WEIGHING RISK FACTORS ASSOCIATED WITH BEE COLONY COLLAPSE DISORDER BY CLASSIFICATION AND REGRESSION TREE ANALYSIS

Dennis vanEngelsdorp[a,b], Niko Speybroeck[c,d], Jay D. Evans[e], Bach Kim Nguyen[f], Chris Mullin[b], Maryann Frazier[b], Jim Frazier[b], Diana Cox-Foster[b], Yanping Chen[e], David R. Tarpy[g], Eric Haubruge[f], Jeffrey S. Pettis[e], and Claude Saegerman[h*]

[a] Bureau of Plant Industry, Pennsylvania Department of Agriculture, 2301 North Cameron Street. Harrisburg PA 17110, TEL: 717-772-5225; FAX:717-705-6518 EMAIL:dennis.vanengelsdorp@gmail.com

[b] Department of Entomology, The Pennsylvania State University, University Park, PA 16802; TEL: 814-865-1895; FAX: 814-865-3048; EMAIL: CAMullin@psu.edu, mfrazier@psu.edu, jfrazier@psu.edu, dxc12@psu.edu, rmul@psu.edu

[c] Institute of Tropical Medicine, Nationalestraat 155, 2000 Antwerp, Belgium; EMAIL: nspeybroeck@itg.be

[d] Institute of Health and Society, Université Catholique de Louvain, Clos Chapelle-aux-Champs, 1200 Bruxelles, Belgium

[e] USDA – ARS Bee Research Laboratory, Bldg. 476 BARC-E, Beltsville, MD 20705 USA EMAIL: Jay.Evans@ars.usda.gov, Judy.chen@ars.usda.gov, Jeffery.Pettis@ars.usda.gov

[f] Department of Functional and Evolutionary Entomology, Agro-Bio Tech, University of Liege -5030 Gembloux, Belgium, TEL: +32-81-62-2287; FAX: +32-81-62-2312; EMAIL: E.Haubruge@ulg.ac.be, nguyen.b@fsagx.ac.be

[g] Department of Entomology, Campus Box 7613, North Carolina State University, Raleigh, NC 27695-7613; TEL: (919) 515-1660; FAX: (919) 515-7746; EMAIL: david_tarpy@ncsu.edu

[h] Department of Infectious and Parasitic Diseases, Epidemiology and Risk analysis applied to the Veterinary Sciences, Faculty of Veterinary Medicine, University of Liege, B-4000 Liege, Belgium, TEL +32-4-366-4579; FAX +32-4-366-4261; EMAIL: Claude.Saegerman@ulg.ac.be

[*] Corresponding author: Department of Infectious and Parasitic Diseases, Epidemiology and Risk analysis applied to the Veterinary Sciences, Faculty of Veterinary Medicine, University of Liege, B-4000 Liege, Belgium, TEL +32-4-366-4579; FAX +32-4-366-4261; EMAIL: Claude.Saegerman@ulg.ac.be

ABSTRACT

Colony Collapse Disorder (CCD), a syndrome whose defining trait is the rapid loss of adult worker honey bees, is thought to be responsible for a minority of the large over-wintering losses experienced by U.S. beekeepers since the winter of 2006-2007. Using the same data set developed to perform a mono-factorial analysis (vanEngelsdorp et al. 2009), we conducted a classification and regression tree (CART) analysis in an attempt to better understand the relative importance and inter-relations among different risk variables in explaining CCD. Fifty-five exploratory variables were used to construct two CART models: one with and one without a cost of misclassifying a CCD-diagnosed colony as a non-CCD colony. The resulting model tree which permitted for misclassification had a sensitivity and specificity of 85% and 59% respectively. While factors measuring colony stress (e.g., adult bee physiological measures such as fluctuating asymmetry or mass of head, and morphological measures such as frames of brood) were important discriminating values, 6 of the 19 variables having the greatest discriminatory value were pesticide levels in different hive matrices. Notably, coumaphos levels in brood (a miticide commonly used by beekeepers) had the highest discriminatory value and were highest in control (healthy) colonies. Our CART analysis provides evidence that CCD is likely the result of several factors acting in concert, making afflicted colonies more susceptible to disease. This analysis highlights several areas that warrant further attention, including the effect of sub-lethal pesticide exposure on pathogen prevalence and the role of variability in bee tolerance to pesticides on colony survivorship.

Keywords: Colony collapse disorder, Epidemiology, Classification and Regression Tree analysis, Pathogens, Apiculture, *Apis mellifera*.

INTRODUCTION

Large-scale losses of managed honey bees (*Apis mellifera* L.) have been reported globally (Haubruge et al. 2006, vanEngelsdorp and Meixner 2010). In the United States, a portion of the dead and dying colonies were characterized by a common set of specific symptoms: (i) the rapid loss of adult worker bees from affected beehives, resulting in weak or dead colonies with excess brood present relative to adult bees; (ii) a noticeable lack of dead worker bees both within and surrounding the hive; and (iii) the delayed invasion of hive pests (e.g., small hive beetles and wax moths) and kleptoparasitism from neighbouring honey bee colonies (Cox-Foster et al. 2007). Subsequently, this syndrome has been termed Colony Collapse Disorder, or CCD, and its case definition has been revised to include (iv) the absence of varroa and nosema loads at levels thought to cause economic damage (vanEngelsdorp et al. 2009).

