION

La méthode de sélection de l'échantillon ne sera pas la même si l'enquête est réalisée par téléphone ou par des interviews. Dans ce dernier cas, les coûts imposent la concentration des interviews dans un certain nombre de zones d'enquêtes.

L'utilisation d'examen clinique ou de test de diagnostic très coûteux conditionne, elle aussi, le choix de la méthode de sondage.

III - TAILLE DE L'ECHANTILLON

Après avoir présenté les différentes procédures de sélection des unités de sondage permettant de garantir la représentativité de l'échantillon, il reste maintenant à déterminer la taille de l'échantillon. Le choix de la procédure d'échantillonnage étant fixé, il convient de répondre à la question "quelle est la taille de l'échantillon nécessaire pour atteindre une précision donnée ?"

Lorsque l'on veut estimer une proportion avec une précision donnée, il y a parfois une certaine ambiguïté dans la formulation du problème.

Par exemple, si l'on cherche à estimer la prévalence des cheptels infectés par le virus de la

maladie d'Aujeszky dans une zone géographique déterminée, "on souhaite estimer cette prévalence (p) avec une précision de 5 %" peut être compris de deux façons différentes. En effet, la valeur de 5 % peut être interprétée :

- Soit comme une **précision absolue** souhaitée $p \pm 5\%$,
- Soit comme une **précision relative** $p \pm 0,05 \times p$

La précision relative s'obtient en divisant la précision absolue par la valeur de la prévalence estimée :

$\Delta R = \frac{\Delta}{p}$ où ΔR est la précision relative, Δ la précision absolue et p la prévalence estimée.

Par exemple, pour $\Delta = 1\%$ et $p = 5\%$, $\Delta R = 20\%$.

La précision d'une estimation se traduit toujours par le calcul d'un intervalle dit de confiance pour une probabilité donnée. Par exemple, sur un échantillon représentatif de 500 animaux on a observé 50 séropositifs. Le but du calcul de l'intervalle de confiance est d'encadrer ce pourcentage observé (10 %) par deux valeurs limites telles qu'on ait une certaine confiance que la vraie valeur du pourcentage dans la population totale soit comprise entre ces deux limites. Le calcul des probabilités propose dans cet exemple 8 % comme limite inférieure et 13 % comme limite supérieure. Cet intervalle (8 %-13 %) est calculé pour un risque d'erreur de 5%, c'est-à-dire qu'on a 5% de risque que la vraie valeur du pourcentage dans la population soit en réalité en dehors de cet intervalle [6].

Plus l'intervalle de confiance est petit plus le sondage est précis. Cette précision dépend à la fois de la taille de l'échantillon et de la dispersion de la caractéristique dans la population étudiée. La dispersion est mesurée par la variance.

Le mode de calcul de la taille de l'échantillon dépend de l'objectif de l'étude et du mode d'échantillonnage choisi. On abordera le calcul de la taille de l'échantillon dans les situations suivantes :

- Estimation de la prévalence des cheptels atteints en distinguant le cas où « le classement » des cheptels (en indemne ou atteint) se base sur un test collectif ou sur des tests appliqués aux animaux,
- Estimation de la prévalence des animaux atteints

Le calcul de la taille de l'échantillon nécessaire pour la qualification d'un élevage ou d'une région sera traité en IV.

A - ESTIMATION DE LA PREVALENCE DES CHEPTELS ATTEINTS

L'objectif est d'estimer la prévalence des cheptels atteints. Deux cas de figure se présentent, le premier correspond au cas où le classement du cheptel en "atteint" ou "indemne" se fonde sur le

résultat d'un test réalisé à l'échelle de l'élevage (test sur le lait de tank par exemple). Le second cas de figure est celui où le classement du cheptel repose sur les résultats individuels d'un test réalisé sur les animaux (test de tuberculination par exemple).

1 - CAS D'UN TEST COLLECTIF

On prendra l'exemple d'une étude descriptive dont l'objectif est de déterminer la proportion de laits de tank contaminés par *Listeria monocytogenes* dans une zone de collecte de lait destiné à la fabrication de produits au lait cru, regroupant 2000 élevages laitiers.

SONDAGE ALEATOIRE SIMPLE

Le nombre d'élevages nécessaire est calculé à partir de la formule suivante [3] :

$$n_i = \frac{z^2 p(1-p)}{\Delta^2} \quad (1)$$

$$n = \frac{1}{\frac{1}{n_i} + \frac{1}{N}}$$

où :

- n : le nombre d'élevages à tirer
- n_i : taille de l'échantillon quand la population est considérée infinie (quand la taille de l'échantillon ne dépasse pas le dixième de la taille de la population).
- p : prévalence
- N : taille de la population
- Δ : précision absolue souhaitée
- z : valeur de la variable normale centrée réduite pour un risque donné (par exemple 1,96 pour 95 %)

Pour notre exemple, en fixant à 5% l'hypothèse de la fréquence de contamination des laits de tank, à 2 % la précision absolue et pour un niveau de confiance de 95 % on trouve $n_i = 456$ et $n = 372$. Pour une certitude pratique (niveau de confiance à 99,73 %, $z=3$) on trouve $n_i = 1069$ et $n = 697$.

Les deux formules de calcul (1) peuvent être combinées en une seule [3] :

$$n = \frac{Np(1-p)}{\frac{(N-1)\Delta^2}{z^2} + p(1-p)} \quad (2)$$

Cette dernière formule peut être exprimée en fonction de l'erreur relative (ΔR) :

$$n = \frac{N(1-p)}{\frac{(N-1)(\Delta R)^2 p}{z^2} + (1-p)} \quad (3)$$

Les tables 1a et 1b donnent des exemples de taille d'échantillons en fonction de l'hypothèse sur la proportion et la précision relative souhaitée. Les prévalences présentées dans les tables 1a et 1b varient entre 1 et 50 %. Le nombre d'unités à tirer pour des prévalences supérieures à 50 % peut être obtenu en lisant dans la colonne correspondant au complément de p (par exemple pour 90 % on utilise la colonne correspondant à 10 %).

SONDAGE ALEATOIRE STRATIFIÉ

Le calcul de la taille d'un échantillon stratifié est plus complexe que dans le cas précédent. On peut utiliser la formule proposée par Farver et al.[3] :

$$n = \frac{\sum_{i=1}^S N_i^2 p_i (1-p_i)}{\frac{N^2 \Delta^2}{z^2} + \sum_{i=1}^S N_i p_i (1-p_i)} \quad (4)$$

où :

- n : le nombre d'élevages à tirer
- S : le nombre de strates
- N_i : taille de la i ème strate
- p_i : prévalence dans la i ème strate
- N : taille de la population
- Δ : précision absolue souhaitée
- z : valeur de la variable normale centrée réduite pour un risque donné
- w_i : poids de répartition pour la strate i

Pour comprendre l'utilisation de cette formule, on va utiliser l'exemple de contamination du lait par *Listeria monocytogenes*. D'après les résultats d'une étude analytique [5], on a observé des risques de contamination selon la taille et le type de stabulation des élevages laitiers. A partir de ces données on peut définir trois strates en fonction de la taille du troupeau où les proportions de contamination sont différentes :

- Strate 1 : < 20 animaux
- Strate 2 : 20 à 40 animaux
- Strate 3 : > 40 animaux

Si l'on suppose que la contamination globale est de 5 % dans la population étudiée, en tenant compte de la répartition des 3 strates (en se fondant sur la base de sondage) et du différentiel de risque de contamination entre les 3 strates, on peut proposer la répartition suivante (tableau I) :

Tableau I : Caractéristiques des strates étudiées

STRATES	FREQUENCE DANS LA POPULATION N_i	PROPORTION DE LAITS CONTAMINES P_i
< 20	400	2,0 %
20 - 40	600	3,5 %
> 40	1000	6,4 %

Si l'on choisit de faire un sondage stratifié proportionnel à la taille des strates les w_i dans la formule (4) sont égaux à la fréquence relative de chacune des strates : $w_i = N_i / N$.

