Science of the Total Environment 753 (2021) 141629

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Prioritizing changes in management practices associated with reduced winter honey bee colony losses for US beekeepers

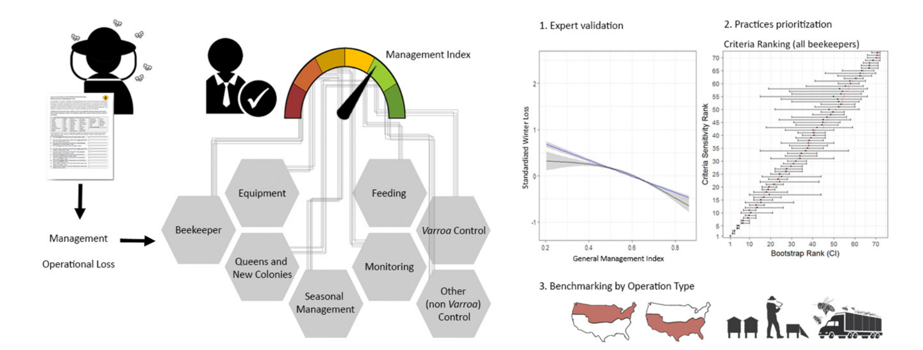
Nathalie Steinhauer [a,*], Dennis vanEngelsdorp [a], Claude Saegerman [b]

[a] Department of Entomology, University of Maryland, College Park, MD 20742, USA
[b] Faculty of Veterinary Medicine, Department of Infectious and Parasitic Diseases, Fundamental and Applied Research for Animal and Health (FARAH) Center, University of Liège, Liege 4000, Belgium

HIGHLIGHTS

- An innovative and integrative approach to assess beekeeping practices using observational data
- Demonstrated the association between management practices quality and overwintering success
- Validated experts' opinion on best practices, and provided new insights on their relative ranking
- Evidence based prioritization of behavior changes tailored to the type of operation

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 May 2020
Received in revised form 6 August 2020
Accepted 9 August 2020
Available online 19 August 2020

Editor: Henner Hollert

Keywords:
Multicriteria analysis
Variable prioritization
Expert elicitation
Best management practices
Beekeeping
Honey bee health

ABSTRACT

Beekeepers attempt to manage their honey bee colonies in ways that optimize colony health. Disentangling the impact of management from other variables affecting colony health is complicated by the diversity of practices used and difficulties handling typically complex and incomplete observational datasets.

We propose a method to 1) compress multi-factored management data into a single index, to holistically investigate the real world impact of management on colony mortality, and 2) simplify said index to identify the core practices for which a change in behavior is associated with the greatest improvement in survivorship.

Experts scored the practices of US beekeepers ($n = 18{,}971$) documented using four years of retrospective surveys (2012–2015). Management Index scores significantly correlated with loss rates, with beekeepers most in line with recommendations suffering lower losses. The highest ranked practices varied by operation type, as recommendations accounted for the current prevalence of practices. These results validate experts' opinion using empirical data, and can help prioritize extension messages. Improving management will not prevent all losses; however, we show that few behavioral changes (in particular related to comb management, sources of new colonies and *Varroa* management) can lead to a non-negligible reduction in risk.

1. Introduction

Management practices impact the health and productivity of livestock. By providing good management, livestock are protected from environmental stressors, have reduced disease burdens and are more productive (Cronin et al., 2014; Huneau-Salauen et al., 2015; Mitchell et al., 2012). Inappropriate management, such as overcrowding, can have negative consequences for livestock health and productivity (Cronin et al., 2014). European honey bees (*Apis mellifera* L.) are a semi-domesticated species (Oldroyd, 2012), housed in artificial structures and subject to human selection, but unrestricted as they forage

* Corresponding author.
E-mail address: nsteinha@umd.edu (N. Steinhauer).

https://doi.org/10.1016/j.scitotenv.2020.141629
0048-9697/

in the surrounding landscape. Managed honey bees are a complex and highly valued study system. Not only do bees produce honey, they provide critical pollination services required by many US crops, a service estimated at over $16 billion annually (2009 US data, Calderone, 2012).

A variety of stressors can cause a honey bee colony to be lost, with major drivers including *Varroa* and its related viruses, pesticides, and nutrition (Potts et al., 2010; Steinhauer et al., 2018; vanEngelsdorp and Meixner, 2010). Further, bees are often exposed to multiple stressors at the same time, which in some cases can act synergistically, exceeding the negative effects either cause individually (Tosi et al., 2017; Straub et al., 2019; van Dooremalen et al., 2018; Johnson et al., 2013; Alaux et al., 2010). While some variables are outside of a beekeeper's control (e.g. weather, external pesticide applications, habitat quality), others can be regulated through management choice (e.g., supplemental feeding, preventative and curative pest controls). Beekeeper management competence (e.g. incorrect or late varroacide application) can greatly influence colony survival (Steinhauer et al., 2018; Giacobino et al., 2015; Jacques et al., 2017; Chauzat et al., 2016). Methods to test the individual or interactive effects of a few management practices on health outcomes are well established. For instance, monitoring *Varroa* load after treatment (to assess treatment effectiveness), disinfecting hive woodenware, and providing supplemental feeding are associated with reduced *Varroa* infestations (Giacobino et al., 2014), while replacing old comb reduces viral prevalence (Molineri et al., 2017) and queen replacement reduces colony mortality overwinter (Giacobino et al., 2016b). However, accurately assessing the relative importance of each of these management strategies remains challenging.

Given the geographic and temporal variation of stressors honey bee colonies face (e.g., winter severity and duration, small hive beetle incidence (*Aethina tumida*) (Lounsberry et al., 2010; Kulhanek and VanEngelsdorp, 2017)), best management strategies vary with region and season. The availability and effectiveness of many management practices are influenced by environmental variables, like temperature. For example, some varroacides, such as formic acid, work optimally in a narrow temperature range (Underwood and Currie, 2003; US EPA, 2017). The complexity of interacting factors that affect bee health make experimental field testing of management impractical and difficult to interpret. Moreover, it makes the experimental evaluation of more than a few practices in a blocked design prohibitive, limiting the development of data supported management plans applicable to all beekeepers. Finally, given the variability of the starting field, with beekeepers engaging in wide ranges of practices, it can be difficult, and subjective, to decide where the largest fraction of extension efforts should be focused.

Honey bee colony mortality (or loss) is an ultimate indicator of the fitness of colonies; as such, it is a practical measure of honey bee health. With high levels of honey bee colony mortalities experienced in the USA (Kulhanek et al., 2017; Lee et al., 2015; Seitz et al., 2015; Spleen et al., 2013; Steinhauer et al., 2014; vanEngelsdorp et al., 2012; vanEngelsdorp et al., 2011; vanEngelsdorp et al., 2010; vanEngelsdorp et al., 2008; vanEngelsdorp et al., 2007), and in many places around the world (Antúnez et al., 2017; Pirk et al., 2014; van der Zee et al., 2014, 2012; Castilhos et al., 2019; Neumann and Carreck, 2010) there is a need for a list of best management practices, or suites of management practices, which optimize colony survivorship (The Pollinator Health Task Force, Vilsack, and McCarthy, 2015). We note that those mortality rates should not be confounded with changes in population size, as there is no indication that populations are in decline in said localities (Food and Agriculture Organization of the United Nations (FAO), 2018).

We have developed a novel method to investigate the association between colony mortality and a wide range of management practices within real-world beekeeping operations. Using a-priori knowledge provided by a team of experts with various backgrounds that range from beekeeping to epidemiology, we summarized over 100 different management practices employed by beekeepers into a simple metric – a Management Index. This index is a direct measure of the proximity of a respondent's answers to the ideal set of practices identified by the panel of experts. To evaluate the management index's performance, we tested the association between the variability in reported management practices and operational colony loss. In a second stage, we optimized the Management Index through sensitivity analyses, an iterative process which allowed us to identify the minimum number of specific management practices that would, if adopted, have the greatest impact on colony loss rates. By performing the simplification of the index in different subsets of beekeepers, we were able to identify different sets of Best Management Practices adapted to different regions and operation types, taking into account the starting practices of those groups. This offers an objective tool to decide what changes in behavior should be promoted, given their association with the largest reduction in risk for the target population.

The data used in the study come from retrospective observational honey bee colony loss and management questionnaires (Bee Informed Partnership, n.d.). Our methodology allows the use of incomplete and semi-structured datasets with hierarchical relations between questions, typical of these surveys, and that results in different response rates between questions. This method should also have utility in other research areas where similarly structured datasets make it inappropriate to use traditional modelling approaches (such as multiple regression or multivariate analyses) that rely on a common set of full answers for all parameters.

Here we describe the step-by-step process of our new multifactorial analytic method as it applies to honey bee management. The specific objectives of this study are to 1) codify experts' opinions regarding the quality of management practices; 2) summarize a wide array of management practices in a management quality index; 3) test the association of that index with our measure of interest, operational winter loss; 4) rank the index components by order of relative importance; 5) simplify the index to its minimal adequate components.

By summarizing complex information into a convenient index reflecting both experts' opinions and empirical data, we identify key management practices for which behavioral change in beekeeper management is the greatest predictor of colony survivorship. This illustrates the utility of this methodology for the complex honey bee system. By optimizing the index to identify specific practices that maximize colony survivorship rates, we are able to make evidence based suggestions to beekeepers seeking to reduce their losses.

2. Methods

Using a-priori knowledge from experts to convert the management practices into an index allows for the evaluation of said opinions. The performance of the resulting index is both a measure of the association as well as a validation of the scoring used. An advantage of this method is in the versatility of the index building process, which can be adjusted based on pre-existing knowledge (we used a relatively simple assumption of additivity, but interactive effects could be added to the model).

2.1. Expert-based management model

First we summarizes complex information about management practices in a simple metric ranging between 0 and 1 – a Management Index – reflecting the quality of the management practices of each respondent compared to an ideal set of practices determined by a panel of experts in the fields of honey bee health and epidemiology (Table 1).

The protocol designed to construct the Management Index was inspired by Humblet et al., 2012. It involved using the management survey responses from each operation to populate a data matrix based on a series of criteria, each representing one unique aspect of beekeeping management practices (e.g. "source of queen replacement", "*Varroa* monitoring technique", etc.) (see Section 2.1.2.2 and Appendix A).

Table 1
Contributing experts.
Background and expertise of the contributing experts (as of 2015, when contribution was granted). Background, field of expertise, experience in the field and keywords were provided by the experts themselves. Contribution to the conversion of criteria options into scores (1) and attribution of weights (2) are described in Methods (Section 2.1.2.3 and Section 2.1.5).

Expert name	Background (title and positions)	Field of expertise	Experience in the field	Keywords	Contribution
Dewey Caron	Emeritus Professor, University of Delaware	Apiculture	47 yrs	Extension specialist, teacher, author, Bee Informed Partnership (BIP) stakeholder committee	(1) (2)
Wayne Esaias	PhD Biological Oceanography, Retired NASA	Remote Sensing of Nectar Phenology and Climate Effects	20–35 yrs., beekeeper 22 yrs	Environmental Effects, Nectar Flow Phenology, Climate Effects, Hive Scales	(2)
Jerry Hayes	Beelogics commercial Lead, Monsanto (now Bayer)	Apiculture, Honey bee health	30 yrs. (8 yrs. Apiary insepctor; 2 yrs. industry)	Apiary Inspector, *Varroa*, disease, commercial beekeeping	(1) (2)
Eugene Lengerich	Professor of Public Health Sciences, Penn State	Epidemiology	20 yrs	Risk factors, Prevention, Mortality, Morbidity, Community-based	(2)
Katie Lee[a]	MS in Entomology, PhD student in Entomology, BIP Tech-Transfer Team Crop Protection Agent (team leader)	Sampling commercial beekeeping	9 yrs	Sampling, *Varroa*, Fieldwork, commercial beekeepers, disease	(1) (2)
Megan Mahoney[a]	BIP Tech-Transfer Team Crop Protection Agent	Beekeeping	8 yrs	Commercial beekeeper, queen rearing	(2)
Jeff Pettis	Research leader, USDA, Beltsville Bee Laboratory	Entomology, honey bee health, toxicology, pathology	30 yrs	Pesticides, queens, disease	(1) (2)
Ben Sallmann[a]	BIP Tech-Transfer Team Crop Protection Agent	Beekeeping	5 yrs	Commercial beekeeping	(2)
Rob Snyder[a]	BIP Tech-Transfer Team Crop Protection Agent	Beekeeping	9 yrs	Commercial beekeeping, queen rearing	(2)
Marla Spivak	Professor and Extension Entomologist, Department of Entomology, Univ Minnesota	Honey bee health, behavior, pathology	30 yrs	Social immunity, breeding, behavior, management	(2)
Liana Tiegen[a]	BIP Tech-Transfer Team Crop Protection Agent	Beekeeping	5 yrs	Tech-Transfer Team, Commercial Beekeeping, University of Florida (UF's HBREL), honey production, pollination	(1) (2)
Ellen Topitshofer[a]	BIP Tech-Transfer Team Crop Protection Agent	Entomology, Apiculture	3 yrs	Scientist-in-training, field technician, educator, lab technician, public speaker	(1) (2)
James Wilkes	PhD in Computer Science, Beekeeper	Computer Science Education, Beekeeping	30 yrs. in computer science, 15 in beekeeping (sideliner with about 100 colonies for past 4 years)	Computing, programming, teaching, software, grants, sideline, honey, sourwood, marketing, farming	(1) (2)
Dan Wyns[a]	BIP Tech-Transfer Team Crop Protection Agent	Commercial Beekeeping, Pollination, Honey Production, Queen Rearing	BIP: <1 yr, Commercial: 8 yrs., Inspector: 3 yrs	Commercial beekeeper, BIP, apiary inspector, pollination, queens	(1) (2)

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the U.S. Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors' affiliated institutions. Use of trade names is for identification only and does not imply endorsement by any of the groups named above.
[a] The persons identified by the asterisks were all field specialists of the Bee Informed Partnership Tech Transfer Team, working with commercial beekeepers as field consultants trained into the monitoring of pests and diseases.