In an attempt to better characterize CCD, an initial descriptive epizootiological study was conducted (vanEngelsdorp et al. 2009). This mono-factorial study focused on identifying and quantifying direct and indirect measures of risk in affected populations and comparing these measures with apparently healthy populations. Some measures of risk differed between apparently healthy and unhealthy populations, although no one factor clearly separated the two groups. Generally, CCD-affected colonies had higher pathogen incidence and pathogen loads, but no pathogen on its own was found in all CCD colonies. This finding suggests that some underlying risk factor or combination of risk factors compromises the immunity of bees and thus decreases a colony’s ability to fight pathogenic infection (vanEngelsdorp et al. 2009). A recent effort found broad changes in gene expression between bees from healthy and collapsed colonies, along with elevated pathogen levels in CCD colonies, but no systematic differences in RNA transcripts for genes implicated in honey bee immunity (Johnson et al. 2009b).

A classification and regression tree (CART) analysis is a useful non-parametric data-mining technique. This analysis is particularly helpful when attempting to investigate which direct and indirect measures of risk are predictive of a newly emerging or complex disease (Saegerman et al. 2004). Contrary to classical regression (which uses linear combinations), CART does not require the data to be linear or additive. Furthermore, CART analysis does not require possible interactions between factors to be pre-specified (Breiman et al. 1984). In essence, the classification trees resulting from a CART analysis accommodate more flexible relationships among variables, missing covariate values, multi-colinearity, and outliers in an intuitive manner (Speybroeck et al. 2004). When values for some predictive factors are missing, they can be estimated using other predictor ("surrogate") variables, permitting the use of incomplete data sets when generating regression trees. Another advantage of a CART analysis (as compared to a classical multivariate regression analysis) is that it allows for the calculation of the overall discriminatory power, or relative importance, of each explanatory variable.

The monofactorial study by vanEngelsdorp and colleagues (2009) investigated more than 200 variables, but only 61 occurred with enough frequency to make meaningful comparisons between diseased (CCD) and apparently healthy populations. Included in this list of variables were 6 that were directly linked with either the operational or refined definition of CCD: frames of bees, ratio of bees to brood, presence of varroa mites (*Varroa destructor*), spore loads and presence of *Nosema* ceranae, *Nosema apis*, or both (see case definition discussion above). While the inclusion of these variables either validated the application of the operational case definition (or justified the revision of the original case definition of CCD), the use of these "case defining" variables in a multi-factorial analysis could skew results as these variables are inherently not independent. In the current study, we preformed a CART analysis to help identify those variables that, independently or in combination, best discriminate CCD from non-CCD populations. However, to avoid creating a circular argument, we included only truly independent variables (n=55) and discarded those (n=6) that were intrinsic to CCD's case definition. This study is the first to apply a CART analysis to honey bee pathology in an attempt to advance the understanding of the underlying causes of CCD.

MATERIALS AND METHODS

Study apiaries and colonies

As outlined in vanEngelsdorp et al. (2009), 91 colonies from 13 apiaries resident in either Florida or California during January and February 2007 had adult bees, brood, wax, and/or beebread (pollen provisions) sampled for further analysis.

Case definition

Select colonies were classified in the field as either (i) not having CCD symptoms (39 'control' colonies) or (ii) having CCD symptoms (52 'CCD' colonies). Colonies were considered to have CCD symptoms when adult bee populations were in obvious rapid decline leaving brood poorly attended, or were dead in an apiary having clear symptoms of CCD. In those CCD colonies where bees remained, there were insufficient number of bees to cover the brood, the remaining worker bees appeared young (i.e., adults bees that were unable to fly), and the queen was present. Notably, both dead and weak colonies in CCD apiaries were not being robbed by other bees despite the lack of bloom in the area, neither were they being attacked by secondary pests despite the presence of honey and beebread in the vacated equipment (vanEngelsdorp et al. 2009).

Explanatory variables

After elimination of six variables inherently linked to defining CCD colonies (vanEngelsdorp et al., 2009, and above), the remaining variables were either indirect measures of colony stress (e.g., adult bee physiological and morphological measures) or direct measures of risk that are thought to directly and adversely affect colony health (e.g., parasite, pathogen, and pesticide loads).