L'utilisation dans ces conditions de la formule (4) en prenant la même précision que dans le cas du sondage aléatoire simple (2 %) fournit un $n = 348$ pour un niveau de confiance de 95 % et un $n = 663$ pour un niveau de confiance de 99,73 %.

Il existe d'autres méthodes permettant d'optimiser la répartition de l'échantillon entre les strates. On peut citer la méthode d'allocation de Neyman qui donne des fractions de sondage proportionnelles aux tailles des strates et des écarts-types du caractère étudié dans chaque strate.

SONDAGE EN GRAPPES

Le calcul de la taille de l'échantillon dans ce type de sondage demande un certain nombre d'informations spécifiques de la population des grappes étudiées. Dans l'exemple de contamination du lait de tank par *Listeria monocytogenes*, on peut considérer que les grappes sont formées par les élevages livrant du lait à la même fromagerie. Donc, au lieu de tirer au sort les élevages on sélectionne les fromageries qui participeront à l'étude. Le calcul du nombre de fromageries à tirer au sort ne peut se faire que si l'on dispose par avance d'une certaine information sur les fromageries de la région. Ces informations sont résumées dans le tableau suivant (tableau II) :

Tableau II : Caractéristiques des grappes
(laiteries) considérées.

LAITERIES	NOMBRE D'ELEVAGES PAR LAITERIE	NOMBRE ATTENDU D'ELEVAGES LIVRANT DU LAIT CONTAMINE
A	150	1
B	200	8
C	104	2
D	300	20
E	98	5
F	124	1
G	117	14
H	180	18
I	213	2
J	114	8
K	223	12
L	177	10

Farver et al.[3] ont proposé la formule suivante pour calculer le nombre de grappes à tirer au sort :

$$m = \frac{Ms_g^2}{\frac{M\Delta^2\bar{N}^2}{z^2} + s_g^2}$$

avec

$$s_g^2 = \frac{1}{M-1} \sum_{i=1}^M (a_i - pN_i)^2 \quad (5)$$

où :

- m : le nombre de laiteries (grappes) à tirer
- s_g^2 : est la variabilité du nombre d'élevages livrant du lait contaminé par laiterie
- M : le nombre de grappes dans la population
- a_i : le nombre d'élevages livrant du lait contaminé dans la laiterie i
- \bar{N} : la taille moyenne des grappes
- Δ : précision absolue souhaitée
- z : valeur de la variable normale centrée réduite pour un risque donné

L'application de cette formule pour une précision absolue de 2 % donne m = 6 pour un niveau de confiance de 95 % et m = 8 pour un niveau de confiance de 99,73 %.

Le tableau III donne les tailles des échantillons requises pour l'estimation de la proportion des laits de tank contaminés par *Listeria monocytogenes* en fonction des différentes stratégies d'échantillonnage. Le sondage aléatoire stratifié demande moins d'élevages que le sondage aléatoire simple. Cependant, le gain en taille d'échantillon n'est pas très élevé (34 et 36 élevages). Le gain serait plus élevé si les

prévalences à l'intérieur des strates étaient davantage différentes.

La difficulté avec le sondage en grappes est que le nombre d'élevages retenus ne peut être déterminé qu'après le tirage au sort. En prenant comme base d'estimation le nombre moyen d'élevages par laiterie, on trouve une taille d'échantillon (1002 et 1336 élevages) beaucoup plus élevée que celle requise par le sondage aléatoire simple. Mais l'enquête est concentrée sur 6 ou 8 laiteries seulement.

Tableau III : Nombre d'élevages à tirer au sort en fonction des méthodes d'échantillonnage : estimation de la prévalence de *Listeria monocytogenes* dans le lait de tank

METHODE D'ECHANTILLONNAGE	NIVEAU DE CONFIANCE	
	95 %	99,73 %
Sondage simple	372	697
Sondage stratifié	348	663
Sondage en grappes	6 grappes : 1002	8 grappes : 1336

2 - CAS DE TESTS INDIVIDUELS

L'unité principale de sondage reste toujours l'élevage. Les modalités d'échantillonnage des élevages ainsi que la taille de l'échantillon seront choisies de la même façon que dans le cas précédent. Les échantillons d'élevages peuvent être constitués selon les méthodes de sondage simple, stratifié, en grappes ou à plusieurs degrés.

La qualification des élevages, par contre, repose sur un ensemble de résultats de tests réalisés sur les animaux. On distingue deux situations : dans la première on teste tous les animaux des cheptels choisis, et dans la seconde on teste seulement un échantillon d'animaux.

Pour déterminer le nombre d'animaux à tester avec un niveau de certitude donné (95 % par exemple), il faut avoir une hypothèse sur la prévalence minimale des animaux atteints par la maladie étudiée à l'intérieur des élevages (prévalence intra-élevage minimale : P.I.M.). Pour une maladie contagieuse donnée, on peut supposer que si la maladie s'installe dans un élevage elle atteint très vite une prévalence minimale, par exemple de 10 à 20 %. Cette valeur seuil de prévalence est à déterminer en fonction des études disponibles ou des hypothèses sur la dynamique de la maladie à l'intérieur des élevages.

Quand le nombre d'animaux à l'intérieur d'un élevage est grand, on peut déduire la taille de l'échantillon à partir de l'inégalité suivante :

$$1 - C \leq (1 - PIM)^n \quad (6)$$

où :

C : le niveau de confiance

n : le nombre d'animaux testés

P.I.M. : la prévalence intra-élevage minimale et $(1 - PIM)^n$ est la probabilité que les n animaux testés provenant d'un élevage dont la prévalence est égale à P.I.M. soient tous indemnes.

L'inégalité (6) propose que cette probabilité soit inférieure ou égale à $(1 - C)$. Si, par exemple, on fixe C à 95 %, cela veut dire que la probabilité de passer à côté d'un élevage atteint doit être inférieure à 5 %.

La taille de l'échantillon doit être alors :

$$n \geq \frac{\log(1 - C)}{\log(1 - PIM)} \quad (7)$$

Cette formule n'est valable que si n est inférieur au dixième de la taille du cheptel.

Cannon et Roe [1] ont proposé une formule permettant d'approcher le nombre minimal d'animaux à examiner pour qualifier des cheptels de petite taille :

$$n \geq \left\{ 1 - (1 - C)^{1/d} \right\} \times \left\{ N - \frac{(d-1)}{2} \right\} \quad (8)$$

où :

C : le niveau de confiance

n : le nombre minimal d'animaux à tester

N : la taille de l'élevage

d : le nombre minimal d'animaux infectés

$$d = N \times PIM$$

La table 2 donne la taille de l'échantillon en fonction de la taille du cheptel et de l'hypothèse sur la prévalence intra-élevage minimale.