Practically, a criterion could encode information from one or more survey questions. For each criterion, the potential answers were grouped and scored by experts (from 0 (worst) to 4 (best)) in accordance to their understanding of the associated benefit and risk of each option. For example, for the criterion "*Varroa* monitoring technique", the answer "alcohol wash" received a score of 4 as the method is deemed highly accurate, whereas "visual inspection of adult bees" received a score of 0 as this method is known to be unreliable. Next, the criteria scores were weighted based on the experts' opinion of the relative importance of each criterion for colony mortality risk (see Section 2.1.5 and Appendix B). For example, the criterion "source of queen replacement" received a higher weight from the experts than "*Varroa* monitoring technique", suggesting this has a greater potential impact on colony mortality risk in our experts' eyes.

2.1.1. Recruitment of the panel of experts

We contacted a group of experts in the fields of honey bee health and epidemiology (authors who frequently publish honey bee research, university professors, extension specialists, industry leaders, and field technicians) from which fourteen responded favorably to our request to contribute to our management scoring model (Table 1). Eight experts contributed their opinions in order to convert criteria into scores (see Section 2.1.2.2 and Appendix A) and all fourteen provided their opinions regarding each of the criteria's relative contribution to colony survivorship, which was used to weight factors when summing criteria scores into the Weighted General Management Index (see Section 2.1.5 and Appendix B).

2.1.2. Summarization of management information

2.1.2.1. Survey design. Our analysis utilized four years of survey data (2011–12, 2012–13, 2013–14, 2014–15), which included a total of 18,971 US-based, unique response sets that provided sufficient information to calculate a valid operational Winter Loss and completed the survey's Management section (with a minimum of 10 valid criteria) (Table S1). The survey design, data acquisition process, and loss rates estimations are detailed in the supplementary materials and respective yearly publication on loss estimates (Spleen et al., 2013; Steinhauer et al., 2014; Lee et al., 2015; Seitz et al., 2015).

2.1.2.2. Encoding of the management information into criteria. The answers to the 100+ survey questions on operational management practices were condensed and encoded into 82 distinct management criteria (Appendix A) defined a priori to capture a comprehensive picture of most aspects of colony management. These 82 criteria were then further sub-divided into 8 domains: 1. Beekeeper; 2. Equipment; 3. Queens and

New Colonies; 4. Seasonal management; 5. Feeding; 6. Monitoring; 7. *Varroa* control strategies; 8. Other disease/pest (non-*Varroa*) control strategies. Each domain was composed of between three and 29 criteria.

We then identified the range of potential answers given to each of our criteria from the Management Survey respondents. This was straightforward for criteria comprised of only one survey question and one set of answers. We standardized the possible responses to criteria derived from more than one survey question so that inconsistent responses could be accounted for. For instance, the criteria "*Varroa* product used" was based on two survey questions: a binary (yes/no) question "Last year, did you use a treatment to try to control *VARROA* MITES in your colonies?" as well as a multiple choice question (multiple selections allowed with an open-ended option) asking respondents to select which product they had applied in their colonies over the preceding year. We identified 3 potential answers: "Yes" with a selection of at least one varroacide, "No" without selection of a varroacide, and two inconsistent combinations: "Yes" without a varroacide being selected (maybe because the respondent didn't remember the name of the product) or "No" with a varroacide selected (maybe because the respondent didn't realize the target of the product used).

2.1.2.3. Conversion of criteria options into scores. Once all the answer options for each of the criteria had been identified, we then assigned each option a score on a five point ordinal scale, from 0: "Greatly decreases chance of survivorship" to 4: "Greatly increases chance of survivorship", based on our understanding of beekeeping and epidemiology (Appendix A). Our panel of experts was then asked to critically evaluate the grouping and scoring values assigned to each criterion. Practically, they recorded that they strongly disagreed, disagreed, agreed, or strongly agreed with the assigned scores. If an expert strongly disagreed with a score they were asked to explain their objection and propose an alternative. Each expert's opinion was recorded individually from the others to avoid peer pressure. In cases where there was disagreement among the experts for a given criterion, the final score was decided by rule of majority.

Using this scoring method, we then converted all of the respondents' management answers into 82 criteria scores ranging from 0 to 4. To continue the example presented above, if the respondent answered "Yes" and selected at least one varroacide product from the list (or in the open-entry), they would receive a score of 4 for that criterion. If they answered "No" and did not select any varroacide product, they would receive a score of 0 for that criteria. If they provided any other combination of answers, therefore showing an inconsistency, they received a score of 2 for that particular criterion.

Missing answers ("NULL") were handled via imputation (see Section 2.1.4). Non-applicable questions ("QNAs") did not receive a score (and were not counted in the denominator), meaning those questions did not count against the respondents. The number of criteria actively included in the index therefore varied by respondent (most between 20 and 40). Respondents who provided fewer than ten valid answers out of the 82 management criteria addressed in this study were removed from the dataset. Most respondents provided enough information to encode between 20 and 40 criteria scores (Fig. 1a).

2.1.3. Missing scores imputation

A respondent's criteria score which could not be determined due to missing information ("NULL") received a placeholder value. To determine which placeholder value was most appropriate we compared the results of four imputations methods in which missing criteria scores were replaced by: the lowest score possible for that criterion; the mean score of all valid respondents for that criterion; the median score of all valid respondents for that criteria; or any score at random from the existing levels for that criteria ("zero", "mean", "median", "random", respectively). Five versions of the General Management Index (without and with imputation for the missing scores) were built and the performance of each was compared to one another (see Section 2.2.1.1).

2.1.4. Weighting of criteria scores

In addition to using experts to convert criteria options into scores, we asked them to determine, in their opinion, the relative importance of each practice on colony survivorship. We used this information to assign unequal weights to criteria scores when summing them into the Weighted General Management Index.

We recorded the experts' opinions regarding the relative importance of management criteria using the "Las Vegas Technique" (Humblet et al., 2012). This involved allocating a pre-determined number of "points" between the criteria grouped in each of the eight domains (CP_i). The number of points available to distribute between criteria within a domain differed to account for variation in number of criteria per domain. Experts were asked to distribute the assigned points within a domain according to their opinion of each criteria's relative importance, where a greater number of points reflected a greater impact of that criterion

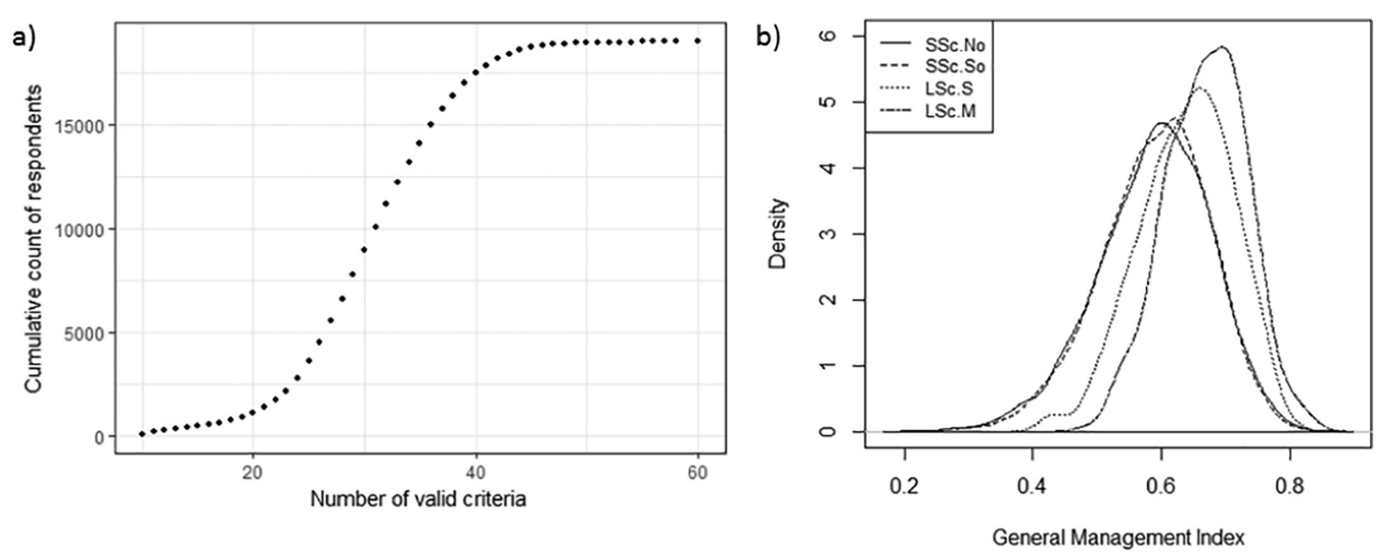


Fig. 1. Response rate and General Management Index scores. a) Cumulative frequency distribution of the number of valid criteria scores used to calculate the General Management Index for each valid respondent; b) probability density distribution of the General Management Index by operation type (all years, $n = 18{,}971$ respondents). Legend for operation type: SSc.No = small-scale North, SSc.So = small-scale South, LSc.S = large-scale single-state, LSc.M = large-scale multi-states.

on colony health. Experts were then asked to give relative weights to each domain, distributing a total of 100 points among the eight domains (DP_i). As with the criteria ranking strategy, we asked experts to assign more points to domains they considered relatively more important with respect to their effect upon colony health.

For each expert's scores we then calculated the relative weight associated with each criterion (W_i, Eq. (1)) by taking into consideration both the relative importance of the domain the criteria was in (DP_i), and the relative importance of the criteria within that domain (CP_i). Before averaging the opinion of all experts, we re-proportioned that quantity so that the sum of all criteria weights for one expert added up to 1000, to ensure that all experts contributed equally. The experts' opinions were aggregated as the average of W_i across all experts (Appendix B).

$$W_i = \frac{CP_i * DP_i}{\sum_{j=1}^{n} \left(CP_j * DP_j\right)} * 1,000 \tag{1}$$

Equation 1: The relative weight of criteria, i, for one expert (W_i), is the product of the number of points allocated to that criterion inside its domain (CP_i) and the number of points allocated to that domain (DP_i), divided by the total number of points distributed by that expert over all criteria (sum of CP_j*DP_j for all criteria j from 1 to $n = 82$) and re-proportioned to a common 1000 points to allow for comparison across experts.

We used the average criteria weight of all contributing experts to modify the criteria scores (original scale of 0 to 4) into weighted criteria scores.

Average domain points (DP_i) and criterion weights (W_i) were subjected to Chi^2 tests to determine whether their distributions were significantly different from an equal distribution, in order to identify which criteria and domains were perceived by the experts as more important drivers of colony success than expected under a null hypothesis.

2.1.5. Criteria exclusion

Before aggregating the weighted criteria into the General Management Index, we tested the robustness of the individual criteria based on self-imposed standards. We imposed a benchmark requiring a 70% minimum response rate, excluding cases where respondents' answers were not applicable (QNA), in order for a criterion to be included in our model. For each criterion, respondents provided answers to survey questions 37 to 100% of the time (excluding non-relevant follow up questions, coded as QNA); four of the 82 criteria failed to reach the required 70% minimum response rate and were therefore excluded from further analyses (Appendix B).

We also excluded criteria lacking contrasts (meaning there were no comparison possible), based on a ruling of requiring a minimum of 30 respondents for at least two scores. Practically, this resulted in the exclusion of six criteria which had low number of responses or were mostly unanimous (or very close so) in our sample population. Those questions were mostly follow-up questions on particular products use and probably too detailed to be relevant to a majority of respondents. As a result, 72 criteria were included in our General Management Index models (response frequency listed for each criteria in Appendix B).

2.2. Aggregating criteria scores into the General Management Index

If the criteria scores are summed with equal weights, they compose the Unweighted General Management Index. Otherwise, the weighted criteria scores were summed up to compose the Weighted General Management Index.

The weighted criteria scores (S_i*W_i, Eq. (2)) were summed to compose the Weighted General Management Index. This Weighted General Management Index reflects the assumption of simple additivity of criteria of unequal relative importance (some criteria likely having a higher influence than others, but all contributing independently to mortality risk).

$$GMI = \sum_{i=1}^{n} \left(S_i * W_i\right) \tag{2}$$

Equation 2: The Weighted General Management Index (WGMI) is the weighted (W) sum of criteria scores (S) for all criteria (where $i = 1$ to 82). Criteria weights (W_i) were assigned by experts (see Section 2.1.5). The Unweighted GMI represents the particular case of Eq. (2) in which all weights are equal to 1.

Comparisons of index scores between beekeeper sub-groups were performed using an analysis of variance (aov, from library "stats" (R Core Team, 2017)) and post-hoc Tukey tests.

2.2.1. Management model selection

2.2.1.1. Selection of Management Index version. The summarization of management information using experts' opinions led us to construct 10 versions of the General Management Index, based on a combination of two weighting methods (weighted and unweighted criteria scores) and five methods of handling missing scores (with and without the four types of imputation). We compared their relative performance using simple correlation (Pearson's moment correlation) between the index and the standardized operational Winter Loss.

2.2.1.2. Management Index performance. We hypothesized that beekeepers whose management practices were more closely aligned with experts' recommendations would experience lower risk of colony mortality. We thereby predicted that higher Management Index values would be associated with lower overwintering losses. This association was tested with a Pearson's product-moment correlation between the Management Index and standardized operational Winter Loss.