Classification and regression tree analysis

A CART (Classification and regression tree) analysis was conducted on the data set, where colony status (CCD or Control) was used as the dependent variable and the 55 direct/indirect measures of risk were used as independent or predictor variables. A CART analysis is a non-linear and non-parametric model that is fitted by binary recursive partitioning of multidimensional covariate space. Using CART 6.0 software (Salford Systems, San Diego, CA, USA), the analysis successively splits the dataset into increasingly homogeneous subsets until it is stratified meet specified criteria (Saegerman et al. 2004, Thang et al. 2008). The Gini index was used as the splitting method, and 10-fold cross-validation was used to test the predictive capacity of the obtained trees. CART performs cross validation by growing maximal trees on subsets of data then calculating error rates based on unused portions of the data set. To accomplish this, CART divides the data set into 10 randomly selected and roughly equal parts, with each "part" containing a similar distribution of data from the populations of interest (i.e., CCD vs. Control). CART then uses the first 9 parts of the data, constructs the largest possible tree, and uses the remaining 1/10 of the data to obtain initial estimates of the error rate of the selected sub-tree. The process is repeated using different combinations of the remaining 9 sub-sets of data and a different 1/10 data sub-set to test the resulting tree. This process is repeated until each 1/10 sub-set of the data has been used as to test a tree that was grown using a 9/10 data sub set. The results of the 10 mini-tests are then combined to calculate error rates for trees of each possible size; these error rates are applied to prune the tree grown using the entire data set.

The consequence of this complex process is a set of fairly reliable estimates of the independent predictive accuracy of the tree, even when some of the data for independent variables are incomplete and/or specific events are either rare or overwhelmingly frequent.

For each node in a CART generated tree, the "primary splitter" is the variable that best splits the node, maximizing the purity of the resulting nodes. When the primary splitting variable is missing for an individual observation, that observation is not discarded but, instead, a surrogate splitting variable is sought. A surrogate splitter is a variable which pattern within the dataset, relative to the outcome variable, is similar to the primary splitter. Thus, the program uses the best *available* information in the face of missing values. In datasets of reasonable quality, this allows all observations to be used. This is a significant advantage of this methodology over more traditional multivariate regression modelling, in which observations which are missing *any* of the predictor variables are often discarded.

In this study, two classification and regression tree models were constructed: one without and one with a cost of misclassifying a CCD diagnosed (positive) colony as an apparently healthy (negative) colony. For the second tree, several possibilities were tested, but the tree generated allowing for a misclassification cost of 2 resulted in the smallest number of misclassified colonies while minimizing the size (complexity) of the resulting tree (cf. Suman et al. 2010 for details). The cost (penalty) is a measure of the likelihood of misclassifying a CCD-diagnosed (positive) colony as an apparently healthy (negative) colony. This classification enabled us to make a distinction between groups of colonies containing at least one colony with CCD from groups of colonies without any CCD-diagnosed colonies. The discriminatory power of each variable included in the analysis was also calculated.

RESULTS

Classification and regression trees analysis without a misclassification cost

The CART analysis without a misclassification cost showed that coumaphos load in brood (p: 100.00) and the fluctuating asymmetry (p: 50.15) were the two predictor variables with the strongest overall discriminating power (Table 1 and Figure 1). Generally, CCD colonies had lower levels of coumaphos in brood and their adult bees were more symmetrical when compared to samples taken from apparently healthy colonies. As indicated by having a discriminatory power of more than 15% , three additional variables—that is, variables that did not act as nodes on the Regression tree (Figure 1)—also had significant discriminating power: loads of esfenvalerate (p: 33.91), coumaphos (p: 29.42), and iprodione (p: 17.65) in the wax (Table 1). Overall, the resulting tree (Figure 1) had a sensitivity of 65% and a specificity of 87%.

Classification and regression trees analysis with a cost of misclassification

When conducting the CART analysis with a misclassification cost of 2, at least five variables distinguished themselves as most important: coumaphos in brood (p: 100.00), coumaphos in beebread (p: 81.11), fluctuating asymmetry (p: 42.48), mass of the head (p: 36.07), coumaphos in wax (p: 27.39), and proteins in the thorax (p: 12.71; Table 2). Some of these variables did not act as splitting nodes in the regression tree (Figure 2). As with the first model, the tree permitting misclassification first segregated the study population based on coumaphos loads in bee brood. A majority of healthy colonies had coumaphos loads in bee brood > 66 ppb. Both of the resulting branches were further split by three other variables (Figure 2) and resulted in five terminal nodes, including one node that contained only CCD colonies. Generally, this model revealed that when compared to CCD colonies, control colonies are best characterized as having higher levels of coumaphos in brood, the adult bees were more asymmetrical, and had heads with a greater mass. This entire tree had a sensitivity of 85% and a specificity of 59%.

4. DISCUSSION

In the United States, overwintering losses of honey bee colonies have averaged around 30% or more over the winters 2006/2007, 2007/2008, and 2008/2009 (vanEngelsdorp et al. 2007, vanEngelsdorp et al. 2008, vanEngelsdorp et al. 2010). While most operations identify known threats as the cause of mortality (e.g., poor queens, colony starvation, and varroa mite parasitism), some of these losses shared symptoms associated with CCD (specifically, no dead bees in affected colonies). Previous attempts to find the cause of CCD failed to identify a single factor that explained all cases of CCD (Cox-Foster et al. 2007, Johnson et al. 2009b, vanEngelsdorp et al. 2009). In an attempt to better characterize CCD following an initial descriptive (and monofactorial) study, we present here the results of a multifactorial CART analysis.

The use of CART analysis in epidemiological studies permits the identification of risk factors that are useful in disease diagnosis (Saegerman et al. 2004) as well as those that may play an important role in disease occurrence (Thang et al. 2008). CART analysis is a valuable tool in epidemiological studies because it generates a non-linear and non-parametric model. In addition, this approach is particularly useful when, as in this case, the dataset includes missing values, because the CART model generates surrogate data points based on relationships identified within the existing data (Saegerman et al. 2004, Thang et al. 2008).