3 - PRISE EN COMPTE DE LA QUALITE DU TEST

La fréquence des résultats positifs observés est la **prévalence apparente**. Elle ne doit pas être confondue avec la prévalence réelle, constituée de la fréquence des individus réellement atteints. La prévalence apparente diffère de la prévalence réelle par la différence entre les erreurs par excès (manque de spécificité) et les erreurs par défaut (manque de sensibilité). La formule

suivante permet de corriger le biais dû aux erreurs de classement :

$$P_c = \frac{PA + Sen - 1}{Sen + Spé - 1} \quad (9)$$

où :

P_c : Taux de prévalence corrigée (prévalence réelle)

PA : Taux de prévalence apparente

Sen : Sensibilité du test

Spé : Spécificité du test

Par exemple, pour un test de sensibilité égale à 95 % et de spécificité égale à 99 %, si l'on observe un taux de prévalence apparente de 8 %, le taux de prévalence corrigée serait égal à 3 %. Les imperfections du test utilisé se traduisent par un écart entre la prévalence réelle que l'on cherche à estimer et la prévalence apparente obtenue.

L'utilisation de la formule (9) n'est possible que si l'unité épidémiologique (c'est-à-dire l'unité utilisée dans le calcul de la prévalence) est aussi l'unité d'observation (l'unité au niveau de laquelle le test est appliqué). C'est le cas, par exemple, lorsque l'objectif est d'estimer la prévalence des cheptels infectés en se fondant sur des tests collectifs.

Lorsque les tests sont individuels, les imperfections du test se traduisent également par un biais que l'on pourrait estimer à partir de la formule (9) en remplaçant la sensibilité et la spécificité individuelles par la sensibilité et la spécificité à l'échelle de l'élevage. Cependant, la sensibilité et la spécificité à l'échelle de l'élevage dépendent de la qualité du test à l'échelle individuelle, du nombre d'animaux testés et de la prévalence intra-élevage. Une sensibilité ou une spécificité moyenne, que l'on peut appeler aussi marginale, ne pourra être estimée que si l'on dispose d'informations relatives à la répartition de la taille des cheptels et à celle de la prévalence intra-élevage dans la population étudiée. La correction du biais nécessite alors des calculs très complexes. Donald [2] a proposé une démarche permettant d'approcher ce biais en tenant compte à la fois de la sensibilité et de la spécificité individuelles du test et de la répartition de la maladie. Mais son modèle reste limité au seul cas où le nombre d'animaux testés par élevage est fixe.

Lorsque le test utilisé a une sensibilité inférieure à 100 %, les formules permettant de calculer la taille des échantillons doivent être modifiées. Dans le cas où l'échantillonnage s'effectue à partir d'une grande population ("infinie"), à partir de (6), on aura :

$$1 - C \leq (1 - PIM \times Sen)^n \tag{10}$$

$$n \geq \left\{ 1 - (1 - C)^{1/(d \times Sen)} \right\} \times \left\{ N - \frac{(d \times Sen) - 1}{2} \right\} \tag{12}$$

D'où

$$n \geq \frac{\log(1 - C)}{\log(1 - PIM \times Sen)} \tag{11}$$

Par analogie, on peut de la même façon adapter la formule (8) de Cannon et Roe [1] :

A partir de cette dernière formule (12), il est clair que la taille de l'échantillon requise augmente avec le manque de sensibilité. Des exemples de calcul sont présentés dans le Tableau IV.

Tableau IV : Effet du manque de sensibilité du test sur la taille de l'échantillon requise pour qualifier un cheptel, pour une prévalence donnée, avec un niveau de confiance fixé à 95%

N	P.I.M. = 3 %			P.I.M. = 30 %			N	P.I.M. = 3 %			P.I.M. = 30 %		
	1,00 ¹	0,90	0,80	1,00	0,90	0,80		1,00	0,90	0,80	1,00	0,90	0,80
10	10	10	10	6	6	7	100	63	66	71	8	9	10
20	20	20	20	7	8	8	150	72	78	84	8	9	11
30	29	29	30	7	8	9	200	78	84	92	8	9	11
40	37	37	38	8	9	10	250	81	89	97	8	9	11
50	43	44	46	8	9	10	300	84	92	101	8	9	11
60	48	50	52	8	9	10	350	86	94	104	8	9	11
70	53	55	58	8	9	10	400	87	96	106	8	9	11
80	57	60	63	8	9	10	450	88	97	108	8	9	11
90	60	63	67	8	9	10	500	89	98	109	8	10	11

¹ Sensibilité du test
P.I.M. : Prévalence intra-élevage minimale

B - ESTIMATION DE LA PREVALENCE DES ANIMAUX ATTEINTS

L'objectif de l'étude est d'évaluer la prévalence des animaux concernés par un événement pathologique donné. Dans le cas des animaux de rente, la population est organisée en élevages. De ce fait, le tirage au sort direct des animaux devient pratiquement impossible. Les sondages les plus courants dans ce type d'études sont le sondage en grappes et le sondage à deux degrés. Dans le premier on tire au sort des élevages et on teste tous les animaux appartenant aux élevages retenus, alors que dans le second on teste seulement un échantillon des animaux appartenant aux élevages choisis. Lorsqu'il s'agit d'un sondage à deux degrés on comprend intuitivement qu'en plus des erreurs d'échantillonnage liées au tirage des élevages s'ajoute une autre catégorie d'erreurs, celles liées au tirage des animaux. Un sondage à deux degrés entraîne toujours une perte de précision par rapport aux autres types de sondages.

Dans le cas d'un sondage en grappes, le principe de sondage et de calcul de la taille de l'échantillon est le même que celui développé

dans la partie précédente. On peut utiliser la formule (5) pour déterminer le nombre d'élevages à tirer. La précision d'un tel sondage dépend essentiellement de la variabilité inter-élevages.

Par contre, un sondage à deux degrés est plus difficile à concevoir et demande une bonne connaissance de la théorie des sondages ainsi que la connaissance de la structure de répartition de la maladie entre les élevages et à l'intérieur des élevages. En effet, il n'existe pas toujours de solutions analytiques permettant de dériver des formules simples de calcul du nombre d'élevages et du nombre d'animaux à tirer. Cependant, on peut développer ici un certain nombre de considérations théoriques permettant de discuter la constitution de l'échantillon.

La précision (variance) de la prévalence "animaux" estimée dans un sondage à deux degrés peut être exprimée sous la forme suivante [4] :

$$V(p) = \frac{A_1}{n} + \frac{A_2}{nm} \tag{13}$$

où :
p : la prévalence "animaux"

$V(p)$: variance de p
 A_1 : variabilité inter-élevages
 A_2 : variabilité intra-élevage
 n : nombre moyen d'animaux à tirer
 par élevage
 m : nombre d'élevages à tirer

Le problème est alors de déterminer le nombre d'animaux à tirer par élevage (n) et le nombre d'élevages (m) qui minimisent la variance de l'estimation de la prévalence. Il faut donc disposer d'un certain nombre d'informations sur la répartition de la maladie pour approcher la taille de l'échantillon. Lorsque la maladie a une répartition aléatoire, les élevages auront des situations très semblables vis-à-vis de la maladie. On tâchera alors de diminuer le nombre d'élevages à visiter et à augmenter le nombre d'animaux par élevage. A l'inverse, si la maladie est très contagieuse, elle se concentre alors sur un nombre moins élevé d'élevages et touchera beaucoup plus d'animaux par élevage. Pour rendre le sondage plus efficace, il faut augmenter le nombre d'élevages et diminuer le nombre d'animaux par élevage.

