

Evaluation de la qualité des données collectées dans le cadre du **dispositif de déclaration obligatoire des avortements** chez les bovins en France

Mathilde Palussière (1), Didier Calavas (1)*, Anne Bronner (anne.bronner@anses.fr) (1)

(1) Anses, Laboratoire de Lyon, France

* Membre de l'équipe opérationnelle de la Plateforme d'épidémiologie en santé animale (Plateforme ESA)

Résumé

La qualité des données collectées dans le cadre d'un dispositif de surveillance en santé animale influence directement l'évaluation de son fonctionnement et de la situation épidémiologique de la maladie. L'objectif ici est de présenter les résultats d'une évaluation quantitative et qualitative de la qualité des données du dispositif national de déclaration obligatoire des avortements chez les bovins (DA). L'évaluation a porté sur l'analyse de la proportion de DA respectant les spécifications générales d'enregistrement, puis de la qualité des données provenant des DA respectant ces spécifications. Au total, 28 indicateurs ont été élaborés et calculés au niveau national et départemental. Malgré une forte variabilité interdépartementale, les résultats de l'évaluation de la qualité des données de DA sont globalement satisfaisants. Toutefois, dans un souci de pouvoir mieux évaluer le fonctionnement du dispositif actuel, des perspectives concrètes d'évolution sont proposées. Elles devront être discutées avec les différents acteurs notamment au sein de la Plateforme ESA.

Mots clés

Avortements, bovins, qualité des données, brucellose, France

Abstract

Assessment of data quality from the mandatory bovine abortion notification system in France

Data quality from an animal health surveillance system influences the assessment of the correct application of this surveillance system and of the disease epidemiological situation. We used a quantitative and qualitative approach to estimate the data quality from the mandatory bovine abortion notification system. The study focused on the proportion of abortion notifications that met the general specifications of data recording and on the quality of data associated with the abortion notifications that met these specifications. A total of 28 specific indicators were constructed and calculated at the local (departmental) and national level. Whereas indicators vary highly between departments, data quality should be considered as quite good. However, some concrete improvement lines are suggested in order to enhance the quality of assessment of this surveillance system; they are going to be discussed with stakeholders, in the framework of the National epidemiological surveillance platform for animal health.

Keywords

Abortion, bovine, data quality, brucellosis, France

L'analyse des données collectées dans le cadre d'un dispositif de surveillance en santé animale permet d'évaluer son fonctionnement, ainsi que la situation épidémiologique du danger sanitaire surveillé. Toutefois, la qualité de cette analyse dépend fortement de la qualité des données recueillies; l'évaluation de la qualité des données est donc un préalable indispensable à l'analyse et à l'interprétation des résultats de surveillance.

Aucun travail dédié spécifiquement à l'évaluation de la qualité des données collectées dans le cadre d'un dispositif de surveillance en santé animale n'a pu être identifié. Il a donc été choisi de développer une méthodologie d'évaluation de la qualité des données, à partir d'une première étude appliquée au dispositif de déclaration obligatoire des avortements (DA) chez les bovins en France. Ce dispositif a été retenu du fait des travaux d'évaluation déjà engagés sur la surveillance des maladies abortives (Bronner *et al.*, 2013a; Bronner *et al.*, 2013b) ainsi que de l'intérêt des différents acteurs à la faire évoluer. En effet, alors que la surveillance des avortements constitue l'une des modalités principales de la surveillance de la brucellose bovine (Anonyme, 1964, 2005, 2008), seul un quart des éleveurs ayant détecté des avortements les déclareraient (Bronner *et al.*, 2013b). Compte tenu de cette faible sensibilité, différentes voies d'amélioration sont actuellement discutées au sein du groupe de suivi « Surveillance des maladies abortives d'intérêt pour l'État en élevage de ruminants » de la Plateforme nationale de surveillance épidémiologique en santé animale (Plateforme ESA).

L'objectif de cet article est de présenter les résultats de l'évaluation de la qualité des données collectées dans le cadre du dispositif de DA, les axes d'amélioration proposés ainsi que les perspectives ouvertes par ce travail.