Among 55 variables used in our CART analysis, one variable stood out as the most important when differentiating CCD from control colonies: coumaphos levels in brood. In both the tree with and without a misclassification cost, colonies from control colonies had the highest level of coumaphos in brood.

The presence of some pesticide products found in hives is not surprising (Bogdanov et al. 1998, Tremolada et al. 2004, Martel et al. 2007). Coumaphos is the active ingredient found in varroa mite control products widely used by U.S. beekeepers. This lipophilic product is known to accumulate in wax. It is therefore not surprising that this product is found extensively in beekeeping operations both in the U.S. and Europe (Mullin et al. 2010, vanEngelsdorp and Meixner 2010). Even one treatment of the organophosphorus miticide coumaphos, marketed as CheckMite+^{TM} (Bayer), can elevate coumaphos levels in brood-chamber honey stores to 60 and 111 ppb (Karazafiris et al. 2008). The discriminatory value of coumaphos in brood suggests that healthy colonies had mite populations that were more aggressively or persistently controlled by the beekeepers. While varroa mite levels were not different between CCD and control populations at the time of sampling (vanEngelsdorp et al. 2009), it is possible that mite populations differed at some time prior to sample collection. CCD may therefore be a consequence of elevated levels of mites—relative to mite levels in control colonies—some time prior to sampling. Clearly, longitudinal studies that monitor the mite levels prior to the onset of CCD are needed to quantify the effect of mite levels prior to colony collapse.

Coumaphos was initially selected as a mite control agent because of its relative low toxicity to honey bees. Despite this low toxicity, chronic sub-lethal exposure to this product can have detrimental effects on colony health (Pettis et al. 2004). Furthermore, the low toxicity of this product also relies, at least in part, on the rapid detoxification of these miticides by the exposed bees (Johnson et al. 2009a). Honey bees, as compared to other insects, have relatively few insecticide detoxifying genes (Claudianos et al. 2006), which may in part explain why honey bees are relatively sensitive to pesticide exposure (Atkins 1992). One gene family in particular, cytochrome P450 mono-oxygenase enzymes (P450), is used by honey bees to detoxify coumaphos (Johnson et al. 2006, Johnson et al. 2009a). As a result, exposure to both products (e.g., coumaphos and fluvalinate) simultaneously has a synergistic effect on toxicity towards bees (Johnson et al. 2009a). While unproven, it does stand to reason that certain populations of honey bees can vary in their tolerance of pesticide exposure as a result of differences in the expression of detoxifying genes. Should this be the case, differences in pesticide resistance could explain the relative importance of some pesticide loads in distinguishing CCD populations from control populations. In the mono-factorial analysis, coumaphos and esfenvalerate in wax were consistently found at higher concentrations in the control colonies (vanEngelsdorp et al. 2009). Pathogenic attack, specifically viral attack, may arrest translation of proteins that mediate pesticide detoxification (Johnson et al. 2009b). Alternatively, since sub-lethal pesticide exposure can increase susceptibility to pathogen attack (Bendahou et al. 1997), it is possible that colonies afflicted with CCD are less tolerant to environmental pesticide exposure and consequently are more susceptible to pathogen attack, which leads to collapse.

While higher levels of coumaphos may benefit colonies by controlling mite populations (vanEngelsdorp et al. 2009), this hypothesis does not explain completely why pesticides *not* used in beekeeping are important discriminating variables when distinguishing control colonies from CCD colonies. As determined by the CART analysis (Tables 1 and 2), the pesticides that are important distinguishing variables come from diverse classes such as coumaphos (an organophosphate), esfenvalerate (a pyrethroid), dicofol (an organochlorine), iprodione and chlorthalonil (two fungicides), and endosulfan (a cyclodiene). More work is needed to explain why some exogenous chemicals are positively associated with CCD while others are negatively associated.

As in the current study, fluctuating asymmetry (FA) was found to discriminate between CCD and non-CCD colonies in our earlier mono-factorial comparisons (vanEngelsdorp et al. 2009). In this current effort, FA was an important discriminating factor in both CART models (without a misclassification cost: 2nd most predictive variable, p = 50.15; with a misclassification cost: 3rd most predictive variable, p = 42.48). FA, defined as random differences in the shape or size of a bilaterally symmetrical character (Palmer and Strobeck 1986), can be an indicator of individual fitness (VanValen 1962) because organisms exposed to stress during their development show less symmetry than unstressed organisms (Tuyttens 2003). Average FA score of worker bees has previously been suggested as a measure of colony level fitness (Schneider et al. 2003). While measuring fluctuating asymmetry is a less sensitive test when it comes to differentiating control colonies from CCD colonies as compared to other variables, it is a more practical test than expensive and time consuming pesticide analyses needed to determine coumaphos levels in brood and beebread. It is not, however, as easily measured as some other discriminating variables (such as head mass). The value of FA as a measure to predict colony health in general and CCD in particular, warrants further investigation.