En pratique, le gain de précision n'est pas le seul argument dans la détermination du couple n et m . Le principal avantage d'un sondage à deux degrés est économique. La concentration de plusieurs tests au même endroit permet une économie de temps et de coûts.

Le coût total d'une enquête utilisant un plan de sondage à deux degrés peut être décomposé de la façon suivante :

$$C = C_0 + mC_1 + nC_2$$

où :

C : coût total de l'enquête
 C_0 : frais généraux
 C_1 : coût de déplacement par élevage
 C_2 : coût de prélèvement et d'analyse par animal

Le problème se pose alors de déterminer n et m qui minimisent la variance pour un coût d'enquête donnée (ou l'inverse : minimiser le coût pour une variance donnée).

En combinant (14) et (13) on trouve la solution suivante :

$$m = \frac{C - C_0}{C_1 - nC_2}; \quad n = \sqrt{\frac{C_1}{C_2} \times \frac{A_2}{A_1}}$$

$$V(p) = \frac{1}{C_1 - C_0} \sqrt{C_1 A_1} + \sqrt{C_2 A_2}$$

(15)

Les résultats en (15) suscitent les commentaires suivants :

- n (nombre d'animaux par élevage) ne dépend pas du niveau de précision
- n est d'autant plus faible que la variabilité inter-élevages est plus élevée que la variabilité intra-élevage. En d'autres termes, lorsque l'effet "élevage" est important (A_2 faible) on a tout intérêt à réduire le nombre d'animaux à observer par élevage et à multiplier le nombre d'élevages.
- n est d'autant plus grand que le rapport coût "élevage" sur coût "animal" est élevé.

IV - QUALIFICATION D'UN ELEVAGE OU D'UNE REGION

On abordera dans cette dernière partie le plan d'échantillonnage utilisé dans l'évaluation de la situation sanitaire d'une région.

Les mesures de lutte contre les maladies réglementées s'appuient sur la qualification des cheptels indemnes et l'assainissement des cheptels qui ont été détectés par le dépistage.

Un cheptel indemne se définit comme un cheptel où aucun animal n'est atteint par la maladie étudiée.

On définit un cheptel atteint comme un cheptel comprenant au moins un animal atteint. La proportion des animaux atteints par élevage dépend de la nature de la maladie et des facteurs d'élevage favorisant l'installation, la transmission ou l'émergence de la maladie dans l'élevage.

Les procédures de qualification des élevages s'appuient sur l'application d'un test de dépistage sur les animaux. L'évaluation de la situation sanitaire d'une région pourrait se faire donc en deux étapes : tirage au sort d'élevages puis qualification des élevages à partir des résultats

des tests appliqués sur les animaux. Afin de limiter le coût de cette évaluation, il est fréquent d'échantillonner les individus dans les élevages tirés au sort.

On se situe dans le cas de l'épidémiosurveillance où l'objectif est de mettre en place un système d'alerte. Pour des raisons économiques évidentes, l'évaluation de la situation sanitaire d'une région est à faire par sondage.

A - QUALIFICATION D'UN ELEVAGE

Le nombre d'animaux à étudier par élevage peut être déterminé à partir :

- De la formule (6), lorsque la taille de l'élevage est grande;
- De la formule (7), ou de la table 2, si l'élevage est de petite taille.

Dans tous les cas, la détermination de la taille de l'échantillon ne peut se faire que si l'on dispose d'hypothèses solides sur la dynamique de la maladie dans les élevages. En effet, le calcul du nombre d'animaux à tester par élevage est fondé sur une hypothèse de prévalence intra-élevage minimale (appelée aussi seuil de prévalence). C'est-à-dire que si la maladie s'installe dans un élevage, on suppose que sa prévalence dépassera très rapidement la prévalence intra-élevage minimale.

L'absence d'animaux atteints dans l'échantillon peut être attribuée :

- A l'absence de la maladie dans l'élevage,
- Aux erreurs d'échantillonnage, dont le risque peut être fixé par l'investigateur (par exemple à 5% ou à 1%),
- A une surévaluation de la prévalence intra-élevage minimale par l'investigateur (la prévalence des animaux atteints dans l'élevage est inférieure au seuil de prévalence fixé par l'investigateur).

Lorsque le test appliqué sur les animaux est parfait, l'absence de réagissant sur l'échantillon permet seulement de conclure que la prévalence des animaux est inférieure au seuil de prévalence fixé au risque d'échantillonnage consenti (5% généralement).

Mais, en général, les tests utilisés présentent des imperfections. L'utilisation d'un test de dépistage conduit inévitablement à des erreurs de classement des animaux qui se répercutent sur la qualification de l'élevage.

Afin d'évaluer cette procédure de qualification des élevages, on peut introduire les critères de sensibilité « cheptel » et de spécificité « cheptel ».

1 - SENSIBILITE « CHEPTEL »

La sensibilité « cheptel » est la probabilité d'obtenir au moins un résultat positif sur un échantillon de n animaux provenant d'un cheptel contaminé. Intuitivement, on comprend que cette sensibilité à l'échelle de l'élevage va dépendre de plusieurs facteurs :

DU RESULTAT DU TIRAGE AU SORT :

C'est-à-dire du nombre d'animaux atteints (y) dans l'échantillon dont la probabilité est fonction du nombre d'animaux tirés au sort et de la prévalence intra-élevage. La sensibilité « cheptel » sera identique à la sensibilité individuelle lorsqu'on tire un seul animal atteint parmi les animaux échantillonnés. En revanche, si l'échantillon comporte plusieurs animaux atteints, la probabilité d'observer au moins un animal positif augmente ; par conséquent, la sensibilité « cheptel » augmente et sera plus élevée que la sensibilité individuelle.

La sensibilité « cheptel » en fonction du nombre d'animaux atteints compris dans l'échantillon peut être calculée à l'aide de la formule suivante :

$$SenCh = 1 - (1 - Sen)^y \quad (16)$$

où :

SenCh : sensibilité « cheptel »
Sen : sensibilité individuelle
 y : nombre d'animaux atteints compris dans l'échantillon

Par exemple, en cas d'utilisation d'un test de sensibilité individuelle égale à 90%, la sensibilité « cheptel » devient très bonne dès lors que l'échantillon contient deux animaux atteints (SenCH = 99%).

Ainsi, la sensibilité de la procédure de qualification peut se révéler excellente, même avec une sensibilité individuelle relativement faible.

□ DE LA SENSIBILITE ET DE LA SPECIFICITE INDIVIDUELLES DU TEST :

La sensibilité « cheptel » dépend à la fois de la sensibilité et de la spécificité individuelles du test. Les réponses positives peuvent être aussi bien observées chez des animaux atteints que chez des animaux indemnes.