Management Index performance was also evaluated with a generalized additive model (gam, library "mgcv") using the standardized operational Winter Loss as the response variable and the Management Index as the predictive variable. Gam models have the advantage of not assuming the shape of the relationship between variables, which allowed us to test for potential curvatures in the relationship without a priori information about the shape of the relationship. We also reported the linear regression results (lm), when we deemed it a conservative estimate of the curvature identified in gam. The impact of covariates, in particular the type of operation and survey year, were also investigated using an analysis of covariance.

2.2.2. Comparison across operation types and regions

We aggregated our respondents into four groups based on a combination of operation size and region. Small-scale beekeepers (i.e. backyard beekeepers, managing fewer than 50 colonies on October 1st) were divided between those managing colonies in Northern States and those managing colonies in Southern States (based on the demarcations of NOAA's US climate regions (Karl and Koss, 1984), with "North" defined as the grouping of states from "Northwest", "West North Central", "East North Central", "Central", and "Northeast", while "South" was defined as the grouping of states from "West", "Southwest", "South", and "Southeast"). Large-scale beekeepers (i.e. sideliners and commercials, managing over 50 colonies on October 1st) were divided between single-state and multi-states if they kept colonies in more than one state over the year.

2.3. Optimization of the Management Index

After confirming the usefulness of the General Management Index as an indicator of management quality associated with reduced colony mortality, we aimed to optimize this index by identifying the key components responsible for driving the association. First we ranked all of the component criteria of the GMI using sensitivity analyses. Then we used this ranking to simplify the GMI one step at a time, in the same logic as for classical model simplification (Crawley, 2007), until an optimum performance was found, which we refer to as the Optimal Management Index.

2.3.1. Sensitivity analyses

Sensitivity methods aim to decompose the total variance of a model's output into the contributions of each input factor, in our case each weighted criteria score. The simplest of those methods consist of varying “One Factor At a Time” (OAT) and measuring the change in the performance of the new model compared to the baseline model (Saltelli et al., 2006). We adapted this technique to rank our management criteria based on their impact on the relationship between our Management Index and standardized Winter Loss.

In practice, we compared the performance of our General Management Index (GMI, Eq. (2), combining N criteria), to N simplified Management Index each combining N-1 criteria (SMI). In other words, we ranked criteria based on how sensitive the performance of our Index was to their removal. The change in the Pearson's moment correlation value (|cor GMI|-|cor SMI|) between the Index and standardized Winter Loss was used as an indication of the contribution of the criteria to the performance of the index. OAT methods are only applicable because our index is linear by design and more complex sensitivity methods should be used if applied to non-linear models (e.g. those where there is an interaction between management practices).

2.3.2. Bootstrapping

The sensitivity analyses were performed with bootstrapping as internal validation method to quantify the uncertainty in the ranking of criteria. We used the library (boot)(Canty and Ripley, 2015), with B = 10,000 bootstrap resamples (with replacement from the original sample, or non-parametric method), n-out-of-n method and bootstrap percentile-t method for confidence interval calculations. We opted to use percentile-t over normal CI to better reflect the asymmetry in the distribution of the bootstrap estimates.

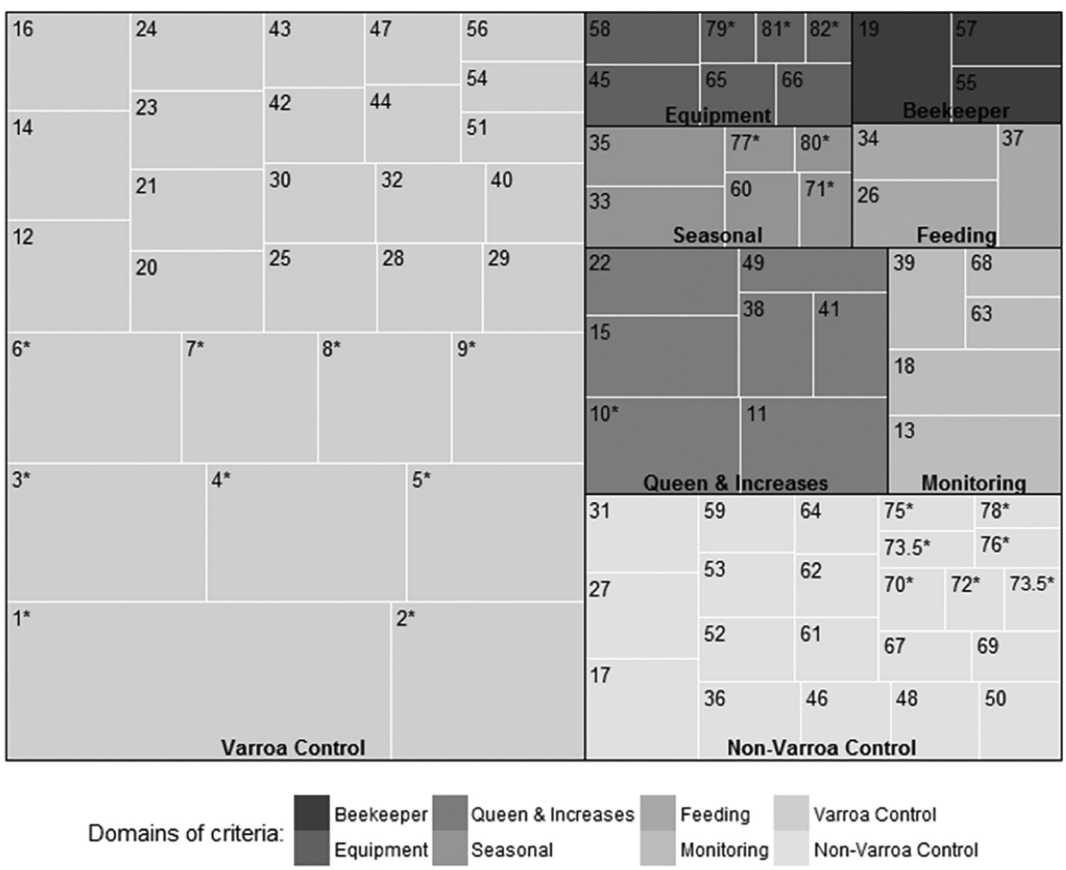


Fig. 2. Tree map of the experts' Criteria Weights (CW). Weights allocated to each of the original criteria ($n = 82$) grouped by domains ($n = 8$), ranked from highest to lowest (with ties) (area size proportional to CW value, Appendix C). The asterisks (*) indicates criteria which received significantly more or less weight than under an equal-weight distribution hypothesis (Chi2 = 707.2, df = 81, $p < .001$). The legends also show the ten criteria excluded from the analysis as greyed out. Legend: PCol (percent colony), Freq (frequency), SBB (Screened Bottom Board), SHB (Small Hive Beetle), FB (Foulbrood), TM (Tracheal mites). Legend: 1* Varroa Treatment (Y/N); 2* Varroa Products Applications (count); 3* Amitraz use (season); 4* Varroa IPM Practices (count); 5* Varroa Products Type (count); 6* Amitraz use (count); 7* Formic Acid use (season); 8* Oxalic Acid use (season); 9* Amitraz use (PCol); 10* Average Queen Age; 11 Queens Replaced (Y/N); 12 Thymol use (season); 13 Varroa Monitoring (Freq); 14 Formic Acid use (Pcol); 15 Queens Replaced (PCol); 16 Oxalic Acid use (PCol); 17 FB Treatment (motive) (excluded); 18 Brood Inspection (Freq); 19 Years of Beekeeping; 20 Thymol use (Pcol); 21 Thymol use (count); 22 Started New Cols (Y/N); 23 Drone Removal (Freq); 24 SBB (PCol); 25 Hop Oil use (season) (excluded); 26 Feeding (Y/N); 27 Terramycin use (season); 28 Drone Removal (PCol); 29 SBB (months); 30 Coumaphos (Varroa) use (season); 31 Nosema Treatment (Y/N); 32 Fluvalinate use (season); 33 Honey Produced (lbs); 34 Feeding (season); 35 Crops (count); 36 SHB Control Technique; 37 Feeding Products Type; 38 Queen Source; 39 Varroa Monitoring Technique; 40 Contraindications (excluded); 41 New Colonies Technique; 42 Hop Oil use (count) (excluded); 43 Fluvalinate use (PCol); 44 Drone Removal Amount; 45 Winter Preparation Technique; 46 MiteAThol use (motive) (excluded); 47 Hop Oil use (PCol); 48 Nosema Products Applications (count); 49 ReQueening Technique; 50 Terramycin use (PCol); 51 Powder Sugar use (months); 52 Fumagilin use (season); 53 SHB Trap type (excluded); 54 Coumaphos (Varroa) use (PCol); 55 Sources of information (count); 56 Coumaphos (Varroa) use (count) (excluded); 57 Beekeeping Education; 58 October Brood Chamber Size; 59 Tylosin use (season); 60 Honey harvest (Y/N); 61 SHB Trap use (month); 62 Fumagilin use (PCol); 63 Nosema Monitoring Technique; 64 SBH Bait type; 65 Average Comb Age; 66 Comb Culling and Storage Technique; 67 Tylosin use (PCol); 68 Nosema Monitoring (Freq); 69 SHB Soil Drench use (month&PCol); 70* Coumaphos (SHB) use (season) (excluded); 71* Moved across State Lines (PCol) (excluded); 72* MiteAThol Use (season); 73.5* Coumaphos (SHB) use (count) (excluded); 73.5* Coumaphos (SHB) use (PCol); 75* MiteAThol use (PCol); 76* Nozevit use (season); 77* States (count); 78* Nozevit use (PCol); 79* Foundation type; 80* Moved across state lines (Y/N); 81* Equipment Type; 82* Action on Deadouts.

2.3.3. Index simplification

Using the sensitivity ranking of criteria as guiding order, we proceeded to simplify the General Management Index step by step, from least sensitive to most sensitive criteria, and observe the change in the model's performance. The performance of the index was measured at each step, and simplification continued until it reached an optimal value – with simpler index structure (with fewest components criteria) being preferred for equally performing indices. The Optimal Management Index (OMI) was identified as the most parsimonious index structure that was best associated with standardized operational Winter Loss.

2.3.4. Comparison across operation types and regions

Next, the bootstrapped sensitivity analyses and index simplification were performed separately for each of the four categories of beekeepers, resulting in region and operation type specific Optimized Management Indices. This allowed us to compare the complexity (number of criteria left after simplification), composition, and predicted effect size between beekeepers subgroups.

3. Results

3.1. Expert-based management model

3.1.1. Experts' criteria weights

The average number of points distributed among the eight domains (DP_i) by the 14 experts varied between 6.36 ± 0.9 for "Equipment" and 23.71 ± 2.0 for "*Varroa* Control" (± SE) (Appendix B). "*Varroa* Control" was the only domain that received more points than would be expected if domain weights were equally distributed ($Chi^2 = 16.932$, $df = 7$, $p = .018$).

The average criteria weights (W_i) allocated by the 14 experts varied from 3.06 ± 0.53 (for the criteria "Action on Deadouts") to a high of 77.01 ± 14.93 (for the criteria "*Varroa* Treatment Y/N") (Appendix B, Fig. 2). This distribution differed from an equal distribution of weights ($Chi^2 = 707.2$, $df = 81$, $p < .001$); experts allocated significantly more weight to 10 criteria (9 of 10 criteria related to *Varroa* control), and significantly less to 13 criteria (marked with an asterisk (*) in Fig. 2).

3.1.2. Encoding of management information into criteria and criteria exclusion

The unweighted General Management Index (GMI) score averaged 0.59, with an asymmetric distribution skewing towards lower GMI scores, and a range from 0.20 to 0.86 (Fig. 1b). This suggests the index expressed a wide range of management quality with which to test its association with overwintering success. It also reflected potential for improving management practices, as most beekeepers' GMI scores were below the maximum limit. A distribution with a peak close to the maximum of the index would have indicated small room for improvement, and possibly a set of experts' recommendations too self-evident and largely already applied by the stakeholder population.

Unweighted GMI scores differed across types of operation (analysis of variance, aov, test by deletion, $df = 3$, Sum of $Sq = 3.4523$, $p < .001$), with large-scale multi-state beekeepers scoring significantly higher than large-scale single-state beekeepers, followed by small-scale beekeepers from both the northern and southern regions. This means that large scale operations reported management practices that were generally more in line with experts' recommendations than small-scale beekeepers.

3.1.3. Management model selection

All versions of the GMI were significantly associated with a reduction in Winter Loss (Pearson's product-moment correlations, $df = 18{,}969$, all p-values <.001, $-21.4 < t < -17.2$, Fig. 3, Table 2), no matter which imputation and weighting methods were used. This indicates that operations with a high GMI score also reported lower colony mortality over the winter, regardless of the imputation and weighting method used to compute the index. The correlation with operational winter loss standardized by year varied between −0.124 (for the unweighted non-imputed data) to −0.154 (unweighted mean imputed data). All 95% CIs of the correlation estimates excluded the value of the null hypothesis.