Head masses between of bees from CCD and non-CCD populations were not significantly different overall (vanEngelsdorp et al. 2009). However, as a discriminating risk factor in CART model with a cost of misclassification, head mass appears to be important. For instance, of the 31 individual colonies that had low coumaphos levels in beebread (≤ 44 ppb), those from control colonies had heavier heads (Figure 2). The heads of winter bees are about 15% lighter than the heads of summer bees (Meyer-Rochow and Vakkuri 2002), which may be the result of reduced hypopharyngeal gland size in winter bees (Fluri et al. 1982) or because summer bees have larger brains (Meyer-Rochow and Vakkuri 2002). The volume of certain brain regions, and presumably the mass of the total bee brain, also changes as summer bees age, with antennal-lobes in forager bees being larger than 4 days old house bees (Brown et al. 2002). As bees age, the size of their hypopharyngeal glands increases for one week and then decreases (Crailsheim and Stolberg 1989). It is therefore possible that the increased head mass in healthy colonies reflects the overall age profile of the bees sampled, as bees remaining in CCD colonies are thought to be young (vanEngelsdorp et al. 2009).

The ability of individual pathogen loads to distinguish CCD and non-CCD colonies was minimal. This confirms previous findings that none of the pathogens quantified by this effort can be implicated as the sole “cause” of CCD. This is not to say, however, that disease agents play no role in CCD, as they clearly do (Cox-Foster et al. 2007, Johnson et al. 2009b, vanEngelsdorp et al. 2009). The use of CART analysis in epidemiological studies permits the identification of risk factors that are useful in disease diagnosis (Saegerman et al. 2004) as well as those that may play an important role in disease occurrence (Thang et al. 2008). This study is the first to apply this analytical tool to bee pathology in general and CCD in particular. It is important to note that this study, being an epizootiological study, did not set out to test a specific hypothesis (Koepsell and Weiss 2003) and so did not intend to identify the cause or causes of CCD. Rather, the results of this analysis are intended to act as a guide for further epidemiological- and hypothesis-driven research. To that end, the CART analysis presented here highlights several areas that warrant further attention, including the effect that sub-lethal pesticide exposure may have on pathogen prevalence, and the potential effect that tolerance to pesticides has on colony survivorship. This analysis also provides further evidence that CCD is likely the result of several factors, acting in concert, which together decrease colony fitness and make affected colonies more susceptible to disease.

ACKNOWLEDGMENTS

This research was funded by the National Honey Board and the USDA-ARS Areawide Program on bee health, the Pennsylvania Department of Agriculture, Penn State Hatch funds, the North Carolina Agriculture Foundation, a grant from the North Carolina Department of Agriculture & Consumer Services, the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service grant number 2007-02281, and the University of Liege Belgium.

References Cited

Atkins, E. 1992. Injury to honey bees by poisoning, pp. 1155-1203. *In* J. M. Graham [ed.], The Hive and the Honey Bee. Revised Edition. Bookcrafters, Hamilton, IL.

Bendahou, N., M. Bounias, and C. Fleche. 1997. Acute toxicity of cypermethrin and fenitrothion on honeybees (*Apis mellifera mellifera*) according to age, formulations and (chronic paralysis virus)/insecticide interaction. Journal of Environmental Biology 18: 55-65.

Bogdanov, S., V. Kilchenmann, and A. Imdorf. 1998. Acaricide residues in some bee products. Journal of Apicultural Research 37: 57-67.

Breiman, I., J. H. Friedman, R. A. Olsen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth, Pacific Growe, CA.

Brown, S., R. Napper, C. Thompson, and A. Mercer. 2002. Stereological analysis reveals striking differences in the structural plasticity of two readily identifiable glomeruli in the antennal lobes of the adult worker honeybee. The Journal of Neuroscience 22: 8514-22.

Claudianos, C., H. Ranson, R. M. Johnson, S. Biswas, M. A. Schuler, M. R. Berenbaum, R. Feyereisen, and J. G. Oakeshott. 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Molecular Biology 15: 615-636.

Cox-Foster, D. L., S. Conlan, E. C. Holmes, G. Palacios, J. D. Evans, N. A. Moran, P. L. Quan, T. Briese, M. Hornig, D. M. Geiser, V. Martinson, D. vanEngelsdorp, A. L. Kalkstein, A. Drysdale, J. Hui, J. H. Zhai, L. W. Cui, S. K. Hutchison, J. F. Simons, M. Egholm, J. S. Pettis, and W. I. Lipkin. 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. Science (Washington) 318: 283-286.

Crailsheim, K., and E. Stolberg. 1989. Influence of diet, age and colony condition upon intestinal proteolytic activity and size of the hypopharyngeal glands in the honeybee (*Apis mellifera* L.). Journal of Insect Physiology 35: 595-602.

Fluri, P., M. Luscher, H. Wille, and L. Gerig. 1982. Changes in weight of pharyngeal gland and haemolymph titres of juvenile hormone, protein and vitellogenin in worker honey bees. Journal of Insect Physiology 28: 61-68.