Le calcul de la sensibilité « cheptel » doit donc tenir compte à la fois de la sensibilité, de la spécificité et du nombre d'animaux atteints dans l'échantillon :

$$SenCh = 1 - (1 - Sen)^y Spé^{n-y} \quad (17)$$

où :

- SenCh : sensibilité « cheptel »
- Sen : sensibilité individuelle
- Spé : spécificité individuelle
- y : nombre d'animaux atteints compris dans l'échantillon
- n : taille de l'échantillon

La possibilité d'observer des réactions par excès sur des animaux provenant d'élevages atteints augmente la sensibilité « cheptel ». Les erreurs par excès compensent alors les erreurs par défaut. La probabilité d'observer des réactions faussement positives augmente quand la prévalence intra-élevage diminue.

□ DE LA STRUCTURE DE REPARTITION DE LA MALADIE :

Les évaluations de la sensibilité « cheptel » que l'on peut faire avec les formules (16) ou (17) ne sont valables que conditionnellement au nombre d'animaux atteints inclus dans l'échantillon. Or, en pratique, les statuts des élevages, la prévalence des animaux atteints et *a fortiori* le résultat du tirage sont inconnus et variables d'un élevage à un autre.

Il faut donc évaluer une sensibilité « cheptel » moyenne (ou marginale) qui dépend des valeurs diagnostiques du test individuel, de la variabilité de la prévalence des animaux atteints entre les élevages atteints et de l'ensemble des possibilités des résultats des tirages au sort des animaux.

La loi de distribution du nombre, y, d'animaux atteints dans un échantillon de taille n issu d'un cheptel de taille N où la prévalence des animaux atteints est égale à p est hypergéométrique. Pour calculer la probabilité d'avoir y animaux atteints dans l'échantillon, il faut connaître la prévalence des animaux atteints et la taille de l'élevage. Ces deux paramètres sont variables d'un élevage à un autre, on calcule alors la probabilité marginale

de y en intégrant sur la distribution de la prévalence intra-élevage.

Pour décrire la variabilité de la prévalence des animaux atteints au sein des élevages, la loi qui s'adapte le plus est la loi bêta. C'est une loi continue, définie entre 0 et 1, qui dépend de deux paramètres réels strictement positifs (π et ρ). Le premier, π , est la moyenne des prévalences intra-élevage dans les cheptels atteints et le second, ρ , est le coefficient de corrélation intra-élevage qui mesure la ressemblance des animaux appartenant au même élevage. Lorsque $\rho=1$ tous les animaux d'un même cheptel ont le même statut, il n'y a dans ce cas que deux types d'élevages : ceux où tous les animaux sont indemnes et ceux où tous les animaux sont atteints. Lorsque $\rho=0$ on est dans le cas où tous les cheptels sont semblables (prévalence intra-élevage constante) et la maladie est répartie aléatoirement dans la population animale indépendamment des élevages. Le paramètre ρ mesure donc la contagiosité de la maladie (s'il s'agit d'une maladie infectieuse), plus ρ est élevé plus la maladie est contagieuse ou encore la variabilité inter-élevages est grande. **Une maladie peut être ainsi décrite par les deux paramètres de la loi bêta .**

En combinant les lois de distribution de p (prévalence intra-élevage) et de y (nombre d'animaux atteints dans l'échantillon) on obtient une loi que l'on peut appeler loi bêta hypergéométrique. La résolution de cette loi pose quelques problèmes analytiques. Mais cette loi se simplifie quand la fraction d'échantillonnage est inférieure à 10% (taille de l'élevage 10 fois plus grande que la taille de l'échantillon). On obtient dans ce cas une loi plus maniable : la loi bêtabinomiale [3].

Le tableau V donne des estimations de la sensibilité « cheptel » en fonction de la sensibilité et de la spécificité du test individuel et des paramètres de distribution de la prévalence intra-élevage. La prise en compte de la répartition de la maladie dans les élevages montre le manque d'efficacité de l'échantillonnage dans la détection des élevages atteints. En effet, lorsque le test est parfait (sensibilité = spécificité = 1), la sensibilité « cheptel » baisse très rapidement avec le paramètre ρ . Par exemple, quand on fixe la moyenne des prévalences intra-élevage à 0,1 et la taille de l'échantillon à 10, la sensibilité « cheptel » est égale à 70 % quand $\rho=0$, à 48 % quand $\rho=0,1$ et à seulement 12 % quand $\rho=0,9$. Cette perte en sensibilité peut être expliquée par

la nature du paramètre ρ . Lorsque ce paramètre augmente on augmente la variabilité de la prévalence intra-élevage autour de sa moyenne π ($=0,10$).

On se retrouve alors avec un nombre plus élevé d'élevages avec une prévalence intra-élevage de plus en plus inférieure à 10 %. Ces élevages à faible prévalence sont plus difficiles à détecter et diminuent la valeur marginale de la sensibilité « cheptel ».

Tableau V : Valeurs de sensibilité « cheptel » en fonction de la sensibilité (Se) et de la spécificité (Sp) du test individuel et des paramètres de distribution de la prévalence intra-élevage (π et ρ). Résultats obtenus par des méthodes de simulation appliquées dans une région où la taille des cheptels est en moyenne de 40 animaux (± 20).

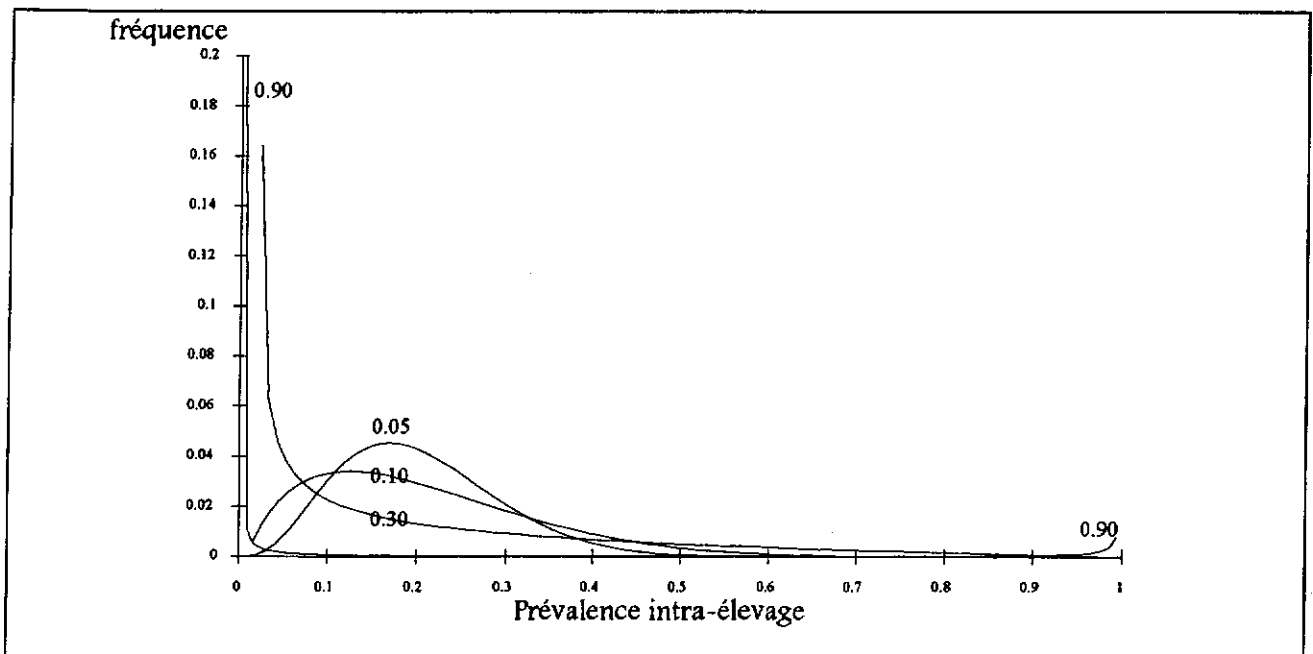
n=10	ρ	Se=Sp=0,9			Se=1 Sp=0,9			Se=0,9 Sp=0,8			Se=Sp=1		
		π			π			π			π		
		0,1	0,2	0,3	0,1	0,2	0,3	0,1	0,2	0,3	0,1	0,2	0,3
0	0,87	0,96	0,99	0,87	0,97	0,99	0,96	0,99	0,99	0,70	0,92	0,98	
0,1	0,81	0,91	0,96	0,82	0,92	0,97	0,94	0,97	0,99	0,48	0,77	0,91	
0,3	0,76	0,84	0,90	0,77	0,85	0,91	0,93	0,95	0,97	0,33	0,57	0,74	
0,5	0,73	0,80	0,85	0,74	0,80	0,86	0,92	0,94	0,95	0,24	0,44	0,60	
0,7	0,71	0,76	0,81	0,71	0,76	0,81	0,91	0,93	0,94	0,18	0,33	0,47	
0,9	0,69	0,73	0,77	0,69	0,73	0,77	0,90	0,92	0,93	0,12	0,24	0,35	