Matériel et méthode

L'évaluation a été conduite aux niveaux national et départemental, incluant les 86 départements ayant enregistré au moins une DA entre le 1^{er} août 2011 et le 31 juillet 2012. L'analyse a été faite au niveau départemental plutôt qu'au niveau de chaque acteur (un laboratoire peut par exemple réaliser les analyses pour différents départements) car la maîtrise d'ouvrage est assurée par la DDecPP, qui est donc responsable de la validation des données dans SIGAL. L'analyse a porté sur les données à visée de surveillance (les données de gestion, comme la catégorie fiscale du véhicule du vétérinaire, n'ont pas été étudiées). Dans un premier temps, elle a été réalisée sur toutes les données disponibles pour la période d'étude afin d'évaluer le respect des spécifications générales d'enregistrement (Encadré) en s'appuyant sur la définition de critères qui ont permis d'élaborer des indicateurs (Tableau 1). Dans un second temps, l'analyse s'est focalisée sur la qualité des données associées aux DA respectant ces spécifications générales.

Plusieurs dimensions de la qualité ont été évaluées, de manière qualitative ou quantitative (Tableau 2).

L'évaluation quantitative de la qualité a été réalisée à deux échelles:

- à l'échelle des données: la qualité de chaque donnée à renseigner (ex. « identifiant bovin » ou « stade de gestation ») a été évaluée selon une ou plusieurs dimensions de la qualité (ex. complétude, format, etc.) jugées pertinentes pour cette donnée. Par exemple, la donnée « identifiant bovin » a été évaluée vis-à-vis des dimensions complétude, format et validité. Ce type d'évaluation a abouti à l'estimation d'indicateurs relatifs à des couples [donnée à renseigner – dimension de qualité];
- à l'échelle des DA: la qualité de chaque DA (c'est-à-dire l'ensemble des données à renseigner se rapportant à une même déclaration) a été évaluée en calculant la proportion de DA de « bonne qualité » selon chaque dimension de la qualité retenue. Par exemple, un indicateur

calculé était la proportion de DA complètement renseignées, c'est-à-dire pour lesquelles les différentes données à renseigner étaient effectivement renseignées.

La cohérence entre certaines données a également été étudiée (Tableau 1). D'autres dimensions ont été évaluées qualitativement: l'objectif était d'affiner l'interprétation des résultats quantitatifs et de

Encadré. Modalités de collecte et de circulation des données dans le cadre du dispositif de DA

Une description générale du dispositif de DA a été présentée dans un précédent article (Fediaevsky *et al.*, 2010), mais plusieurs points méritent d'être précisés ici. La circulation des données peut être résumée ainsi:

- le vétérinaire sanitaire (VS) collecte les commémoratifs d'intervention (incluant notamment l'identifiant du bovin, la durée de gestation, la date d'avortement);
- le laboratoire saisit tout ou partie des commémoratifs d'intervention et des résultats d'analyse dans son système d'information (LIMS, *Laboratory information management system*);
- les données saisies par le laboratoire sont exportées dans SIGAL (système d'information du ministère de l'agriculture) sous la forme d'un RAI (résultat d'analyses informatisé);
- ce RAI est utilisé comme base pour la création dans SIGAL d'une intervention, qui regroupe toutes les données disponibles se rapportant à la DA;
- la DDecPP (ou le GDS par délégation de service public) peut saisir dans Sigal certaines informations qui n'auraient pas été saisies ou transmises par le laboratoire.

De plus, la recherche sérologique réalisée par le laboratoire doit se faire via un test ELISA ou une épreuve à l'antigène tamponné (EAT), puis une fixation du complément (FC) si la première analyse est positive (Anonyme, 2010).

Définitions

Qualité d'une donnée: concept multidimensionnel (Berti-Equille *et al.*, 2011; Kerr and Norris, 2004); une donnée est de qualité élevée quand elle est *convenable pour l'usage que l'on souhaite en faire* (« *fitness for use* ») (Kerr *et al.*, 2007; Wang and Strong, 1996).

DA respectant les spécifications générales d'enregistrement: DA pour laquelle une intervention enregistrée dans SIGAL a été associée à un seul bovin ayant avorté à un moment donné, et à un seul RAI complet; la date de visite du vétérinaire devait en outre être antérieure à la date d'exportation des résultats. Ces spécifications générales d'enregistrement ont été définies afin de garantir une analyse fiable des données (et notamment, pouvoir disposer d'une durée de gestation et d'une date d'avortement pour chaque femelle ayant avorté).

mettre en évidence des facteurs influençant la qualité des données, et notamment ceux relatifs aux modes de fonctionnement des acteurs. Des entretiens ont été menés auprès de huit VS, quatre personnes travaillant dans un laboratoire, deux dans un GDS et quatre en DDecPP, localisés dans trois départements (Rhône, Puy-de-Dôme et Haute-Savoie) (Tableau 2). Il leur a été demandé d'attribuer, pour chaque donnée, une note entre 0 et 3 par dimension. Les indicateurs correspondent, pour chaque donnée et par dimension, aux moyennes par type d'acteur et à la moyenne globale.