With the exception of the "minimum imputation" method, the various imputation methods represented a marginal improvement over the non-imputed GMI (Fig. 3). All imputation methods were consistent in terms of directionality and significance of the association. Subsequent analyses used the best performing method of imputation using average values ("mean" imputation method). The use of criteria weights (W_i) in GMI construction did not improve index performance compared to unweighted GMI (all correlation estimates

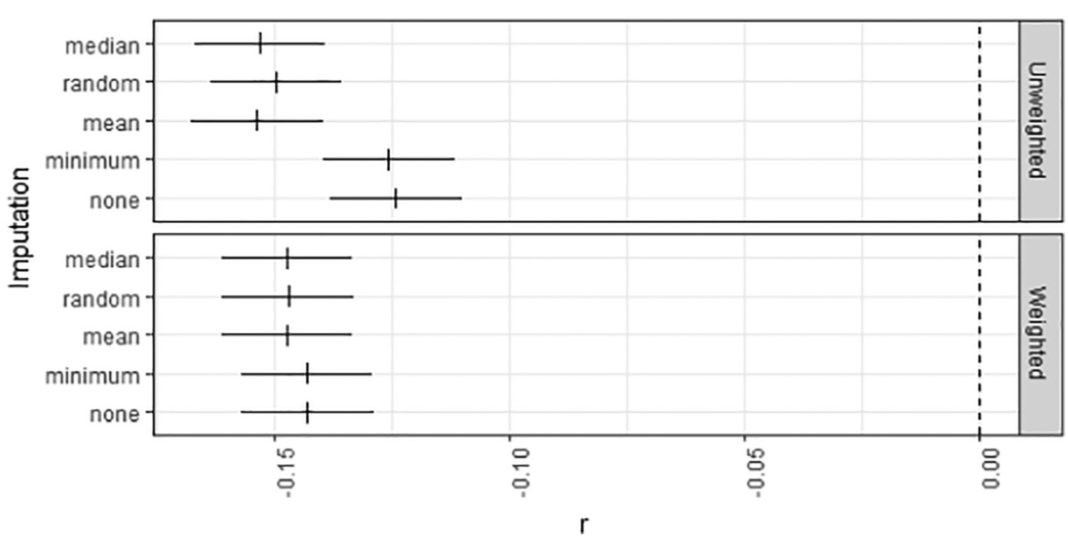


Fig. 3. Model comparison between imputation and weighting methods. Relative performance of the various versions of the General Management Index (with 10 different computations based on imputation methods for missing values and weightings of criteria scores) as tested by the GMI's association to operational winter loss standardized across years (Pearson's product-moment correlations (r and 95% CI), $df = 18{,}969$, all p-values < .0001, $-21.408 < t < -17.213$).

Table 2
Relative performance of 10 versions of the General Management Index (regarding criteria weightings and imputation) as tested by GMI's association to the operational standardized Winter Loss.

Imputation	Weight	Pearson's product-moment correlation (df = 18,969, all p-values < .0001)	
		t	Cor r [95% CI]
None	Unweighted	−17.213	−0.124 [−0.138, −0.11]
None	Weighted	−19.888	−0.143 [−0.157, −0.129]
Minimum	Unweighted	−17.436	−0.126 [−0.14, −0.112]
Minimum	Weighted	−19.918	−0.143 [−0.157, −0.129]
Mean	Unweighted	−21.408	−0.154 [−0.168, −0.14]
Mean	Weighted	−20.503	−0.147 [−0.161, −0.133]
Random	Unweighted	−20.839	−0.15 [−0.164, −0.136]
Random	Weighted	−20.472	−0.147 [−0.161, −0.133]
Median	Unweighted	−21.347	−0.153 [−0.167, −0.139]
Median	Weighted	−20.506	−0.147 [−0.161, −0.133]

95% CI overlapping, Fig. 3). We therefore used the most parsimonious index construction method based on equally weighted criteria scores. All results reported hereafter were obtained using the GMI scores calculated from implementing the "mean" imputation method and non-weighted criteria or domains. Because the weighted index did not perform better than the unweighted index, we proceeded with the more parsimonious hypothesis of simple additivity of criteria (equally-weighted criteria composing the index).

3.2. General Management Index performance

The GMI showed significant curvature in relationship to the standardized Winter Loss (generalized additive model, gam, Y = s(X), edf = 2.87, F = 134.3, p < .001) (Fig. 4A, black smooth line). The shape of the curve indicated a threshold (~0.5 score) under which management quality is not associated with overwintering loss rates. However, above this threshold, a strong association between GMI scores and decreased mortality risk is observed. The linear regression appears as a conservative estimate of the slope of the relationship above threshold. The majority (75%) of our respondents had GMI scores above threshold. Assuming a linear relationship in the section of the curve above threshold, GMI scores were significantly associated to a reduction in standardized Winter Loss (lm, Y = 1.04–1.76X, F = 458.3, df = 1, p < .001) (Fig. 4A, blue regression line). So, after a minimal number of good practices are in place (indices ≥0.5), beekeepers that reported management practices more in line with the experts' "ideal" set observed lower overwinter colony mortalities compared to the average level of loss that year. Given the slope of this relationship, we can expect that for an improvement of 0.1 in the GMI score, an operation would reduce its risk of overwintering colony loss by 0.176 standard deviations, which represents a reduction of between 5.3 and 6.6 percentage points depending on overall losses in a given year included in this study.

Both "survey year" and "operation type" were significant interacting covariates in the relationship between the index and standardized winter loss (lm, test by deletion of 2-way interaction between index and operation type: F = 7.0422,df = 3,p < .001; test by deletion of 2-way interaction between operation type and survey year: F = 11.572,df = 9, p < .001)(Fig. 4B,C). The interaction between operation type and survey year was expected, as it had already been shown in previous publications that operation types were not systematically associated with differences in winter loss across years (see yearly surveys, e.g. Kulhanek et al., 2017). The interaction between the index and operation type indicates that the slope of the relationship – the strength of the association between the index and loss – is dependent on the type of operation. The absence of a third-level interaction between the index, operation type and survey year (lm, test by deletion of 3-way interaction between index, operation type and survey year: F = 0.590, df = 9, p = .807) reassured us that the slope of the index was consistent across years for a specific type of operation (Fig. 4B,C), indicating the index created performed equally well across years.

3.3. Ranking of criteria through sensitivity analyses

The ranking represents the relative importance of the specific criteria to the performance of the management index. We interpret high ranking criteria as management practices for which a change in behavior is associated with the highest reduction in risk of overwintering colony loss. The width of the bootstrap confidence interval indicates the consistency of the ranking across multiple bootstrap resamples, therefore narrow CI indicates a consistent ranking across our respondents' population.

When performed over all respondents, the bootstrapped sensitivity analysis revealed a ranking of the management criteria clustered in three levels: a very consistent set of top management practices, followed by a wide array of interchangeable criteria whose rank, though always mediocre, was highly dependent on the subset of respondent selected by the bootstrap, and, finally, a small set of criteria consistently ranked last across all bootstrap samples (Appendix C, Fig. 5a). The lowest ranking cluster can be seen as management criteria associated with the smallest potential to improve survivorship. This could be due to either a low impact of the criteria (no difference in success between the various options for that criteria) or an already established high prevalence of the "best" behavior in the population. This three-tiered structure of the criteria ranks holds true for both small-scale subsets of beekeepers: the profile and rankings from northern and southern small-scale beekeepers were very similar to the original ranking (Appendix C, Fig. 5bc), which was expected as they represent a majority of the survey respondents. Large-scale beekeepers obtained visually different profiles and rankings, which clearly illustrates the need to consider operation types separately when prioritizing management practices. Large-scale beekeepers presented rankings that were more variable according to the random subsets of respondents considered (Fig. 5de).

The overall ranking of criteria (all participants combined) was significantly correlated with all four of the rankings by type of operation (Spearman's rank correlation, S-values between 10,542 and 31,890, p-values <.05, rho between 0.49 and 0.83) who were also all largely correlated to each other (Spearman's rank correlation, S-values between 26,246 and 44,428, p-values <.05, rho between 0.28 and 0.58). This suggests that the grand lines of the ranking hold no matter the operation type. The two small-scale subsets' rankings were the closest to one another, and to the overall ranking, which was expected as they represent the most frequent types of beekeepers in our sampled population. The two large-scale subsets' rankings were also most comparable to one another (Appendix C). Both small-scale groups (north and south) displayed a similar three-tiered structure of a small number of consistently high ranking criteria, small number of consistently low ranking criteria, with the majority of criteria having variable impacts on model performance (Fig. 5b,c).

Large-scale beekeepers (Fig. 5d,e) presented a less structured ranking profile, which rank estimates displaying larger 95% CI. Those wider intervals indicate a higher variability of the criteria's ranking between the bootstrap resamples. This is partially explained by the smaller number of large scale beekeepers who provided responses to our survey. This indicates that if "top recommendations" can be easily identified for small-scale beekeepers, such a generalization is harder to make for large-scale beekeepers. Large-scale beekeepers started with a higher average index score than small-scale beekeepers, indicating how large-scale beekeepers' practices were generally more aligned with experts' recommendations than small-scale beekeepers. Our analysis indicates that criteria holding the most potential for improving large-scale beekeepers are more variable, highlighting the need for specialist consultants to work at a more individual level with these beekeepers.

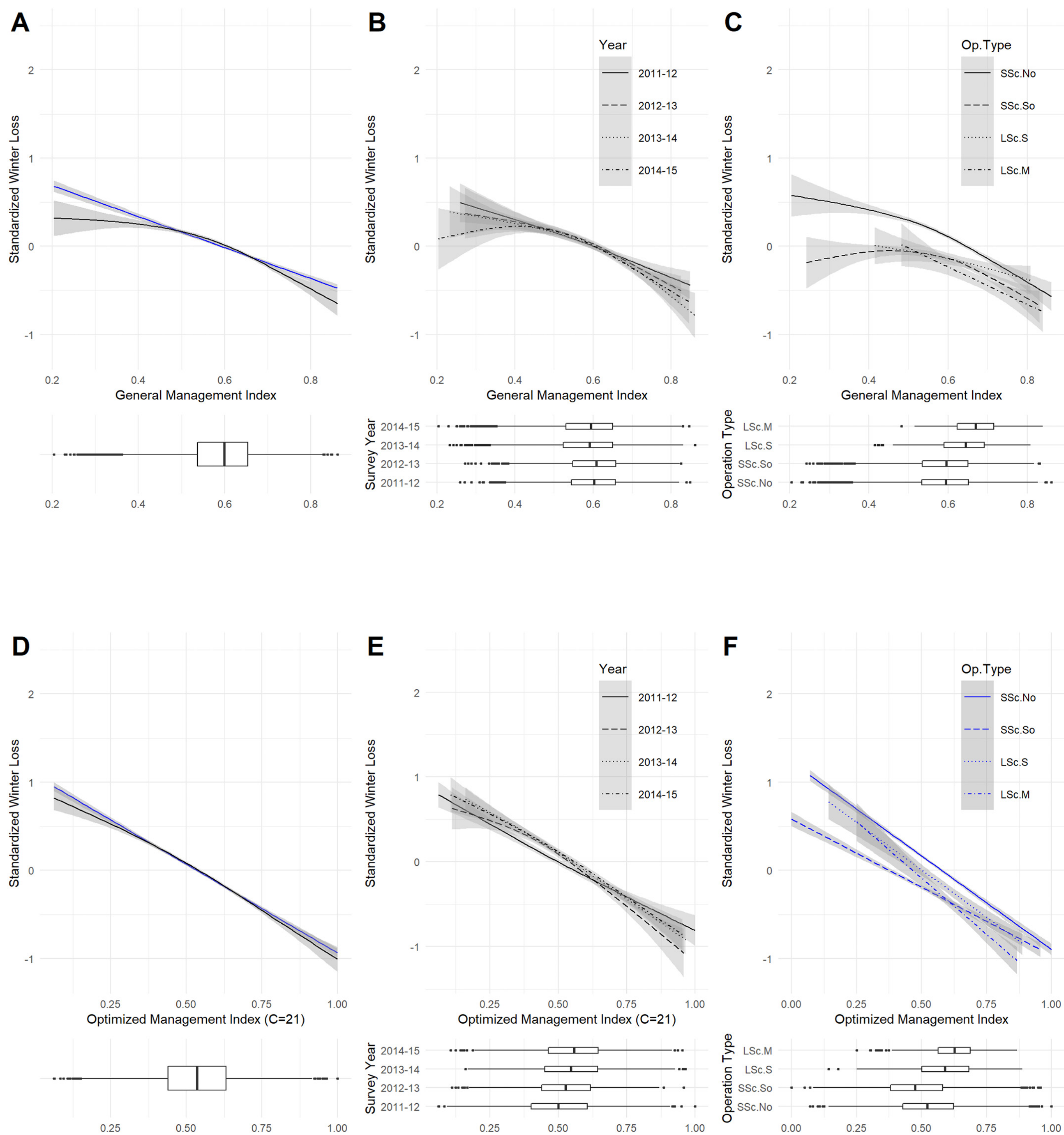


Fig. 4. Performance of the General Management Index model (A, B, C) and the Optimized Management Index model (D, E, F). A) Smoother estimate (generalized additive model, gam, in black) and linear correlation (in blue), with 95% confidence intervals, of the Standardized Winter Loss by the General Management Index ($N = 72$ criteria), across all respondents; B) gam, by survey year; C) gam, by type of operation; D) gam (black) and linear correlation (blue) of the Optimized Index ($N = 21$ criteria) across all respondents; E) gam, by survey year of the Optimized Index ($N = 21$ criteria) across all respondents; F) linear correlation of the Optimized Indices by operation type ($N = 16, 9, 15, 25$ criteria respectively). Legend for operation type: SSc.No = small-scale North, SSc.So = small-scale South, LSc.S = large-scale single-state, LSc.M = large-scale multi-states. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Management Index simplification

3.4.1. All respondents

Over all respondents, the correlation was optimized for the index composed of the 21 most sensitive criteria (Fig. 5f). Despite its simpler structure, this Optimized Management Index (OMI) model's performance was superior to the General Management Index model (Fig. 4A,D): the correlation value was higher and the relationship more linear all throughout the range of index scores. Though the OMI still presented evidence of curvature in its relationship to standardized Winter Loss (gam, $Y = s(X)$, edf $= 2.30$, $F = 4.6703$, $p = .0099$) (Fig. 4D, black smooth line), the linear relationship can be considered