Afin d'illustrer l'effet du paramètre ρ sur la distribution de p entre les élevages, on a tracé sur le même graphique plusieurs lois bêta de même moyenne ($\pi=0,20$) avec des valeurs de ρ différentes (figure 3). On remarque que plus le paramètre ρ augmente plus les valeurs extrêmes de p sont plus fréquentes. Pour $\rho=0,90$ la distribution de la prévalence intra-élevage à la forme en "J" inversé, où la majorité des élevages

ont une prévalence intra-élevage inférieure à 0,10.

Lorsque le test manque de spécificité, on observe paradoxalement une amélioration de la sensibilité « cheptel ». Les cheptels sont déclarés "atteints" grâce aux résultats faussement positifs. Une augmentation de la sensibilité individuelle n'entraîne, par contre, qu'une faible amélioration de la sensibilité cheptel (Tableau V).

Figure 3 : Exemples de lois bêta décrivant la distribution de la prévalence intra-élevage : ($\pi=0,20$ et pour différentes valeurs de ρ : 0,05 ; 0,10 ; 0,30 et 0,90)



2 - SPECIFICITE « CHEPTEL »

La spécificité « cheptel » est la probabilité de n'obtenir aucune réaction positive sur un échantillon de taille n tiré d'un élevage indemne.

La spécificité « cheptel » est plus facile à déterminer que la sensibilité « cheptel ». Elle dépend seulement de la spécificité du test individuel et de la taille de l'échantillon :

$$SpéCh = Spé^n \quad (18)$$

où :

SpéCh : spécificité « cheptel »
Spé : spécificité individuelle
n : taille de l'échantillon

D'après la formule (18), on comprend que plus on teste d'animaux plus le risque d'erreurs par excès est élevé. La spécificité « cheptel » est donc inférieure à la spécificité individuelle, et d'autant plus faible que la taille de l'échantillon est plus élevée.

A titre d'exemple, l'utilisation d'un test de spécificité individuelle de 95 %, sur un échantillon de 10 animaux provenant d'un élevage indemne, donne une spécificité « cheptel » de 59,87 %. Quand la taille de l'échantillon est doublée, la spécificité « cheptel » passe à 35,85 %.

Une augmentation de la spécificité du test individuel entraîne une amélioration significative de la spécificité « cheptel ». Avec une spécificité individuelle de 99 % on arrive à une spécificité de 90,44 % quand la taille de l'échantillon est égale à 10, et à 81,79 % quand l'échantillon est de 20 animaux.

L'équation (18) n'est valable que lorsque les réactions non spécifiques ont une répartition aléatoire dans la population animale, c'est-à-dire qu'il n'existe pas de regroupements particuliers de ces réactions dans certains types d'élevages, auquel cas il faudrait tenir compte de cette répartition dans la détermination de la spécificité « cheptel » comme on l'a fait pour la sensibilité cheptel. Par analogie avec la sensibilité « cheptel », on peut dire que l'on détectera moins bien les réactions non spécifiques lorsque leur répartition n'est pas aléatoire, ce qui correspond à une amélioration de la spécificité « cheptel ». Les évaluations sous l'hypothèse d'une répartition aléatoire sont par conséquent les plus défavorables pour la spécificité « cheptel ».

3 - INTERPRETATION DES RESULTATS

L'interprétation des résultats des procédures de qualification des cheptels ne peut se faire correctement que si l'on connaît avec précision les valeurs de sensibilité « cheptel » et de spécificité « cheptel » en fonction de la nature de la maladie et des paramètres d'échantillonnage.

Les résultats devront être interprétés en fonction du contexte épidémiologique de la région où se trouve l'élevage. On peut utiliser les valeurs prédictives « cheptel » comme aide à la décision sur le statut de l'élevage. Les valeurs prédictives seront calculées de la façon suivante :

$$VPPCh = \frac{Pch \times SenCh}{Pch \times SenCh + (1 - Pch) \times (1 - SpéCh)}$$

$$VPNCh = \frac{(1 - Pch) \times SpéCh}{(1 - Pch) \times SpéCh + Pch \times (1 - SenCh)} \quad (19)$$

où :

VPPCh : valeur prédictive positive « cheptel »
VPNCh : valeur prédictive négative
« cheptel »
Pch : prévalence des cheptels atteints
SpéCh : spécificité « cheptel »
SenCh : sensibilité « cheptel »

L'utilisation et l'interprétation des valeurs prédictives « cheptel » est identique à celle des valeurs prédictives calculées à l'échelle individuelle [7].

B - QUALIFICATION D'UNE REGION

Le principe du calcul du nombre d'élevages à tirer est le même que celui développé dans le cadre de qualification des élevages. On utilisera la formule (6) lorsque la population des élevages dans la région est grande (fraction de sondage inférieure à 10 %) et la formule (7) ou la table 3 lorsque le nombre d'élevages requis dépasse les 10 % de la population. Le seuil de prévalence (P.I.M.) des formules (6) et (7) correspond dans le cas de la qualification d'une région à la prévalence des cheptels atteints minimale.

L'échantillonnage a pour objectif de vérifier si la prévalence des cheptels atteints dans la région étudiée se trouve au-dessous de la prévalence des cheptels atteints minimale. Le choix de cette prévalence minimale est fonction de la nature de la maladie, des moyens disponibles et de la spécificité du test de dépistage à l'échelle de l'élevage.

Lorsque le seuil de la prévalence des cheptels atteints est fixé au dessus de la valeur (1-Spécificité « cheptel »), toutes les zones seront classées atteintes, même les zones où la maladie est réellement absente. La limite (1-Spécificité « cheptel ») n'est autre que la prévalence cheptel apparente minimale.

La "sensibilité" du système d'alerte dépendra alors, paradoxalement, de la spécificité de la

procédure de qualification des cheptels et non pas de sa sensibilité.

Un résumé des principaux objectifs d'échantillonnage en épidémiologie descriptive animale avec indication des sources d'information pour la détermination de la taille de l'échantillon est présenté dans le tableau VI.