Haubruge, E., B. K. Nguyen, J. Widart, J.-P. Thomé, P. Fickers, and E. Depauw. 2006. Le dépérissement de l'abeille domestique, *Apis mellifera* L., 1758 (Hymenoptera : Apidae) : faits et causes probables. Notes fauniques de Gembloux 56: 3-21.

Johnson, R. M., Z. Wen, M. A. Schuler, and M. R. Berenbaum. 2006. Mediation of pyrethroid insecticide toxicity to honey bees (Hymenoptera: Apidae) by cytochrome P450 monooxygenases. Journal of Economic Entomology 99: 1046-1050.

Johnson, R. M., H. S. Pollock, and M. R. Berenbaum. 2009a. Synergistic interactions between in-hive miticides in *Apis mellifera*. Journal of Economic Entomology 102: 474-479.

Johnson, R. M., J. D. Evans, G. E. Robinson, and M. R. Berenbaum. 2009b. Changes in transcript abundance relating to colony collapse disorder in honey bees (Apis mellifera). Proceedings of the National Academy of Sciences 106: 14790-14795.

Karazafiris, E., C. Tananaki, U. Menkissoglu-Spiroudi, and A. Thrasyvoulou. 2008. Residue distribution of the acaricide coumaphos in honey following application of a new slow-release formulation. Pest Management Science 64: 165-171.

Koepsell, T. D., and N. S. Weiss. 2003. Epidemiologic methods: Studying the occurrence of illness. Oxford University Press, New York.

Martel, A. C., S. Zeggane, C. Aurières, P. Drajnudel, J. P. Faucon, and M. Aubert. 2007. Acaricide residues in honey and wax after treatment of honey bee colonies with Apivar(R) or Asuntol(R)50. Apidologie 38: 534-544.

Meyer-Rochow, V. B., and O. Vakkuri. 2002. Honeybee heads weigh less in winter than in summer: a possible explanation. Ethology, Ecology & Evolution 14: 69-71.

Mullin, C. A., M. Frazier, J. L. Frazier, S. Ashcraft, R. Simonds, D. vanEngelsdorp, and J. S. Pettis. 2010. High levels of miticides and agrochemicals in North American apiaries: Implications for honey bee health. PloS ONE 5: e9754.

Palmer, A. R., and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns; Augmented title: review. Annual Review of Ecology and Systematics 17: 391-421.

Pettis, J. S., A. M. Collins, R. Wilbanks, and M. F. Feldlaufer. 2004. Effects of coumaphos on queen rearing in the honey bee, *Apis mellifera*. Apidologie 35: 605-610.

Saegerman, C., N. Speybroeck, S. Roels, E. Vanopdenbosch, E. Thiry, and D. Berkvens. 2004. Decision support tools for clinical diagnosis of disease in cows with suspected bovine spongiform encephalopathy. Journal of Clinical Microbiology 42: 172-178.

Schneider, S. S., L. J. Leamy, L. A. Lewis, and G. DeGrandi-Hoffman. 2003. The influence of hybridization between African and European honey bees, *Apis mellifera*, on asymmetries in wing size and shape. Evolution 57: 2350-2364.

Speybroeck, N., D. Berkvens, A. Mfoukou-Ntsakala, M. Aerts, N. Hens, G. v. Huylenbroeck, and E. Thys. 2004. Classification trees versus multinomial models in the analysis of urban farming systems in Central Africa. Agricultural Systems 80: 133-149.

Suman, P., Thys E., Mfoukou-Ntsakala A., Ali L., Ouedraogo M., Van den Bossche P., Van Huylenbroeck G., Berkvens D., and N. Speybroeck. 2010. Methodology for assessing determinants of manure use in urban areas of Africa. Waste Management Research doi:10.1177/0734242X09356016.

Thang, N. D., A. Erhart, N. Speybroeck, L. Hung, T. Le Khanh, H. Cong Trinh, K. Pham Van, M. Coosemans, and U. D'Alessandro. 2008. Malaria in central Vietnam: analysis of risk factors by multivariate analysis and classification tree models. Malaria Journal 7: (30 January 2008).

Tremolada, P., I. Bernardinelli, M. Colombo, M. Spreafico, and M. Vighi. 2004. Coumaphos distribution in the hive ecosystem: case study for modeling applications. Ecotoxicology 13: 589-601.

Tuyttens, F. A. M. 2003. Measures of developmental instability as integrated, *A Posteriori* indicators of farm animal welfare: a review. Animal Welfare 12: 535-540.

vanEngelsdorp, D., and M. D. Meixner. 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. Journal of Invertebrate Pathology 103: S80-S95.

vanEngelsdorp, D., R. Underwood, D. Caron, and J. Hayes, Jr. 2007. An estimate of managed colony losses in the winter of 2006-2007: a report commissioned by the Apiary Inspectors of America. American Bee Journal 147: 599-603.

vanEngelsdorp, D., J. Hayes, Jr., R. M. Underwood, and J. Pettis. 2008. A survey of honey bee colony losses in the U.S., Fall 2007 to Spring 2008. PLoS ONE 3: e4071.

vanEngelsdorp, D., J. H. Jr., R. Underwood, and J. S. Pettis. 2010. A survey of honey bee colony losses in the U.S., fall 2008 to spring 2009. Journal of Apicultural Research 49: 7-14.

vanEngelsdorp, D., J. D. Evans, C. Saegerman, C. Mullin, E. Haubruge, B. K. Nguyen, M. Frazier, J. Frazier, D. Cox-Foster, Y. Chen, R. Underwood, D. R. Tarpy, and J. S. Pettis. 2009. Colony Collapse Disorder: A descriptive study. PloS ONE 4: e6481.