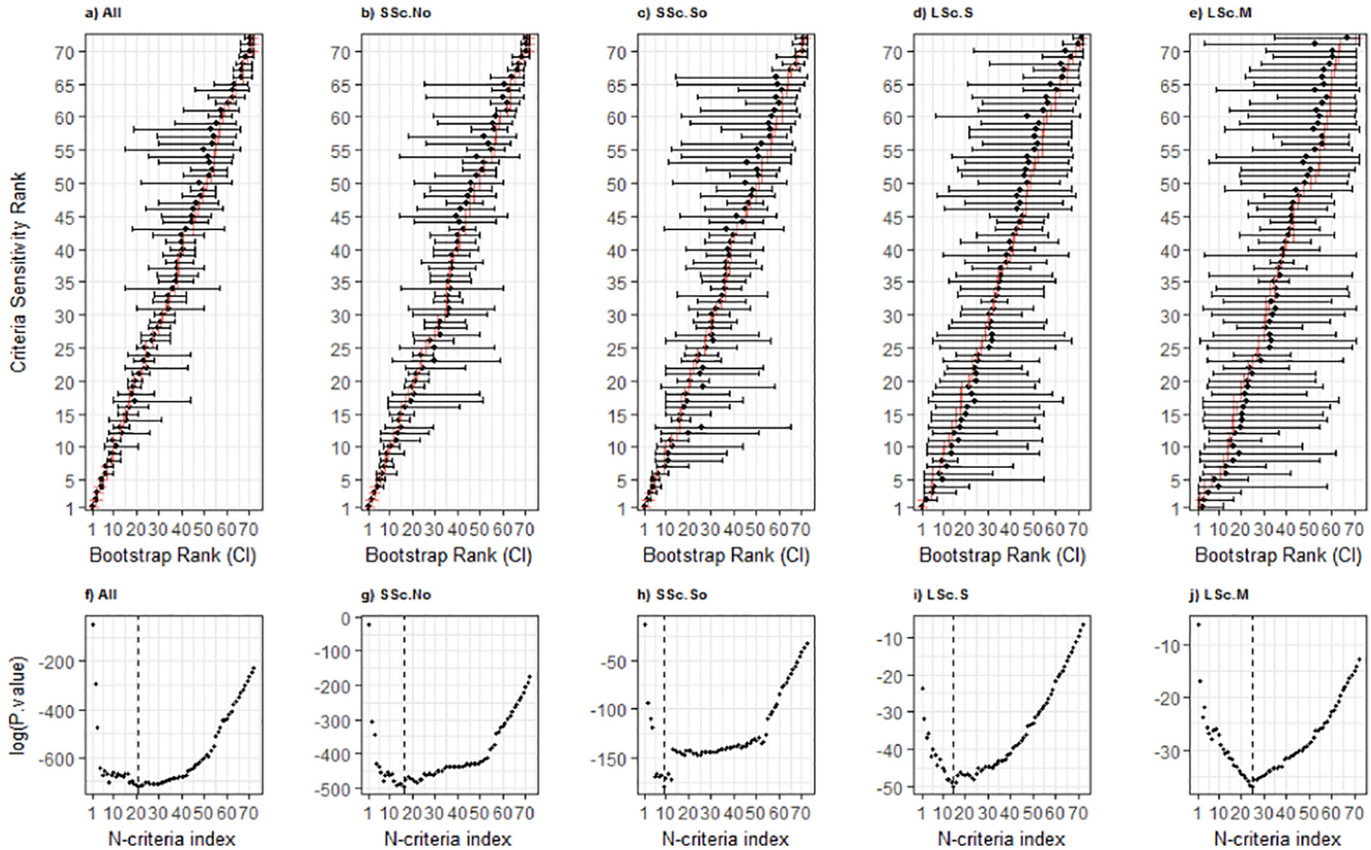


Fig. 5. Sensitivity Analyses on the General Management Index: a-e) bootstrap estimates of the rank (mean estimates of b = 10,000 bootstraps with 95%CI based on percentile distribution, bootstrap median indicated in red plus sign) compared to original ranking (for a) all participants and b-e) subsets by operation type). f-j) optimum index performance curve: p-value of Pearson correlation of the index (at various stages of simplification) to standardized Winter Loss, by increasing number of criteria in the indices (from N = 1,single criteria, to N = 72, General Management Index). Vertical dash bar indicates optimum. Legend: All = all beekeepers included, SSc.No = small-scale North, SSc.So = small-scale South, LSc.S = large-scale single-state, LSc.M = large-scale multi-states. See criteria name in Appendix D, by sensitivity rank. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a good and conservative approximation of the relationship for all but the extreme values of the index. Assuming a linear relationship, the OMI for all respondents, all years, was significantly negatively associated to the standardized Winter Loss (lm, Y = 1.07–2.01X, F = 1489, df = 1, $p < .001$) (Fig. 4D, blue regression line). This indicates that an improvement of 0.1 in the OMI score would reduce the risk of overwintering colony loss by 0.20 standard deviations, which represents a reduction in risk of between 6.1 and 7.6 percentage points between the years of our survey. More importantly, the recommendations of improved management practice that would drive this reduction in risk have been reduced to 21. The 21 management criteria included in the OMI for all respondents belonged to five different domains, though most (n = 10 of 21) represented the domain "*Varroa* Control".

Survey year and operation type were significant covariates within the OMI model (lm, test by deletion of 2-way interaction between index and operation type: F = 2,7991,df = 3,p = .0385; test by deletion of 2-way interaction between operation type and survey year: F = 13.107,df = 9, $p < .001$). However, these two covariates did not interact with each other (lm, test by deletion of 3-way interaction between index, operation type and survey year: F = 0.3784, df = 9, p = .3784) (Fig. 4E,F). This indicates that the improvement in colony survivorship associated with the increase in management quality varied between operation types and were more marked certain years than others.

3.4.2. By operation type

The same simplification process was performed for each of the subsets of beekeepers according to operation type. The optimization curve were generally similar for all four groups of beekeepers; however, the number of criteria included in the OMI varied, ranging from a low of nine criteria for small-scale southern beekeepers to a high of 25 criteria for large-scale multi-state beekeepers (Appendix D, criteria marked by *). The four OMI showed variable slopes in relation to colony mortality, but the effect sizes ranged from a 3.9 to 8.1 percentage point reduction in colony loss for each improvement of 0.1 in the index (Table 3).

All 21 criteria that were included in the overall OMI were present in at least one of the four operation-type specific OMI. In addition to those 21 criteria, another 15 criteria were included in at least one of the operation-type specific OMI (Appendix D). All told, 36 criteria appeared in at least one of the operation-type specific OMI. Two criteria were retained in all four operation types: "New Colonies Technique" and "Crops (count)", i.e. what technique was used to increase the number of colonies in the operation (e.g. in-house splits vs purchase of packages) and how many agricultural crops were present around the apiary during the active season. Another seven criteria were retained in three of the four subsets: "Action on Deadouts", "*Varroa* Treatment (Y/N)", "Comb Culling and Storage Technique", "*Varroa* Products Type (count)", "Honey produced (lbs)", "Average Comb Age", "Screened Bottom Board (PCol)". Nine criteria were retained in any two subsets and the last 18 criteria were retained in only one of the four subsets of beekeepers.

4. Discussion

We successfully implemented an original method to summarize and analyze complex management information according to the opinion of experts to produce evidence based best management practices for targeted populations. Identified practices, if implemented, could reduce

Table 3
Expected improvements by increased Optimized Management Index (OMI) score.
Effect size of an improvement of 0.1 in the OMI by beekeeper group. The change in standard deviation of operational winter loss (Δ WL StdDev) has been converted in percentage points and compared to the average operational Winter Loss observed (Obs. Av WL) by the specific group of beekeeper for that year. The adjusted average Winter Loss (Adj. Av WL) represents the risk of colony mortality if beekeepers had improved their practices by 0.1.

Beekeeper typology		OMI n criteria	Effect size: for each improvement of Index by 0.1				
			Δ WL StdDev	Year	Δ WL (%)	Obs. Av WL (%)	Adj. Av WL (%)
Small-scale	Northern	16	−0.212	2011–12	−6.647	24.03	17.39
				2012–13	−7.762	48.00	40.24
				2013–14	−8.122	49.30	41.18
				2014–15	−7.821	46.28	38.46
Small-scale	Southern	9	−0.154	2011–12	−4.654	23.76	19.11
				2012–13	−5.473	37.71	32.23
				2013–14	−5.402	32.49	27.09
				2014–15	−5.408	36.01	30.61
Large-scale	Single-state	15	−0.215	2011–12	−3.913	21.11	17.20
				2012–13	−5.570	35.65	30.08
				2013–14	−5.989	41.10	35.11
				2014–15	−4.757	28.59	23.83
Large-scale	Multi-states	25	−0.255	2011–12	−3.945	20.31	16.37
				2012–13	−5.727	31.63	25.91
				2013–14	−4.861	22.35	17.49
				2014–15	−6.104	28.71	22.61

colony loss rates suffered by US beekeepers by 5 or more percentage points per year. As discussed below, this work has implications for beekeepers, extension agents and policy makers, as well as for others attempting to prioritize recommendations from “messy” survey data in other production systems.

4.1. Improved management practices

Through the ranking and simplification process, we found that, globally, a majority of beekeepers could expect the greatest reduction in mortality risk by modifying their behavior in terms of comb management, source of new colonies and *Varroa* management. This holds particularly true for small-scale beekeepers, which represents the majority of beekeepers in our respondent pool and in the stakeholder community.

Concretely, small-scale beekeepers should adopt a more active beekeeping management, actively replacing their deadouts throughout the active season (Action on Deadouts). When brood comb was taken out of production, it should ideally not been reused unless frozen for a period of time (Comb culling and storage). The benefits of comb management support previous research that showed that newer comb better support honey bee colony health and reproduction (Berry and Delaplane, 2001). Beeswax has been shown to accumulate pesticide residues (Calatayud-Vernich et al., 2018; El Agrebi et al., 2020a), to levels that could result in increased bee mortality (El Agrebi et al., 2020b). Though colony-level effects (growth and survivorship) from wax contamination had not been confirmed through field trials (Payne et al., 2019), where the authors noted that *Varroa* levels were a far stronger predictor of colony failure.

Small-scale beekeepers starting their colonies from packages should expect a higher level of loss over the winter (New Colonies Technique) compared to the ideal situation consisting of making splits from existing colonies. Colonies started from packages are also more likely to start on undrawn foundation, which adds to the amount of honey the colony will need as they build their wax, but also could slow colony growth in the start. The production of splits (or nuclei) is also typically recommended as a swarming prevention method, which if unprevented can result in the loss of the largest fraction of the worker population, decreased production and increase risk of queen events should the requeening fail. Finally, splitting colonies is known to help reduce the *Varroa* pressure in mother colonies (Maucourt et al., 2018). It is not unusual for complete beginners to start from packages, but years of beekeeping experience did not appear as a high predictor in our small scale groups. This could appear to contradict European's findings (Jacques et al., 2017), but only because they contrasted small apiaries (with little experience) to professional beekeepers.

Finally, the importance of *Varroa* control is reflected by more than one top ranking criteria (among others, *Varroa* Treatment Y/N, *Varroa* products types (count), and various products use), highlighting the benefits of applying a strict *Varroa* control program. This suggests that some variability exists in the optimum *Varroa* control methods, but in any case, the use of any type of *Varroa* control treatment is highly associated with reduction of colony mortality risk compared to the no-treatment option. A separate paper has been dedicated to the investigation of *Varroa* control using this dataset (Haber et al., 2019). This data has also been used to highlight differences in attitudes between respondents who use *Varroa* treatments or not (Thoms et al., 2018; Underwood et al., 2019). Though this work and others have highlighted the importance of *Varroa* control, outreach specialists should still consider that “hands-off” management are associated with core beliefs that are unlikely to change without addressing their fundamental concerns (Thoms et al., 2018). Knowing the limitations and drawbacks of chemical treatments (Johnson et al., 2013; Rinkevich, 2020; Rinkevich et al., 2017), outreach should still focus on comprehensive approaches that do not rely on single “silver bullets”, but include preventive management as well as curative, for example improvement of honey bee lines through selection for hygienic behavior (Bixby et al., 2017; Evans and Spivak, 2010; Wagoner et al., 2018).

For large-scale beekeepers, practices were less generally associated with a reduction in risk. Honey production ranked highly among large-scale beekeepers' management criteria. This could indicate the importance of placing colonies in an environment conducive to good honey production. Though *Varroa* control was also associated with a high potential for risk reduction in single-state beekeepers, multi-state beekeepers' top recommendation regarding *Varroa* would be to use appropriate monitoring techniques.

Among others, our methodology brought forward unexpected results which could be translated into research opportunities. For example, the apparent lack of impact of supplemental feeding, or the varying potential of queen renewal and queen age. Though nutrition is known to be an important factor in honey bee health (Alaux et al., 2010; Di Pasquale et al., 2013), the effects of supplementation are less clear (Mortensen et al., 2019; Branchiccela et al., 2019; Giacobino et al., 2018). Queen age and queen replacement had independently been found to be associated with better colony outcomes (Oberreiter and Brodschneider, 2020; Giacobino et al., 2016b). Those results however contrast with data from the same survey reporting on beekeeper's perceived causes of colony death, in which starvation and queen failure are often cited by participants (Kulhanek et al., 2017).