Tableau VI : Résumé des principaux objectifs de l'échantillonnage en épidémiologie descriptive animale et source des informations

POPULATION ETUDIEE	OBJECTIF	TAILLE DE L'ECHANTILLON
ELEVAGE	Estimer le taux de prévalence des animaux atteints	Table 1a ou 1b
	Détecter la présence de la maladie au dessus d'un certain niveau	Table 2
	Qualifier un élevage « indemne » : vérifier que le taux de prévalence est inférieur à un niveau défini	Table 2
REGION	Estimer le taux de prévalence des élevages atteints	. Echantillon aléatoire simple : table 1 . Echantillon aléatoire stratifié : formule(4) . Echantillon aléatoire en grappes : formule(5)
	Estimer le taux de prévalence des animaux	. Echantillon en grappes : « consulter un statisticien » . Echantillon à plusieurs degrés : « consulter un statisticien »
	Qualifier une région « indemne » : vérifier que le taux de prévalence des cheptels atteints est inférieur à un niveau défini	Table 3

V - CONCLUSION

En conclusion, la qualité des résultats d'une enquête dépend de la méthode d'échantillonnage et de la taille de l'échantillon. Mais le choix du type d'échantillonnage n'est qu'une étape de l'enquête épidémiologique et il est lié à toutes les autres composantes de l'enquête, à savoir les outils d'observation (tests biologiques, le questionnaire et son mode d'administration), et le traitement statistique des données qui feront l'objet de présentations spécifiques dans ce numéro spécial.

VI - BIBLIOGRAPHIE

- [1] Cannon R.M., Roe R.T.- Livestock disease surveys : A field manual for veterinarians. Australian Bureau of Animal Health, Camberra, 1982, pp. 14-17.
- [2] Donald A.- Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. *Prev. vet. Med.*, 1993, **15** : 167-187.
- [3] Farver T.B., Thomas C., Edson R.K.- An application of sampling theory in animal disease prevalence survey design. *Prev. vet. Med.*, 1985, **3** : 463-473.
- [4] Farver T.B.- Disease prevalence estimation in animal population using two-stage sampling designs. *Prev. vet. Med.*, 1987, **5**, 1-20.
- [5] Sanaa M., Poutrel B., Ménard J.L., Serieys F.- Risk factors associated with contamination of raw milk by *Listeria monocytogenes* in dairy farms. *J. Dairy Sci.*, 1993, **76**, 2891-2898.
- [6] Schwartz D.- Méthodes statistiques à l'usage des médecins et des biologistes, 1969, Flammarion, Paris, 318 p.
- [7] Toma B., Bénet J.J., Dufour B., Eloit M., Moutou F., Sanaa M.- Glossaire d'épidémiologie animale. Maisons-Alfort, Éditions du Point Vétérinaire, 1991, 365 p.

Table 1a : Estimation d'une proportion : taille de l'échantillon nécessaire pour estimer une proportion p avec une erreur absolue Δ et un niveau de confiance de 95 %

N	p																										
	1%		2%		3%		4%		5%		10%		20%		30%		40%		50%								
	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%							
	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
100	94	97	89	98	92	74	94	79	63	95	83	67	98	90	69	94	80	49	90	70	36	86	60	27	80	49	20
150	137	143	126	146	133	98	137	107	79	139	113	87	144	128	89	137	108	59	129	90	41	120	74	30	108	59	21
200	177	188	159	192	170	117	177	130	91	181	140	101	190	163	104	177	132	65	164	106	44	149	84	31	132	65	22
250	215	231	188	237	205	133	214	150	100	220	162	113	234	194	116	216	152	70	196	119	46	175	92	32	152	70	22
300	251	273	215	282	237	145	250	166	107	258	181	122	277	223	126	251	169	73	225	129	48	198	98	33	169	73	23
350	285	314	239	325	267	156	283	180	112	294	199	129	318	250	134	286	184	76	252	137	49	218	103	33	184	76	23
400	317	354	262	368	295	165	315	192	117	329	214	135	359	274	141	318	196	78	277	144	50	237	106	34	196	78	23
450	348	392	282	409	321	173	345	203	121	361	227	140	399	296	146	349	208	80	300	150	50	253	110	34	208	80	23
500	377	429	301	450	346	180	374	213	124	393	239	145	437	317	151	378	218	81	321	155	51	268	112	34	218	81	23
600	431	501	334	529	391	191	427	229	129	452	260	152	512	355	159	432	235	83	360	164	52	294	117	34	235	83	24
700	480	568	363	606	431	200	475	242	133	506	277	158	583	387	166	481	248	85	393	170	52	316	120	35	248	85	24
800	525	633	388	679	467	208	519	253	137	557	291	162	650	416	171	527	260	86	423	175	53	335	123	35	260	86	24
900	566	693	410	750	499	214	559	262	139	603	303	166	715	441	175	568	270	87	449	180	53	352	125	35	270	87	24
1000	604	751	430	818	528	219	596	270	141	646	314	169	776	464	178	606	278	88	473	184	54	366	126	35	278	88	24
1500	756	1002	502	1124	641	236	744	297	148	824	350	179	1047	549	189	759	306	91	561	195	54	417	132	36	306	91	24
2000	864	1202	548	1382	718	246	849	312	152	954	372	185	1267	604	195	869	323	92	619	202	55	448	135	36	323	92	24
2500	946	1366	579	1604	773	252	928	322	154	1055	386	188	1451	643	199	952	333	93	660	206	55	469	137	36	333	93	24
3000	1010	1503	602	1796	815	257	989	329	156	1135	396	191	1606	671	202	1016	341	94	690	209	55	484	138	36	341	94	24
Infinie	1521	3011	753	4470	1118	280	1475	369	164	1824	456	203	3456	864	216	1536	384	96	896	224	56	576	144	36	384	96	24

Exemple :

Pour estimer une proportion de 10% avec une erreur absolue de 1% dans une population de taille 300, il faut tirer au sort au moins 277 individus.

Table 1b : Estimation d'une proportion :
taille de l'échantillon nécessaire pour estimer une proportion p avec une erreur absolue Δ et un niveau de confiance de 99.73 %