VanValen, L. 1962. A study of fluctuating asymmetry. Evolution 16: 125-142.

Table 1. Ranking of CCD colony risk factors by overall discriminatory power without a cost of misclassifying a CCD-diagnosed colony as a non-CCD colony

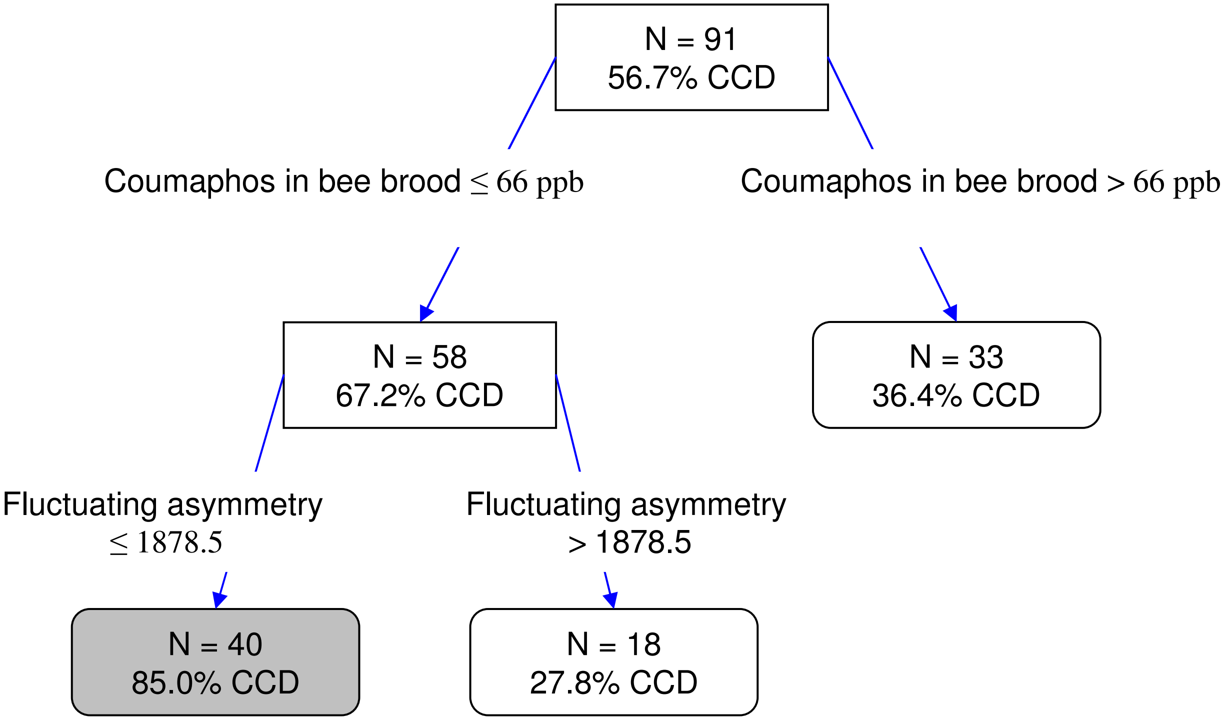
Variable	Power
Coumaphos in brood	100.00
Fluctuating asymmetry	50.15
Esfenvalerate in wax	33.91
Coumaphos in wax	29.42
Iprodione in wax	17.65
Dicofol in breebread	7.65
Chronic bee paralysis virus (CBPV)	6.77
Centriod size	5.74
Chlorothalonil in wax	5.03
Protein in the abdomen	4.49
Acute bee paralysis virus (ABPV)	3.58
Endosulfan in beebread	2.89