In this paper, we evaluated the strength of association between management and loss rates, and estimated the effect size that can be expected by improving management practices alone. This study provides additional observational support that management practices can affect beekeeping operation success by mitigating or intensifying colony mortality risk. Though this study does not assign a cause to effect, as no observational study can, it does show a credible association between management and colony loss (i.e. that improvement in survivorship can be expected only after a minimum of management quality is reached, as suggested by the threshold in the association) and realistic effect size in real world conditions.

4.2. Management survey respondents

The Bee Informed Partnership (BIP) survey is the largest annual honey bee survey focusing on colony mortality and management practices in the US. However, there are several factors that may limit how representative the results are of the overall US beekeeping population. First, survey respondents represented a non-random subset of the target population that were successfully reached and agreed to participate. Second, results may be biased or inaccurate because they are self-reported up to one year after the fact. Third, large-scale beekeepers were more likely to complete the survey's Management section than small-scale beekeepers (of both regions) ($Chi^2 = 18.7$, $df = 3$, p-value $= .0003$). As large-scale beekeepers typically report lower levels of Winter Loss, their overrepresentation in the beekeeper pool who completed the management survey portion could explain the slightly lower Winter Losses in this group compared to beekeepers who completed the Loss survey portion (Spleen et al., 2013; Steinhauer et al., 2014; Lee et al., 2015; Seitz et al., 2015). Alternatively, beekeepers more successful in overwintering their colonies than the average beekeeper may be more likely to remember or report their practices, which could also skew this class towards lower Winter loss levels. However, the directionality of this potential bias would result in a reduction of the variability of loss reported in the survey, and a reduction of power to detect an association with management quality, which would make our association a conservative estimate rather than the opposite. Despite these limitations, our results provide insights into the range of management practices commonly used and how those practices are associated with colony loss, as the association itself should not be affected by bias.

4.3. Experts' criteria weights

The experts that helped develop our index combined a great deal of experience in the fields of beekeeping, bee health, or epidemiology. They represent all sides of the community: university researchers, field consultants, beekeepers, and private companies (Appendix A). Using expert knowledge elicitation is a recognized methodology for synthetizing published and unpublished knowledge in an area, and the recommended basis of a recent conceptual framework aimed at characterizing the Best Management Practices in beekeeping (Sperandio et al., 2019).

Experts clearly favored management criteria relating to the control of *Varroa* (Fig. 2). This is consistent with pre-established knowledge that *Varroa* poses the single greatest biological stress to honey bees, where *Varroa* is established (Genersch, 2010; vanEngelsdorp and Meixner, 2010). Due to the severity of the effects of *Varroa* and its associated viruses (Kevill et al., 2019; Wilfert et al., 2016; Dainat et al., 2012), coupled with its high prevalence (Traynor et al., 2016), beekeepers must regularly intervene with various management practices to keep *Varroa* populations low. Most untreated colonies in temperate climates collapse from *Varroa* pressure in less than 3 years (Rosenkranz et al., 2010). Research on *Varroa* control methods are often restricted to chemical treatment applications, and most focus on establishing efficacy rates of various chemical regiments (Giacobino et al., 2015; Giovenazzo and Dubreuil, 2011; Jesus Gracia et al., 2017). The experts solicited in this study seemed to share this point of view with eight of the nine practices revolving around the use of one or more chemical products. Experts did however also ranked highly non-chemical regimens (inaptly named “IPM practices” in the survey). Those results also prompted us to dedicate a full report on the use of *Varroa* control methods (chemical and nonchemical) published separately (Haber et al., 2019).

On the other side of the spectrum, the 13 criteria that received significantly less than equal weight belonged to the domains “Equipment”, “Seasonal management” and “Non-*Varroa* control” (specifically relating to the use of 3 products for the control of small-hive beetles (*Aethina tumida*), Nosema (*Nosema* sp.) and tracheal mites (*Acarapis woodi*)).

As the weighted index did not represent a significant improvement on the non-weighted index, the optimized model did not end up reflecting the experts' a-priori ranking of criteria. Indeed, the ranking resulting from our sensitivity analyses did not correspond to the experts' ranking from weights. It is true that the experts were asked what they thought were the most important factors impacting loss, whereas the sensitivity analyses does take into consideration the current practices from which a change in behavior would be associated with a reduced risk of loss.

4.4. Encoding management information into criteria and criteria exclusion

There was high variability in the response rate across survey questions, and therefore, management criteria. This was partly due to the hierarchical structure of the questionnaire, with not all questions relevant to all respondents, but also because of missing answers or failure to provide acceptable response to one or more question – which is not unusual in observational surveys (Brick and Kalton, 1996). The advantages of using survey data is the relatively low cost of gathering observational data and its direct applicability to real world conditions as opposed to experimental setups. These types of questionnaires are frequently used in epidemiological, veterinary and agricultural studies, but make building and comparing traditional models difficult because of different response rates between parameters. This made traditional statistical modelling impractical because it relies on complete datasets. We used imputation to deal with missing answers, but not with “non-applicable” questions, which were dealt with by adjusting the denominator of the scoring. In our method, incomplete answers were still able to be used, as “non-applicable” questions did not count in the management score of the respondent.

4.5. Selection of the imputation and weighting method

The stability of the association with regard to the imputation method used indicates the handling of missing values did not impact the interpretation of the association. The effect of weighting was inconsistent, with no general improvement of index performance compared to the unweighted versions (Fig. 3), possibly due to the relatively low disparities between criteria weights (CW_i), as experts promoted few factors ahead of a cohort of similarly-low-weight criteria. We would recommend that future uses of this methodology insist on stronger expert discrimination between criteria.

4.6. General Management Index performance

The negative association between the index and Winter Loss indicates operations that reached a high GMI score also reported a lower level of colony mortality over the winter. This confirms that some variability in colony survivorship is associated with variability in management practices. Because this survey is observational in nature, a causal relationship cannot be drawn. This association, however, is based on real world observations in a wide array of environmental and societal conditions, which gives it practical meaning.

The index allows us to investigate the impact of management practices in a comprehensive manner. Various aspects of management have been previously associated with colony health outcomes in reference to a specific, individual health stressor. For example, nutritional supplementation can partially compensate for nutritional deficits in periods of dearth (Brodschneider and Crailsheim, 2010). Most notably, given the high impact of *Varroa*, it is believed most colonies in temperate climates would collapse within two to three years without periodic treatments to control it (Rosenkranz et al., 2010). Some aspects of management, such as disease control, have been identified as risk factors (Asensio et al., 2016; Chauzat et al., 2016; Jacques et al., 2017) for either high *Varroa* levels, or reduced treatment efficacy (Giacobino et al., 2016a, 2015). In a more recent study, various indicators of beekeeper type, background and experience were strong indicators of colony survivorship (Jacques et al., 2017). This is the first study to attempt to quantify the impact of management practices on colony mortality in a holistic way.

There was strong evidence of a curvature in this relationship between management index and loss (Fig. 4A, gam), indicating that the benefit of improved management practices would be stronger after surpassing a minimum threshold. This means that for very low scores of the index – with management practices far from the experts' recommendations – a lot of improvement is required before seeing a reduction in mortality risk. On the other hand, beekeepers starting with a medium index (which correspond to the 75th percentile of the population of our respondents) can expect a sharper reduction in risk with a small improvement in practices. The linear regression appears as a conservative estimate of the slope of the gam smooth for those median to high index respondents. This indicates that improvement of 0.1 in the management index within this model range is associated with a reduction in the risk of overwintering colony loss of 0.176 standard deviation. This represents a reduction of between 5.3 and 6.6 percentage point between the years of our survey, which is not trivial for a beekeeping operation. So even if the correlation value might seem weak, its effect size is in a range both meaningful and realistic.

4.7. Management Index simplification

The minimum adequate index over all respondents was composed of 21 criteria(Fig. 5f, Appendix D). The performance of the OMI was superior to the GMI, with a more manageable set of criteria. The investigation of the relationship for curvature also indicated a less marked threshold effect, which might be explained by the successful removal of the “noise” of management criteria for which a behavior change was not associated with risk reduction.

For the most part, the criteria composing the OMI did not correspond to the criteria favored by the experts. This indicates that experts were successful at identifying the “best” from the “worst” options for each particular management criterion (as verified by the good performance of the indices), but not at prioritizing criteria as to which behavior change would be most rewarding for the beekeepers. Two criteria were particularly misjudged by the experts: “Action on deadouts” and “States (count)” both received less weight from the experts than they would have under an equal weight distribution, but were high ranking in the sensitivity analysis and therefore retained in the OMI.

The 21 criteria composing the OMI represented five domains of management criteria (Equipment, Queens and New Colonies, Seasonal management, *Varroa* Control strategies and Non-*Varroa* Control strategies). The only domains completely removed from the Simplified Management Index were “Feeding” and “Monitoring”, though some of those practices were included in operation-type specific indices.

4.8. Comparison of OMI performance across operation types and regions

We expected to see differences between the practices recommended by our model between the four operation types contrasted. This is because their starting practices were already markedly different, which would then influence which adoption would be more beneficial for most beekeepers from that typology. The optimization process led us to different optimum indices for each of the four categories of beekeepers. The number of criteria retained ranged from 9 for southern small-scale beekeepers (Fig. 5g, Appendix D), to 25 for multi-state large-scale beekeepers (Fig. 5j, Appendix D). That the complexity of the index was higher for large-scale operations indicates that recommendations for behavior change were less generic than for small-scale operations for which a small number of behavior changes can be broadly recommended for the majority of the beekeepers. This reinforces the need for tailored practices for large scale beekeepers as previously mentioned.

Each of the operation-type specific OMI showed a significant association with standardized operation loss and represented an improvement on the performance of the General Index for that specific group. For each, the effect size of the associations were relatively low, but highly significant, with realistic and non-trivial reductions in risks of colony mortality.

There was however overlap among the practices identified as most significant by our model among the four beekeeper types. All 21 criteria from all respondents' OMI were accounted for in one or the other of the subset-specific OMI, with an additional 15 criteria that were not initially retained in the index generated for the whole respondent population but appeared when looking at the four subsets individually. Only a couple of criteria were retained in all four subset-specific OMI, and only a few more in most of the subsets. In addition, a comparison of the first five criteria in each of the operation-type specific OMI reveals that if certain management recommendations can be generalized across all beekeeper types, the largest fraction need to be operation-type specific.

For example, the criterion “*Varroa* Treatment (Y/N)”, a simple three level criterion addressing the use (or not) of a chemical product for the control of *Varroa*, ranked highly in all subsets of beekeepers except large-scale multi-state. This apparent contradiction can be explained by looking at the answer profile of beekeepers for this specific criteria: over 90% of large-scale multi-state beekeepers reported using a *Varroa* treatment product, compared to between 37 and 73% in the other three subsets of beekeepers (all years combined). As virtually all large-scale multi-state beekeepers reported the expert-identified “best” practice for that specific criterion, it is unsurprising that it did not register as one of the most sensitive components of their index, because it would not represent a large improvement of colony mortality risk for that population. Our method would therefore avoid the futile recommendation of a behavior already largely implemented in that group.

By contrast, the criterion identified with the second highest potential for large-scale multi-states operations was “*Varroa* monitoring technique”. The two most prevalent methods of *Varroa* monitoring for this group were visual inspection of bees (selected by 47.9% of large-scale multi-state beekeepers) and visual inspection of drone brood (selected by 57.5%). Though beekeepers were allowed to select multiple monitoring methods, it remains that a high proportion of those beekeepers rely, at least partly, on monitoring methods known to be highly unreliable (Honey Bee Health Coalition, 2015). Though this same sub-population exhibits a treatment regimen close to experts' recommendations, their choice of monitoring technique might imperil their abilities to detect damaging pests' levels, thereby leaving them unaware of potential reinfestations or treatment failures. The importance of *Varroa* monitoring after treatment had already been highlighted as a risk factor for elevated *Varroa* levels (Giacobino et al., 2014), though the method of monitoring was not specified.

The ranking of criteria should not be seen as identifying the most influential management practices on overwinter colony loss, but rather prioritizing criteria for which a change of behavior would be associated with the greatest reduction in the risk of overwinter loss at the population level. Our methodology opens the door to a systematic benchmarking of beekeepers, comparing their management practices to analogous operations. We have demonstrated the use of this methodology in four specific subsets of beekeepers based on type of operation and region, but with increasing participation, the level of refinement of the subsets could be improved. In particular, the subdivision into regions was limited to two levels, but could be implemented to more localized subgroups if the sample size allows it.

4.9. Future directions

The empirical recommendations identified in this work (in particular the top 4 practices identified for Northern small scale beekeepers) were subsequently put to the test in a multi-season field-trial from which the results will be published separately (Kulhanek et al., under review).

Though the effect size of our model appears meaningful enough to encourage efforts to improve management practices, they also realistically reminded us that, though management is one of the most actionable factors affecting colony health, it is not the only factor. The use of a comprehensive, yet simple, numeric metric to document and qualify management practices could be used in other studies to partition the relative importance of management with other sources of variation in honey bee colony mortality, such as environmental conditions, disease prevalence and loads, and pesticide exposure. Having a simple numeric index reflective of management quality could now be used to characterize and account for management practices in studies wanting to investigate other health stressors of bees in realistic field settings. This could help reduce the noise of field studies trying to associate specific stressors to colony-level health outcomes. It might also be interesting to investigate how those other factors might co-vary or constrain the effect of improving management practices, and therefore identify in which conditions management itself becomes a limiting factor of colony survivorship.