Résultats

Les résultats de certains indicateurs calculés à l'échelle départementale sont présentés sous forme de box plot, dont les règles de lecture sont présentées en Figure 1.

L'analyse a porté sur 67 261 DA, parmi lesquelles 94 % (n=63 073) respectaient les spécifications générales d'enregistrement. La suite de l'analyse n'a concerné que les DA qui respectaient les spécifications générales d'enregistrement.

Concernant la vitesse de circulation des données, les moyennes nationales des délais entre la date d'avortement et la date de validation des résultats, ainsi qu'entre la date de validation des résultats et la date d'exportation des données étaient respectivement de 11,7 et 2,9 jours, avec une forte variabilité interdépartementale (Figure 2).

Pour l'identifiant du bovin (Figure 3.a), l'indicateur associé à la complétude était compris entre 99,6 et 100 %, celui associé au format présentait une importante variabilité, et celui de la validité était le plus bas des trois, avec une médiane autour de 93 %. Un identifiant bovin valide a été défini comme un identifiant retrouvé dans la Base de données nationale d'identification des bovins (BDNI), qui recense l'ensemble des mouvements de bovins et leurs lieux de détention. La date de visite et l'identifiant de l'exploitation étant considérées comme des données présentant très peu d'erreurs de collecte, le bovin devait par ailleurs être présent dans l'exploitation déclarée le jour de la visite. La démarche étant progressive (la validité n'a été étudiée que sur les données au bon format, elle-même étudiée que sur les données renseignées), il en résulte que l'identifiant du bovin n'était pas valide pour environ un cinquième des DA respectant les spécifications générales d'enregistrement.

Pour la durée de gestation et la date d'avortement, les indicateurs associés à leur complétude étaient d'un niveau assez variable entre les

Tableau 1. Modalités de calcul des indicateurs pour chaque critère et dimension de la qualité retenue

Critères / dimensions		Indicateurs	
DA respectant les spécifications générales d'enregistrement	Un bovin par intervention	Proportion (en %) d'interventions ou d'avortements respectant le critère étudié	
	Une intervention par avortement		
	Un RAI par intervention		
	Date de visite < date de RAI		
Evaluation à l'échelle des données	Complétude de la durée de gestation et de la date d'avortement	Proportion (en %) de DA ou d'analyses pour lesquelles la donnée est présente	
	Complétude de l'interprétation et de l'identifiant du bovin	Proportion (en %) de DA complètes avec la donnée au bon format	
	Format de l'identifiant du bovin, de la date d'avortement et de la durée de gestation		
	Validité du prélèvement, de la matrice analysée, de la méthode et du résultat		Proportion (en %) de DA ou d'analyses complètes et au bon format pour lesquelles la donnée est valide (appartenant à une liste de valeurs)
Evaluation à l'échelle des DA	Complétude	Proportion (en %) de DA avec toutes les données renseignées	
	Validité	Proportion (en %) de DA complètes avec toutes les données valides	
	Vitesse de circulation des données	Délai d'obtention des résultats	Délai moyen entre la date d'avortement et la date de validation des résultats (en jours)
		Disponibilité des données	Délai moyen entre la date de validation des résultats et la date d'envoi du RAI (en jours)
Cohérence	Date d'avortement - date de visite Nature du prélèvement - matrice analysée Matrice analysée - analyte - méthode Analyse - motif de non analyse Méthode - « confirmation » Entre les différentes analyses Résultats d'un RAI - interprétation	Proportion (en %) de DA avec les données étudiées cohérentes	

Tableau 2. Définition retenue pour chaque dimension de la qualité étudiée

	Dimension	Définition
Evaluation quantitative	Vitesse de circulation des données	Délai entre la date de l'événement ayant généré la donnée et la date à partir de laquelle elle est disponible
	Complétude	Absence de valeur nulle (Berti-Equille <i>et al.</i> , 2011)
	Format	Absence d'erreur de syntaxe (orthographe ou format), pour les données renseignées
	Validité	Concordance entre la donnée collectée et la réalité, pour les données au bon format (Berti-Equille <i>et al.</i> , 2011)
	Cohérence	Cohérence de plusieurs données valides entre elles
Evaluation qualitative	Pertinence	Intérêt de la donnée collectée (Pipino <i>et al.</i> , 2002)
	Crédibilité	Niveau de confiance dans la validité de la donnée collectée (Pipino <i>et al.</i> , 2002)
	Objectivité	Niveau d'impartialité perçu dans la collecte de la donnée (Pipino <i>et al.</i> , 2002)