N	P																										
	1%		2%		3%		4%		5%		10%		20%		30%		40%		50%								
	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%	Δ%							
100	98	99	100	97	87	98	90	80	98	92	83	99	96	84	98	90	70	96	84	57	94	78	46	90	70	37	
150	144	147	139	148	123	144	128	108	145	132	114	148	140	116	144	129	90	140	117	71	135	104	55	129	90	42	
200	190	195	180	197	186	154	190	163	132	192	169	141	196	183	144	190	164	106	183	145	80	175	126	60	164	106	45
250	234	242	219	245	229	181	234	194	152	237	203	164	243	223	168	234	196	119	224	170	87	211	144	64	196	119	47
300	277	288	257	292	270	206	277	223	169	281	235	184	290	262	189	277	225	129	263	191	92	246	159	67	225	129	48
350	319	334	293	339	309	229	318	250	184	324	264	202	336	299	208	319	252	137	300	210	96	278	172	69	252	137	50
400	360	379	327	386	348	249	359	274	196	366	292	218	382	335	224	360	277	144	336	228	100	309	184	71	277	144	50
450	400	424	359	432	385	267	399	296	208	408	317	232	427	369	239	400	300	150	371	243	103	338	194	72	300	150	51
500	439	467	390	478	420	284	437	317	218	448	341	244	471	401	252	440	322	156	404	257	105	365	202	73	322	156	52
600	514	553	448	568	489	314	512	355	235	527	385	266	559	463	275	515	360	164	467	280	109	416	217	75	360	164	53
700	586	637	502	657	553	339	583	387	248	602	424	283	645	521	295	587	394	171	525	300	112	461	228	76	394	171	53
800	654	719	551	744	613	361	650	416	260	674	458	299	729	574	311	655	424	176	580	317	114	503	238	77	424	176	54
900	719	799	596	829	670	380	715	441	270	744	489	311	810	624	325	720	450	180	630	332	116	540	246	78	450	180	54
1000	781	876	639	913	724	396	776	464	278	811	517	323	891	670	337	783	474	184	678	345	117	575	253	79	474	184	54
1500	1056	1238	811	1313	954	456	1047	549	306	1111	625	361	1266	862	379	1059	563	196	875	389	122	711	276	81	563	196	55
2000	1282	1559	938	1680	1135	494	1267	604	323	1363	697	384	1604	1007	405	1286	621	203	1025	416	124	806	290	82	621	203	56
2500	1470	1846	1035	2019	1280	520	1451	643	333	1578	749	400	1911	1119	422	1476	662	207	1142	434	126	877	298	83	662	207	56
3000	1629	2106	1111	2333	1399	538	1606	671	341	1763	789	411	2190	1209	434	1637	693	210	1236	447	127	932	304	83	693	210	56
Infinitie	3564	7056	1764	10476	2619	655	3456	864	384	4275	1069	475	8100	2025	507	3600	900	225	2100	525	132	1350	338	85	900	225	57

Exemple :

Pour estimer une proportion de 10% avec une erreur absolue de 1% dans une population de taille 300, il faut tirer au sort au moins 290 individus.

Table 2 : Nombre d'animaux à tirer au sort pour détecter la présence de la maladie dans un cheptel en fonction de la taille du cheptel, de la prévalence intra-élevage minimale et du niveau de certitude.

Taille du cheptel	PREVALENCE INTRA-ELEVAGE MINIMALE																	
	1%			5%			10%			20%			30%					
	Certitude			Certitude			Certitude			Certitude			Certitude					
	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%			
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5			
10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10			
15	15	15	14	15	15	12	13	14	8	9	11	5	6	8	7			
20	20	20	18	19	20	13	15	18	8	10	13	6	7	9	8			
25	25	25	21	23	24	15	17	20	8	10	14	6	7	10	9			
30	30	30	23	26	28	16	18	23	9	11	15	6	7	10	10			
35	35	35	25	28	32	16	19	25	9	11	15	6	8	11	11			
40	40	40	27	31	36	17	20	26	9	11	16	6	8	11	11			
45	45	45	28	33	39	17	21	28	9	12	16	6	8	11	11			
50	50	50	30	34	41	18	22	29	9	12	17	6	8	11	11			
55	54	55	31	36	44	18	22	30	9	12	17	6	8	12	12			
60	59	60	32	37	46	18	23	31	10	12	17	6	8	12	12			
65	63	65	32	38	48	19	23	32	10	12	18	6	8	12	12			
70	68	69	33	40	50	19	23	32	10	12	18	6	8	12	12			
75	72	74	34	41	52	19	24	33	10	12	18	6	8	12	12			
80	76	78	34	41	54	19	24	33	10	12	18	6	8	12	12			
85	79	83	35	42	55	19	24	34	10	12	18	6	8	12	12			
90	83	87	35	43	57	19	24	34	10	12	18	6	8	12	12			
95	87	91	36	44	58	20	25	35	10	13	19	6	8	12	12			
100	90	95	36	44	59	20	25	35	10	13	19	6	8	12	12			
125	105	114	38	47	64	20	25	37	10	13	19	6	8	12	12			
150	117	129	39	48	67	20	26	38	10	13	19	6	8	12	12			
175	128	143	40	50	70	21	26	39	10	13	19	6	8	13	13			
200	136	155	40	51	72	21	27	39	10	13	20	6	8	13	13			

Exemple :
Pour détecter la maladie dans un cheptel de 50 animaux où la prévalence minimale est égale à 10%, le nombre d'animaux à tirer au sort est égale 18, 22 ou 29 en fonction du niveau de certitude souhaité (respectivement 90%, 95% et 99%). On choisit en général un niveau de certitude égal à 95%.

Table 3 : Nombre d'élevages à tirer au sort pour détecter la présence de la maladie dans une région en fonction du nombre d'élevages dans la région, du taux de prévalence des élevages atteints et du niveau de certitude. Cette table peut être également utilisée dans le cadre de la qualification d'une région "indemne".

Nombre d'élevages*	Taux de prévalence des cheptels atteints																	
	0,1%		0,5%		1%		0,5%		1%		1%		2%		3%			
	Certitude		Certitude		Certitude		Certitude		Certitude		Certitude		Certitude		Certitude			
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%		
200	200	200	180	190	198	136	155	180	87	105	136	63	78	106				
300	300	300	235	259	286	160	189	235	95	117	159	67	84	119				
500	495	499	500	349	420	184	224	300	102	128	183	70	89	130				
1000	900	950	990	368	450	601	205	258	367	108	138	204	73	94	140			
1500	1177	1296	1430	396	493	687	212	270	395	110	141	211	74	95	144			
2000	1367	1552	1800	410	517	736	216	277	409	111	143	215	74	96	146			
2500	1504	1745	2103	420	532	769	219	281	419	111	144	218	74	96	147			
3000	1607	1894	2353	426	542	791	221	284	425	112	145	219	75	97	147			
3500	1687	2012	2560	430	549	808	222	286	429	112	145	221	75	97	148			
4000	1750	2108	2734	434	555	821	223	287	433	112	146	222	75	97	148			
4500	1802	2187	2882	437	560	831	223	288	436	113	146	222	75	97	149			
5000	1844	2253	3008	439	563	839	224	289	438	113	146	223	75	97	149			
5500	1881	2309	3118	441	566	846	224	290	440	113	146	223	75	97	149			
6000	1911	2357	3214	442	569	852	225	291	441	113	146	224	75	98	149			
6500	1938	2399	3298	444	571	857	225	291	442	113	147	224	75	98	149			
7000	1961	2436	3373	445	573	861	225	292	443	113	147	224	75	98	150			
7500	1982	2469	3440	446	574	865	226	292	444	113	147	225	75	98	150			
8000	2000	2498	3500	446	576	868	226	293	445	113	147	225	75	98	150			
8500	2016	2524	3554	447	577	871	226	293	446	113	147	225	75	98	150			
9000	2031	2547	3603	448	578	873	226	293	447	113	147	225	75	98	150			
9500	2044	2568	3648	448	579	876	226	293	447	113	147	225	75	98	150			
10000	2056	2587	3689	449	580	878	227	294	448	113	147	225	75	98	150			

Exemple :

Pour détecter la maladie dans une région de 2.000 élevages où la prévalence minimale est égale à 0,1%, le nombre d'élevages à tirer au sort est égal à 1.367, 1.552 ou 1.800 en fonction du niveau de certitude souhaité (respectivement 90%, 95% et 99%). On choisit en général un niveau de certitude égal à 95%.

* Nombre d'élevages dans la région