Our results can also be taken as a validation of the opinion of experts, as the scores of competing options for each criterion was based on their judgment. It should be noted that this validation of expert opinion is global but not punctual (criterion by criterion). A possible further study should test the influence of permutations of the criteria's levels to identify the optimal options for each criterion and confront those data-based optima with the opinion of experts. As this could be computationally intensive, the use of Augmented Intelligence might be appropriate.

5. Conclusion

Year after year, US beekeepers are experiencing high and recurring levels of losses (Kulhanek et al., 2017; Lee et al., 2015; Seitz et al., 2015; Spleen et al., 2013; Steinhauer et al., 2014; vanEngelsdorp et al., 2012; vanEngelsdorp et al., 2011; vanEngelsdorp et al., 2010; vanEngelsdorp et al., 2008; vanEngelsdorp et al., 2007). Those losses are concerning, not only for the sustainability of those beekeeping operations, but because of the high dependence of agricultural crops for honey bee pollination (Calderone, 2012), which links honey bee health to food security and diversity. Many factors are impacting honey bee health (Potts et al., 2010; Steinhauer et al., 2018; vanEngelsdorp and Meixner, 2010), some of which are effectively out of the hands of their beekeepers. However, others can be successfully prevented or remedied through beekeeping practices. This is the first study to report an association between the overall quality of management practices in beekeeping operations and colony loss. Previous research has addressed various aspects of management one at a time, mostly concerning pest control or feeding supplementation, but none had considered management in a comprehensive way. This type of holistic approach is essential to estimate the importance of management in relation to the other stressors known to impact honey bee health, and therefore the scale of potential improvement in colony loss that can be expected from improving management practices alone.

This paper can be of interest as a practical implementation of a methodology to build on expert knowledge to summarize a wide array of qualitative information in a simple metric with a measure of success (here, honey bee colony overwintering mortality). This methodology's particular strengths are its applicability to noisy and incomplete datasets that are typical in observational studies, but also its flexibility, as the complexity of the index structure could evolve to reflect the continuous improved understanding of the system.

Our results confirmed the association between the quality of management practices and success of colony overwintering. Another way to look at those results is as a validation of experts' opinions, by confronting their recommendations with real world observational data, transcending a wide array of environmental and societal conditions. Beekeepers who reported management practices of higher quality, according to the opinion of experts, were more likely to report lower mortality rates of colonies over the winter. This indicates that improvement in management can be, at least after a minimum quality level is reached, associated with increased odds of colonies surviving the winter. The expected improvement in colony survivorship are modest (5 to 6 percentage points for each 0.1 improvement in the index) but still practically meaningful enough to encourage efforts to improve management practice. A small number of practices were associated with the largest increase in colony survivorship for most beekeepers, but rankings varied by operation types. These results provide actionable evidence that can be used to design extension programs tailored to the target audience. Though management is only one of the factors impacting colony loss, it is almost entirely under the control of beekeepers. Improving management will not prevent all colony losses; however, our results indicate a non-negligible reduction in risk that could alleviate stress on both bees and beekeepers.

CRediT authorship contribution statement

Nathalie Steinhauer: Methodology, Software, Investigation, Data Curation, Validation, Formal analysis, Visualization, Writing – Original Draft. **Dennis vanEngelsdorp**: Conceptualization, Writing – Review & Editing, Supervision, Funding acquisition. **Claude Saegerman**: Conceptualization, Methodology, Writing – Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project was funded by a Coordinated Agricultural Product (CAP) grant (N° 2011-67007-30355) from US Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) and the Bee Informed Partnership Inc. We thank the reviewers for their insightful comments. We thank all respondent beekeepers for their participation in our surveys. We thank the University of Maryland bee lab personnel and undergraduates students for their help entering the answers from paper surveys. We thank Karen Rennich and Jeri Parrent for their help in editing the manuscript. We thank Marla Spivak, David Tarpy, Ramesh Sagili, James Wilkes, Jonathan Engelsma, and Bee Informed Partnership, Inc. for obtaining the data supporting the results of this manuscript. We thank Mikayla Wilson for her work in software development and data curation. Our gratitude goes out to the many beekeeping organizations, industry leaders, and beekeeping clubs that forwarded our appeal for participation. A special thank you is owed to USDA APHIS, the Apiary Inspectors of America, the American Honey Producers Association, Eastern Apiculture Society, the American Beekeeping Federation, Brushy Mountain Bee Farm, Bee Culture magazine, Project Apis m. and American Bee Journal for sending out participation requests to their online audiences. We thank in particular our experts who contributed to the scoring of the management practices: Dewey Caron, Wayne Esaias, Jerry Hayes, Eugene Lengerich, Katie Lee, Megan Mahoney, Jeff Pettis, Ben Sallman, Rob Snyder, Marla Spivak, Liana Tiegen, Ellen Topitshofer, James Wilkes and Dan Wyns. I would also like to thank the members of my thesis committee, Dr. David Hawthorne, Dr. William Lamp and the departed and greatly missed Dr. Bahram Momen.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.141629.

References

Alaux, C., Ducloz, F., Crauser, D., Le Conte, Y., 2010. Diet effects on honeybee immunocompetence. Biol. Lett. 6 (4), 562–565 [Online]. Available from. https://doi.org/10.1098/rsbl.2009.0986.

Antúnez, K., Invernizzi, C., Mendoza, Y., vanEngelsdorp, D., et al., 2017. Honeybee colony losses in Uruguay during 2013–2014. Apidologie 48 (3), 364–370 [Online]. Available from. https://doi.org/10.1007/s13592-016-0482-2.

Asensio, I., Vicente-Rubiano, M., Muñoz, M.J., Fernández-Carrión, E., et al., 2016. Importance of ecological factors and colony handling for optimizing health status of apiaries in Mediterranean ecosystems. PLoS One 11 (10), e0164205 [Online]. Available from:. https://doi.org/10.1371/journal.pone.0164205.

Bee Informed Partnership (n.d.) Loss & management survey – bee informed partnership. [Online]. Available from: https://beeinformed.org/citizen-science/loss-and-management-survey/ [Accessed: 27 November 2019].

Berry, J.A., Delaplane, K.S., 2001. Effects of comb age on honey bee colony growth and brood survivorship. J. Apic. Res. 40 (1), 3–8 [Online]. Available from. https://doi.org/10.1080/00218839.2001.11101042.

Bixby, M., Baylis, K., Hoover, S.E., Currie, R.W., et al., 2017. A bio-economic case study of Canadian honey bee (Hymenoptera: Apidae) colonies: marker-assisted selection (MAS) in queen breeding affects beekeeper profits. J. Econ. Entomol. 110 (3), 816–825 [Online]. Available from. https://doi.org/10.1093/jee/tox077.

Branchiccela, B., Castelli, L., Corona, M., Díaz-Cetti, S., et al., 2019. Impact of nutritional stress on the honeybee colony health. Sci. Rep. 9 (1), 10156 [Online]. Available from. https://doi.org/10.1038/s41598-019-46453-9.

Brick, J.M., Kalton, G., 1996. Handling missing data in survey research. Stat. Methods Med. Res. 5 (3), 215–238 [Online]. Available from. https://doi.org/10.1177/096228029600500302.

Brodschneider, R., Crailsheim, K., 2010. Nutrition and health in honey bees. Apidologie 41 (3), 278–294 [Online]. Available from. https://doi.org/10.1051/apido/2010012.

Calatayud-Vernich, P., Calatayud, F., Simó, E., Picó, Y., 2018. Pesticide residues in honey bees, pollen and beeswax: assessing beehive exposure. Environ. Pollut. 241, 106–114 [Online]. Available from. https://doi.org/10.1016/j.envpol.2018.05.062.

Calderone, N.W., 2012. Insect pollinated crops, insect pollinators and US agriculture: trend analysis of aggregate data for the period 1992-2009. PLoS One 7 (5). https://doi.org/10.1371/journal.pone.0037235 [Online]. Available from.

Canty, A., Ripley, B., 2015. Boot: Bootstrap R (S-Plus) Functions.

Castilhos, D., Bergamo, G.C., Gramacho, K.P. & Gonçalves, L.S. (2019) Bee colony losses in Brazil: a 5-year online survey. Apidologie. [Online] Available from: doi:https://doi.org/10.1007/s13592-019-00642-7 [Accessed: 18 June 2019].

Chauzat, M.-P., Jacques, A., Consortium, E, Laurent, M., et al., 2016. Risk indicators affecting honeybee colony survival in Europe: one year of surveillance. Apidologie 47 (3), 348–378 [Online]. Available from. https://doi.org/10.1007/s13592-016-0440-z.

Crawley, M.J., 2007. Generalized linear models. The R Book. John Wiley & Sons, Ltd , pp. 511–526 [Online]. Available from. http://onlinelibrary.wiley.com/doi/10.1002/9780470515075.ch13/summary. (Accessed 15 April 2014). .

Cronin, G.M., Rault, J.-L., Glatz, P.C., 2014. Lessons learned from past experience with intensive livestock management systems. Revue Scientifique Et Technique-Office International Des Epizooties 33 (1), 139–151.

Dainat, B., Evans, J.D., Chen, Y.P., Gauthier, L., et al., 2012. Dead or alive: deformed wing virus and Varroa destructor reduce the life span of winter honeybees. Appl. Environ. Microbiol. 78 (4), 981–987 [Online]. Available from. https://doi.org/10.1128/AEM.06537-11.

Di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L.P., et al., 2013. Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? PLoS One 8 (8), e72016 [Online]. Available from. https://doi.org/10.1371/journal.pone.0072016.

El Agrebi, N., Tosi, S., Wilmart, O., Scippo, M.-L., et al., 2020a. Honeybee and consumer's exposure and risk characterisation to glyphosate-based herbicide (GBH) and its degradation product (AMPA): residues in beebread, wax, and honey. Sci. Total Environ. 704, 135312 [Online]. Available from. https://doi.org/10.1016/j.scitotenv.2019.135312.

El Agrebi, N., Traynor, K., Wilmart, O., Tosi, S., et al., 2020b. Pesticide and veterinary drug residues in Belgian beeswax: occurrence, toxicity, and risk to honey bees. Sci. Total Environ. 745, 141036 [Online]. Available from. https://doi.org/10.1016/j.scitotenv.2020.141036.

Evans, J.D., Spivak, M., 2010. Socialized medicine: individual and communal disease barriers in honey bees. J. Invertebr. Pathol. 103, S62–S72 [Online]. Available from. https://doi.org/10.1016/j.jip.2009.06.019.

Food and Agriculture Organization of the United Nations (FAO), S.D. (ESS), 2018. FAOSTAT: Expanded and FAO Commodity List (FCL): Beehives (1961–2016). [Online]. , p. 2018 Available from. http://www.fao.org/faostat/en/. (Accessed 5 September 2018). .

Genersch, E., 2010. Honey bee pathology: current threats to honey bees and beekeeping. Appl. Microbiol. Biotechnol. 87 (1), 87–97 [Online]. Available from. https://doi.org/10.1007/s00253-010-2573-8.

Giacobino, A., Cagnolo, N.B., Merke, J., Orellano, E., et al., 2014. Risk factors associated with the presence of Varroa destructor in honey bee colonies from east-central Argentina. Preventive Veterinary Medicine. 115 (3–4), 280–287 [Online]. Available from:. https://doi.org/10.1016/j.prevetmed.2014.04.002.

Giacobino, A., Molineri, A., Cagnolo, N.B., Merke, J., et al., 2015. Risk factors associated with failures of Varroa treatments in honey bee colonies without broodless period. Apidologie 46 (5), 573–582 [Online]. Available from. https://doi.org/10.1007/s13592-015-0347-0.

Giacobino, A., Molineri, A., Bulacio Cagnolo, N., Merke, J., et al., 2016a. Key management practices to prevent high infestation levels of Varroa destructor in honey bee colonies at the beginning of the honey yield season. Preventive Veterinary Medicine 131, 95–102 [Online]. Available from. https://doi.org/10.1016/j.prevetmed.2016.07.013.

Giacobino, A., Molineri, A., Cagnolo, N.B., Merke, J., et al., 2016b. Queen replacement: the key to prevent winter colony losses in Argentina. J. Apic. Res. 55 (4), 335–341 [Online]. Available from. https://doi.org/10.1080/00218839.2016.1238595.

Giacobino, A., Pacini, A., Molineri, A., Rodríguez, G., et al., 2018. Potential associations between the mite Varroa destructor and other stressors in honeybee colonies (Apis mellifera L.) in temperate and subtropical climate from Argentina. Preventive Veterinary Medicine 159, 143–152 [Online]. Available from. https://doi.org/10.1016/j.prevetmed.2018.09.011.

Giovenazzo, P., Dubreuil, P., 2011. Evaluation of spring organic treatments against Varroa destructor (Acari: Varroidae) in honey bee Apis mellifera (Hymenoptera: Apidae) colonies in eastern Canada. Exp. Appl. Acarol. 55 (1), 65–76 [Online]. Available from. https://doi.org/10.1007/s10493-011-9447-3.

Haber, A.I., Steinhauer, N.A., vanEngelsdorp, D., 2019. Use of chemical and nonchemical methods for the control of Varroa destructor (Acari: Varroidae) and associated winter Colony losses in U.S. beekeeping operations. J. Econ. Entomol. https://doi.org/10.1093/jee/toz088 [Online] Available from.

Honey Bee Health Coalition, 2015. Tools for Varroa management - a guide to effective varroa sampling and control. [Online]. Available from. http://honeybeehealthcoalition.org/wp-content/uploads/2015/08/HBHC-Guide_Varroa-Interactive-PDF.pdf. (Accessed 14 August 2015). .

Humblet, M.-F., Vandeputte, S., Albert, A., Gosset, C., et al., 2012. Multidisciplinary and evidence-based method for prioritizing diseases of food-producing animals and zoonoses. Emerg. Infect. Dis. 18 (4). https://doi.org/10.3201/eid1804.111151 [Online]. Available from.