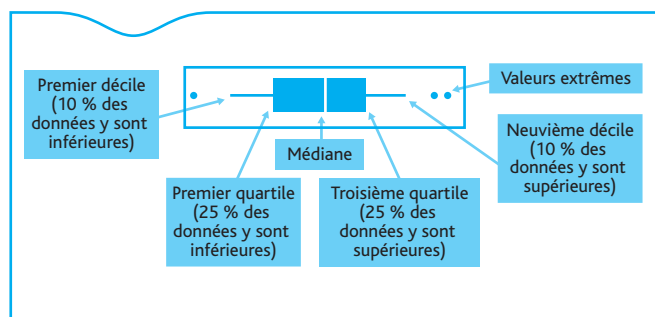


Figure 1. Modalités d'interprétation d'un « box plot »

départements, contrairement à ceux associés à leur format (Figure 3.b et 3.c). L'indicateur associé à la validité de la durée de gestation (une donnée valide correspondant à une durée de gestation comprise entre 1 et 10 mois) était également élevé et peu variable.

Le prélèvement, la matrice, la méthode d'analyse et les résultats ont été considérés valides (c'est-à-dire appartenant à une liste de valeurs possibles) dans plus de 96,9 % des analyses au niveau national. Pour ces trois dernières données, quelques départements se démarquaient fortement (départements pour lesquels les indicateurs de la validité étaient compris entre 50 et 80 %).

À l'échelle des DA, 83,6 % des DA étaient complètement renseignées, et parmi ces DA, 80,7 % étaient complètement valides. Au total, 67,4 % (=83,6%*80,7%) des DA étaient donc complètement renseignées et valides.

La majorité des indicateurs associés à la cohérence entre des données (Tableau 1) étaient globalement très bons (moyennes nationales situées entre 96,3 % et 99,7 %) et peu variables entre départements, à l'exception de l'indicateur associé à la cohérence entre les différentes analyses réalisées par le laboratoire. Cet indicateur a montré une variabilité très élevée, avec 48 départements en dessous de 5 % et 32 au-dessus de 95 % ; au niveau national, des analyses supplémentaires étaient réalisées pour 80 % des avortements (par exemple, la FC était systématiquement réalisée en complément de l'EAT ou de l'ELISA).

Enfin, concernant l'évaluation qualitative, les personnes interrogées ont jugé la collecte de la durée de gestation et de la date d'avortement peu pertinente, et la qualité de ces données peu crédible et objective : en effet, la collecte de ces données est demandée (et réalisée) alors que souvent elles ne peuvent pas être connues de manière précise. Les données relatives aux analyses de laboratoire ont été jugées très crédibles et objectives par tous, mais leur collecte parfois peu pertinente, du fait de leur absence d'utilisation.

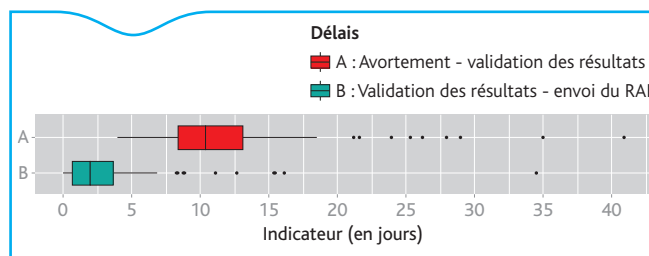
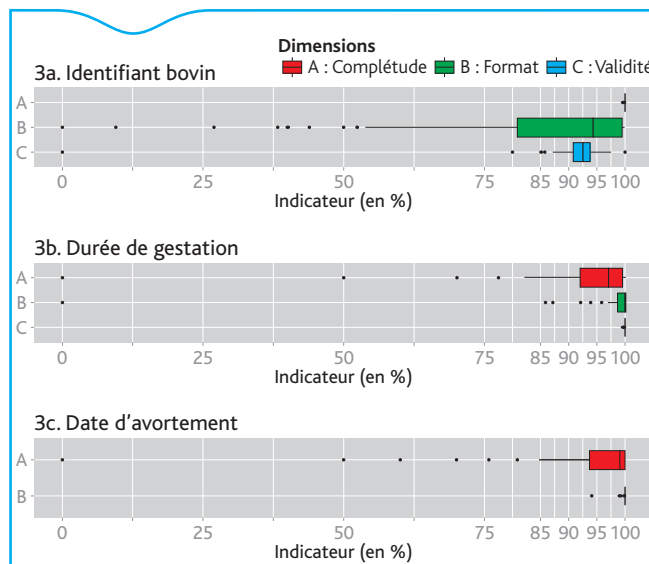