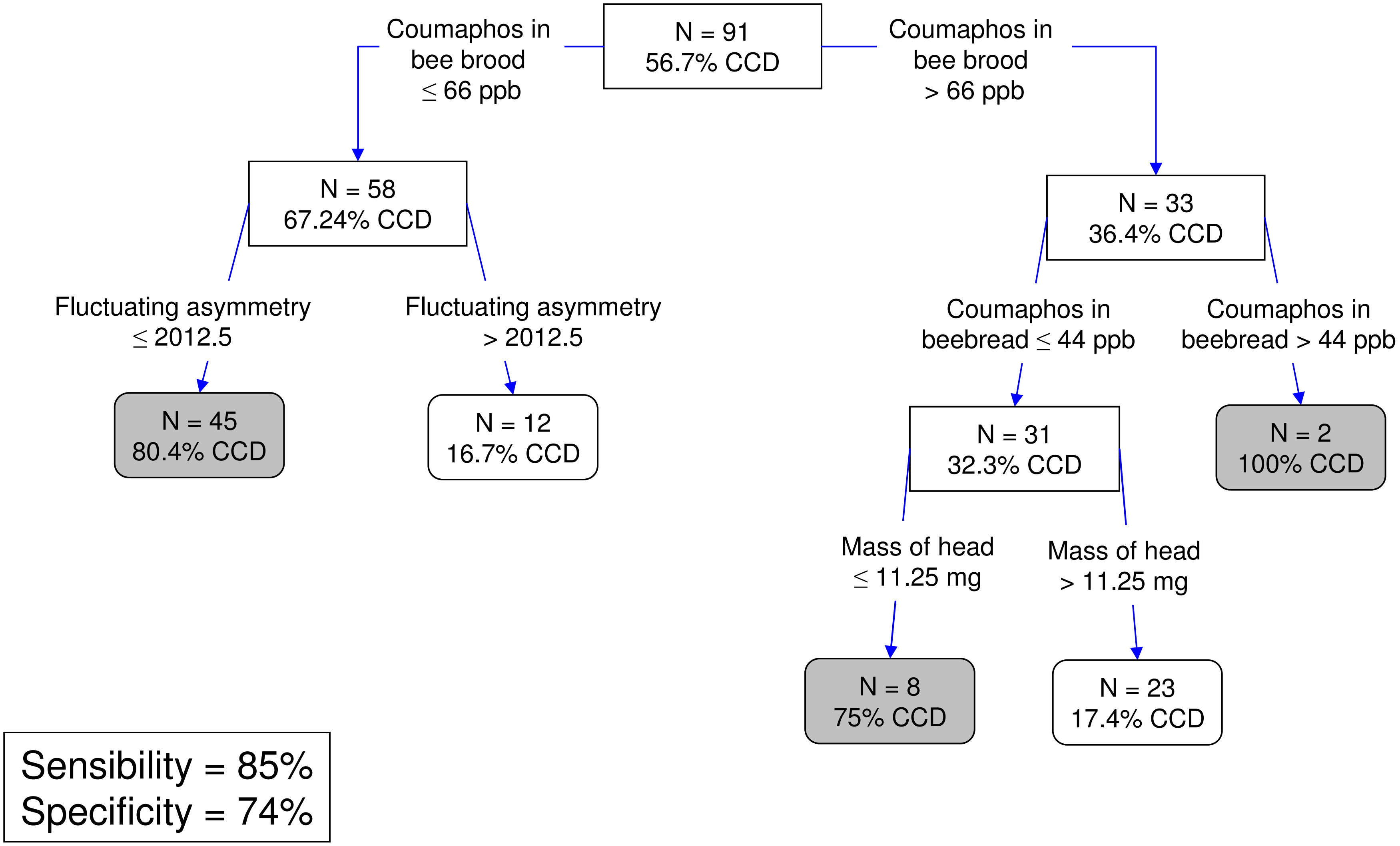
Table 2. Ranking of CCD colony risk factors by overall discriminatory power with a cost of 2 for misclassifying a CCD-diagnosed colony as a non-CCD colony

Variable	Power
Coumaphos in brood	100.00
Coumaphos in beebread	81.11
Fluctuating asymmetry	42.48
Mass of the head	36.07
Coumaphos in wax	27.39
Proteins in the thorax	12.71
Proteins in the abdomen	9.66
Acute bee paralysis virus (ABPV)	8.76
Dicofol in beebread	7.54
Proteins in the head	6.16
Centriod size	5.57
Total proteins	4.75
Chlorothalonil in wax	4.31
Mass of the abdomen	3.75
Endosulfan in beebread	2.71
Ratio proteins in the thorax / Mass of the thorax	2.57
Ratio proteins in the abdomen / Mass of the abdomen	1.91
Frames of brood	1.64
Ratio total proteins / Total mass	1.04

Figure 1. Classification tree of the risk factors for CCD colonies without a cost of misclassifying a CCD-diagnosed colony as a non-CCD colony

Figure 2. Classification and regression tree of the risk factors for CCD colonies with a cost of 1.8 points for misclassifying a CCD-diagnosed colony as a non-CCD colony


N = 91
56.7% CCD
Coumaphos in bee brood ≤ 66 ppb
Coumaphos in bee brood > 66 ppb
N = 58
67.2% CCD
N = 33
36.4% CCD
Fluctuating asymmetry ≤ 1878.5
Fluctuating asymmetry > 1878.5
N = 40
85.0% CCD
N = 18
27.8% CCD


N = 91
56.7% CCD
Coumaphos in bee brood ≤ 66 ppb
Coumaphos in bee brood > 66 ppb
N = 58
67.24% CCD
Fluctuating asymmetry ≤ 2012.5
Fluctuating asymmetry > 2012.5
N = 45
80.4% CCD
N = 12
16.7% CCD
N = 33
36.4% CCD
Coumaphos in beebread ≤ 44 ppb
Coumaphos in beebread > 44 ppb
N = 31
32.3% CCD
N = 2
100% CCD
Mass of head ≤ 11.25 mg
Mass of head > 11.25 mg
N = 8
75% CCD
N = 23
17.4% CCD
Sensibility = 85%
Specificity = 74%