Huneau-Salauen, A., Bougeard, S., Balaine, L., Eono, F., et al., 2015. Husbandry factors and health conditions influencing the productivity of French rabbit farms. World Rabbit Sci. 23 (1), 27–37 [Online]. Available from. https://doi.org/10.4995/wrs.2015.3076.

Jacques, A., Laurent, M., Consortium, E, Ribière-Chabert, M., et al., 2017. A pan-European epidemiological study reveals honey bee colony survival depends on beekeeper education and disease control. PLoS One 12 (3), e0172591 [Online]. Available from. https://doi.org/10.1371/journal.pone.0172591.

Jesus Gracia, M., Moreno, C., Ferrer, M., Sanz, A., et al., 2017. Field efficacy of acaricides against Varroa destructor. PLoS One 12 (2), e0171633 Online. Available from. https://doi.org/10.1371/journal.pone.0171633.

Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS One 8 (1), e54092 [Online]. Available from. https://doi.org/10.1371/journal.pone.0054092.

Karl, T., Koss, W.J., 1984. Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895–1983. Historical Climatology Series 4-3National Climatic Data Center, Asheville, NC.

Kevill, J.L., de Souza, F.S., Sharples, C., Oliver, R., et al. (2019) DWV-A lethal to honey bees (Apis mellifera): a colony level survey of DWV variants (A, B, and C) in England, Wales, and 32 states across the US. *Viruses*. [Online] vol. 11 (5). Available from: doi: https://doi.org/10.3390/v11050426 [Accessed: 24 June 2019].

Kulhanek, K., VanEngelsdorp, D., 2017. Le petit coléoptère des ruches (Aethina tumida) aux États-Unis, parasite ravageur des colonies d'Apis mellifera. Bulletin Epidemiologique 81 (8), 5 Numero special abeilles.

Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D.M., et al., 2017. A National Survey of Managed Honey Bee 2015–2016 Annual Colony Losses in the USA. vol. 56 pp. 328–340.

Kulhanek, K., Steinhauer, N., Wilkes, J., Wilson, M., Spivak, M., Sagili, R., Tarpy, D.R., et al. Survey-derived best beekeeping management practices improve colony health and reduce mortality. PLOS ONE under review (n.d.).

Lee, K.V., Steinhauer, N., Rennich, K., Wilson, M.E., et al., 2015. A national survey of managed honey bee 2013–2014 annual colony losses in the USA. Apidologie, 1–14 [Online]. Available from. https://doi.org/10.1007/s13592-015-0356-z.

Lounsberry, Z., Spiewok, S., Pernal, S.F., Sonstegard, T.S., et al., 2010. Worldwide diaspora of Aethina tumida (Coleoptera: Nitidulidae), a nest parasite of honey bees. Ann. Entomol. Soc. Am. 103 (4), 671–677 [Online]. Available from. https://doi.org/10.1603/AN10027.

Maucourt, S., Fournier, V., Giovenazzo, P., 2018. Comparison of three methods to multiply honey bee (Apis mellifera) colonies. Apidologie 49 (3), 314–324 [Online]. Available from. https://doi.org/10.1007/s13592-017-0556-9.

Mitchell, E.S.E., Smith, R.P., Ellis-Iversen, J., 2012. Husbandry risk factors associated with subclinical coccidiosis in young cattle. Vet. J. 193 (1), 119–123 [Online]. Available from. https://doi.org/10.1016/j.tvjl.2011.09.017.

Molineri, A., Giacobino, A., Pacini, A., Bulacio Cagnolo, N., et al., 2017. Risk factors for the presence of deformed wing virus and acute bee paralysis virus under temperate and subtropical climate in Argentinian bee colonies. Preventive Veterinary Medicine 140, 106–115 [Online]. Available from. https://doi.org/10.1016/j.prevetmed.2017.02.019.

Mortensen, A.N., Jack, C.J., Bustamante, T.A., Schmehl, D.R., et al., 2019. Effects of supplemental pollen feeding on honey bee (Hymenoptera: Apidae) Colony strength and Nosema spp. infection. J. Econ. Entomol. 112 (1), 60–66 [Online]. Available from. https://doi.org/10.1093/jee/toy341.

Neumann, P., Carreck, N.L., 2010. Honey bee colony losses. J. Apic. Res. 49 (1), 1–6 Online. Available from. https://doi.org/10.3896/IBRA.1.49.1.01.

Oberreiter, H., Brodschneider, R., 2020. Austrian COLOSS survey of honey bee colony winter losses 2018/19 and analysis of hive management practices. Diversity 12 (3), 99 [Online]. Available from. https://doi.org/10.3390/d12030099.

Oldroyd, B.P., 2012. Domestication of honey bees was associated with expansion of genetic diversity. Mol. Ecol. 21 (18), 4409–4411 [Online]. Available from. https://doi.org/10.1111/j.1365-294X.2012.05641.x.

Payne, A.N., Walsh, E.M., Rangel, J., 2019. Initial exposure of wax foundation to agrochemicals causes negligible effects on the growth and winter survival of incipient honey bee (Apis mellifera) colonies. Insects 10 (1), 19 Online. Available from. https://doi.org/10.3390/insects10010019.

Pirk, C.W.W., Human, H., Crewe, R.M., vanEngelsdorp, D., 2014. A survey of managed honey bee colony losses in the Republic of South Africa - 2009 to 2011. J. Apic. Res. 53 (1), 35–42 [Online]. Available from. https://doi.org/10.3896/IBRA.1.53.1.03.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., et al., 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol. Evol. 25 (6), 345–353 [Online]. Available from. https://doi.org/10.1016/j.tree.2010.01.007.

R Core Team (2017) R: A Language and Environment for Statistical Computing. [Online]. Vienna, Austria, R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Rinkevich, F.D., 2020. Detection of amitraz resistance and reduced treatment efficacy in the Varroa mite, Varroa destructor, within commercial beekeeping operations. PLoS One 15 (1), e0227264 [Online]. Available from. https://doi.org/10.1371/journal.pone.0227264.

Rinkevich, F.D., Danka, R.G., Healy, K.B., 2017. Influence of Varroa mite (Varroa destructor) management practices on insecticide sensitivity in the honey bee (Apis mellifera). Insects 8 (1), 9 [Online]. Available from. https://doi.org/10.3390/insects8010009.

Rosenkranz, P., Aumeier, P., Ziegelmann, B., 2010. Biology and control of Varroa destructor. J. Invertebr. Pathol. 103 (Supplement), S96–S119 [Online]. Available from. https://doi.org/10.1016/j.jip.2009.07.016.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2006. Sensitivity analysis practices: strategies for model-based inference. Reliability Engineering & System Safety. 91 (10−11), 1109–1125 Online. Available from. https://doi.org/10.1016/j.ress.2005.11.014.

Seitz, N., Traynor, K.S., Steinhauer, N., Rennich, K., et al., 2015. A national survey of managed honey bee 2014–2015 annual colony losses in the USA. J. Apic. Res. 54 (4), 292–304 [Online]. Available from. https://doi.org/10.1080/00218839.2016.1153294.

Sperandio, G., Simonetto, A., Carnesecchi, E., Costa, C., et al., 2019. Beekeeping and honey bee colony health: a review and conceptualization of beekeeping management practices implemented in Europe. Sci. Total Environ. 696, 133795 Online. Available from. https://doi.org/10.1016/j.scitotenv.2019.133795.

Spleen, A.M., Lengerich, E.J., Rennich, K., Caron, D., et al., 2013. A national survey of managed honey bee 2011-12 winter colony losses in the United States: results from the Bee Informed Partnership. J. Apic. Res. 52 (2), 44–53 Online. Available from. https://doi.org/10.3896/IBRA.1.52.2.07.

Steinhauer, N.A., Rennich, K., Wilson, M.E., Caron, D.M., et al., 2014. A national survey of managed honey bee 2012-2013 annual colony losses in the USA: results from the Bee Informed Partnership. J. Apic. Res. 53 (1), 1–18 [Online]. Available from. https://doi.org/10.3896/IBRA.1.53.1.01.

Steinhauer, N., Kulhanek, K., Antúnez, K., Human, H., et al., 2018. Drivers of colony losses. Current Opinion in Insect Science 26, 142–148 [Online]. Available from. https://doi.org/10.1016/j.cois.2018.02.004.

Straub, L., Williams, G.R., Vidondo, B., Khongphinitbunjong, K., et al., 2019. Neonicotinoids and ectoparasitic mites synergistically impact honeybees. Sci. Rep. 9, 8159 [Online]. Available from. https://doi.org/10.1038/s41598-019-44207-1.

The Pollinator Health Task Force, Vilsack, T., McCarthy, G., 2015. National strategy to promote the health of honey bees and other pollinators. [Online]. Available from. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/Pollinator%20Health%20Strategy%202015.pdf. (Accessed 3 August 2017). .

Thoms, C.A., Nelson, K.C., Kubas, A., Steinhauer, N., et al., 2018. Beekeeper stewardship, colony loss, and Varroa destructor management. Ambio https://doi.org/10.1007/s13280-018-1130-z [Online] Available from. (Accessed: 26 November 2018).

Tosi, S., Nieh, J.C., Sgolastra, F., Cabbri, R., et al., 2017. Neonicotinoid pesticides and nutritional stress synergistically reduce survival in honey bees. Proc. R. Soc. B 284 (1869), 20171711 [Online]. Available from. https://doi.org/10.1098/rspb.2017.1711.

Traynor, K.S., Rennich, K., Forsgren, E., Rose, R., et al., 2016. Multiyear survey targeting disease incidence in US honey bees. Apidologie 47 (3), 325–347 [Online]. Available from:. https://doi.org/10.1007/s13592-016-0431-0.

Underwood, R.M., Currie, R.W., 2003. The effects of temperature and dose of formic acid on treatment efficacy against Varroa destructor (Acari: Varroidae), a parasite of Apis mellifera (Hymenoptera: Apidae). Experimental & Applied Acarology 29 (3), 303 [Online]. Available from. https://doi.org/10.1023/A:1025892906393.

Underwood, R.M., Traver, B.E., López-Uribe, M.M., 2019. Beekeeping management practices are associated with operation size and Beekeepers' philosophy towards in-hive chemicals. Insects 10 (1), 10 [Online]. Available from. https://doi.org/10.3390/insects10010010.

US EPA, 2017. Receipt of application for notificiant under Pesticide Registration Notice: Product Name: Mite-Away Quick Strips. [Online]. Available from. https://www3.epa.gov/pesticides/chem_search/ppls/075710-00002-20170217.pdf.

van der Zee, R., Pisa, L., Andonov, S., Brodschneider, R., et al., 2012. Managed honey bee colony losses in Canada, China, Europe, Israel and Turkey, for the winters of 2008-9 and 2009-10. J. Apic. Res. 51 (1), 91–114 [Online]. Available from. https://doi.org/10.3896/IBRA.1.51.1.12.

van der Zee, R., Brodschneider, R., Brusbardis, V., Charriere, J.-D., et al., 2014. Results of international standardised beekeeper surveys of colony losses for winter 2012-2013: analysis of winter loss rates and mixed effects modelling of risk factors for winter loss. J. Apic. Res. 53 (1), 19–34 [Online]. Available from. https://doi.org/10.3896/IBRA.1.53.1.02.

van Dooremalen, C., Cornelissen, B., Poleij-Hok-Ahin, C., Blacquiere, T., 2018. Single and interactive effects of Varroa destructor, Nosema spp., and imidacloprid on honey bee colonies (Apis mellifera). Ecosphere 9 (8), e02378 [Online]. Available from. https://doi.org/10.1002/ecs2.2378.

vanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95 [Online]. Available from. https://doi.org/10.1016/j.jip.2009.06.011.

vanEngelsdorp, D., Underwood, R., Caron, D., Hayes, J., 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 (7), 599–603.

vanEngelsdorp, D., Hayes, J., Underwood, R.M., Pettis, J., 2008. A survey of honey bee colony losses in the U.S., Fall 2007 to Spring 2008. PLoS One 3 (12). https://doi.org/10.1371/journal.pone.0004071 Online. Available from. (Accessed: 29 September 2013).

vanEngelsdorp, D., Hayes, J., Underwood, R.M., Pettis, J.S., 2010. A survey of honey bee colony losses in the United States, fall 2008 to spring 2009. J. Apic. Res. 49 (1), 7–14 [Online]. Available from. https://doi.org/10.3896/IBRA.1.49.1.03.

vanEngelsdorp, D., Hayes, J., Underwood, R.M., Caron, D., et al., 2011. A survey of managed honey bee colony losses in the USA, fall 2009 to winter 2010. J. Apic. Res. 50 (1), 1–10 [Online]. Available from. https://doi.org/10.3896/IBRA.1.50.1.01.

vanEngelsdorp, D., Caron, D., Hayes, J., Underwood, R., et al., 2012. A national survey of managed honey bee 2010-11 winter colony losses in the USA: results from the Bee Informed Partnership. J. Apic. Res. 51 (1), 115–124 [Online]. Available from:. https://doi.org/10.3896/IBRA.1.51.1.14.

Wagoner, K.M., Spivak, M., Rueppell, O., 2018. Brood affects hygienic behavior in the honey bee (Hymenoptera: Apidae). J. Econ. Entomol. 111 (6), 2520–2530 [Online]. Available from. https://doi.org/10.1093/jee/toy266.

Wilfert, L., Long, G., Leggett, H.C., Schmid-Hempel, P., et al., 2016. Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. Science 351 (6273), 594–597 Online. Available from. https://doi.org/10.1126/science.aac9976.