Figure 2. Distribution des indicateurs témoignant de la vitesse de circulation des données calculés à l'échelle départementale (données tronquées à 43 jours)



Figures 3. Distribution des indicateurs témoignant de la qualité des données pour l'identifiant bovin (3a), la durée de gestation (3b), la date d'avortement (3c), calculés à l'échelle départementale

Discussion

L'évaluation de la qualité des données du dispositif de DA présente quelques limites, en particulier en ce qui concerne l'étude de la validité des données. Celle-ci avait comme objectif d'évaluer la concordance entre les données et la réalité. Cependant, hormis l'identifiant du bovin pour lequel un contrôle était possible (en utilisant comme base de référence la BDNI, même si des erreurs de notification existent), cette validité n'a pu être étudiée que de façon incomplète, en s'assurant que la donnée appartenait bien à un ensemble de valeurs considérées comme « valides » et en recueillant la perception qu'en avaient les acteurs (au travers de la dimension « crédibilité »). De plus, elle n'a pas pu être étudiée pour certaines données, comme la date d'avortement, alors même que les entretiens avec les acteurs de terrain ont révélé que cette date pouvait être saisie par défaut (lorsqu'elle était non renseignée par le vétérinaire) comme identique à la date de visite. Le délai moyen d'obtention des résultats est donc réduit, et l'indicateur relatif au délai moyen d'obtention des résultats surestimé. En outre, malgré le souhait de limiter le nombre d'indicateurs, ceux-ci restent nombreux ; il serait ainsi intéressant d'élaborer des indicateurs synthétiques, permettant d'avoir une vision plus globale de la qualité des données, à l'échelle nationale et départementale. Ceux-ci pourraient être calculés comme le rapport entre le nombre d'indicateurs pour lesquels une valeur cible (préalablement fixée) serait atteinte ou dépassée et le nombre d'indicateurs mesurés. Les valeurs cibles restent à définir avec les acteurs du dispositif.

Malgré ces limites, cette évaluation permet de disposer d'un premier état des lieux de la qualité des données du dispositif de DA. Les résultats de cette évaluation sont globalement satisfaisants et il y a peu de départements pour lesquels la qualité des données peut être considérée comme insuffisante. Toutefois, une marge de progrès existe : à l'échelle nationale, deux tiers des DA (sur 63 073 DA) étaient entièrement complètes et valides. Les différents défauts de qualité des

données mis en évidence ne sont pas préjudiciables à la capacité de détecter précocement un foyer de brucellose lorsqu'un avortement est déclaré. Toutefois, ils impactent la qualité de l'évaluation du dispositif et la connaissance de la population surveillée: l'identifiant bovin permet de s'assurer que l'avortement a bien eu lieu, la durée de gestation de connaître les types d'avortements déclarés, par exemple. En outre, un délai d'obtention des résultats dans SIGAL élevé (ce délai moyen étant supérieur à 15 jours dans un cinquième des départements) limite certainement les possibilités de correction à apporter aux données mal ou non renseignées. Enfin, certains résultats de cette étude permettent de mettre en évidence certains dysfonctionnements dans la mise en œuvre du dispositif, tels que ceux relatifs à la non-conformité des résultats d'analyse (des analyses excédentaires étant fréquemment réalisées).

Cette étude a permis d'identifier certains facteurs influençant la qualité des données et ainsi certains axes d'amélioration, regroupés en trois catégories: le matériel, la méthode et les acteurs. Les facteurs relatifs au matériel portent sur la feuille de commémoratifs utilisée (avec parfois des informations demandées mais non utilisées), qui mériterait d'être harmonisée au niveau national, et sur le fait que SIGAL a été conçu comme un outil informatique imposant peu de contraintes de saisie (relatives aux relations entre données, et au renseignement et format de certaines données). En outre, la forte disparité interdépartementale qui a été observée peut s'expliquer pour partie par la diversité des LIMS des laboratoires et de leur paramétrage. En effet, les différents LIMS utilisés en France sont nombreux (plus de 25 prestataires recensés) et une partie de leur programmation relève des laboratoires eux-mêmes. Ainsi on peut considérer que chaque laboratoire a son propre LIMS.

Concernant la méthode, les modalités de collecte, de saisie et de validation des données sont précisées pour partie dans un ordre de méthode national, mais reposent fortement sur le réseau d'acteurs (vétérinaires sanitaires, laboratoires agréés, coordinateurs SIGAL régionaux (CoSiR), agents des DDecPP et des GDS), et mériteraient d'être rappelées dans des procédures nationales.

Enfin, même si certains résultats attestent du souci qu'ont les acteurs à renseigner à leur niveau des données de « bonne qualité », l'harmonisation nationale de la qualité des données pourrait être améliorée par une meilleure valorisation des données de DA et une plus forte sensibilisation de l'ensemble des acteurs. Plus globalement, il s'agirait ainsi de renforcer l'animation du dispositif sur la qualité des données (incluant une sensibilisation des acteurs, un suivi de la qualité des données et un retour d'information).

Conclusion

L'évaluation de la qualité des données issues du dispositif de DA a montré une qualité très satisfaisante pour un grand nombre de points. Toutefois, compte tenu de certains résultats et de l'impact direct de la qualité des données sur l'évaluation du fonctionnement du dispositif, il semble important d'intégrer dans les réflexions menées actuellement au sein de la Plateforme ESA des modalités d'amélioration de cette qualité. Au préalable, il s'agirait de préciser les objectifs attendus en terme de qualité optimale des données en tenant compte des différentes contraintes, notamment informatiques et de terrain (temps et ressources humaines disponibles). Les résultats de ce travail seront utilisés pour élaborer un guide générique d'évaluation de la qualité des données, qui pourrait être utilisé par les partenaires de la Plateforme ESA pour d'autres dispositifs de surveillance en santé animale.

Remerciements

Cette étude a été réalisée dans le cadre d'un stage de première année à l'École nationale des services vétérinaires.

Les auteurs tiennent à remercier particulièrement les agents des GDS, des DDecPP, des laboratoires et les vétérinaires sanitaires rencontrés dans le Rhône, le Puy-de-Dôme et la Haute-Savoie, André Gauffier pour son décryptage du système d'information SIGAL et Anne Van de Wiele pour sa connaissance sur les échanges informatiques SIGAL-laboratoires.

Références bibliographiques

Anonyme 1964. Directive du Conseil du 26 Juin 1964 relative à des problèmes de police sanitaire en matière d'échanges intra-communautaires d'animaux des espèces bovine et porcine.

Anonyme 2005. Décision de la Commission du 28 octobre 2005.

Anonyme 2008. Arrêté du 22 avril 2008 fixant les mesures techniques et administratives relatives à la prophylaxie collective et à la police sanitaire de la brucellose des bovinés.

Anonyme 2010. Note de service du 31 août 2010 relative à la brucellose des bovinés: application de l'arrêté du 22 avril 2008 révisé.

Berti-Equille, L., Comyn-Wattiau, I., Cosquer, M., Kedad, Z., Nugier, S., Peralta, V. Si-Said Cherfi, S. Thion V., 2011. Assessment and analysis of information quality: a multidimensional model and case studies. *IJIQ 2*, 300-323.

Bronner, A., Hénaux, V., Fortané, N., Calavas, D., 2013a. Identification des facteurs influençant la déclaration des avortements chez les bovins par les éleveurs et les vétérinaires. *Bull. Epid. Santé Anim. Alim.* 57, 5-8.

Bronner, A., Henaux, V., Vergne, T., Vinard, J.L., Mornat, E., Hendrikx, P., Calavas, D., Gay, E., 2013b. Assessing the mandatory bovine abortion notification system in France using unilist capture-recapture approach. *PLoS one* 8, e63246.

Fediaevsky, A., Garin-Bastuji, B., Moutou, F., 2010. Bilan de la surveillance de la brucellose bovine en 2009: des contraintes de surveillance dans une situation assainie. *Bull. Epid. Santé Anim. Alim.* 40, 9-12.

Kerr, K., Norris, T., 2004. The development of a healthcare data quality framework and strategy. *IQ2004*.

Kerr, K., Norris, T., Stockdale, R. 2007. Data quality information and decision making: a healthcare case study. In: 18th Australasian Conference on Information Systems, Toowoomba, 5-7.

Pipino, L.L., Lee, Y.W., Wang, R.Y., 2002. Data quality assessment. *Communications of the ACM* 45, 211-218.

Wang, R.Y., Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers? *J. Manag. Inform. Syst.*, 5-33.