Short Note

Towards the development of an index for the holistic assessment of the health status of a honey bee colony

Gianni Gilioli\textsuperscript{a}, Giorgio Sperandio\textsuperscript{a,b}, Fani Hatjina\textsuperscript{c}, Anna Simonetto\textsuperscript{a,b,c,1}

\textsuperscript{a} Dep. of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
\textsuperscript{b} Dep. of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy
\textsuperscript{c} Division of Apiculture, Hellenic Agricultural Organisation DEMETER, Nea Moudania, Greece

\textbf{A B S T R A C T}

Honey bees play an important role in the maintenance of both, biodiversity and food security through pollination services and also represent an important source of income for rural areas. Despite several studies and monitoring projects gathering a large amount of data on the main factors/stressors influencing honey bee colonies, there is a lack of holistic and multidimensional statistical tools integrating different aspects which define honey bee colony health. Such tools are important to support the sustainable management of honey bees. In this study, we designed and tested a methodological framework based on Structural Equation Models for the development of a honey bee Health Status Index. The index accounts for the main abiotic (e.g. pesticide contamination, landscape characteristics) and biotic (e.g. parasitosis, virosis) factors influencing honey bee health, including the beekeepers role in managing the colony. The proposed methodology was validated against a dataset representing seven scenarios generated through Expert Knowledge Elicitation. The validation procedure showed that the Health Status Index integrates different sources of data and it can quantify the health status of a honey bee colony based on the characteristics of each scenario. Furthermore, it allows the investigation of the influence of different drivers/stressors on the health of the honey bee colony. The Health Status Index provides flexibility in the selection of variables, making it a valuable tool for holistic and harmonised assessment of honey bee health. Once validated, the index can support the evaluation at different spatial (from local- to area-wide management) and temporal (medium- to long-term management) scales to support stakeholders’ (e.g. beekeepers, risk assessors, risk managers) decision-making.

1. Introduction

Honey bees play a key role in the maintenance of important ecosystems services, such as pollination of cultivated (Bommarco et al., 2012; Bradbear, 2009; Klein et al., 2007) and wild plants (Aguilera et al., 2006; Ashman et al., 2004). Both of which are important for ensuring food security (Bommarco et al., 2013). In addition, honey bees represent an important source of income for beekeepers (Formato and Smulders, 2011; Mizrahi and Lenaky, 2013; Schmidt, 1997) contributing to an increase in economic value of rural areas (Deloitte, 2013). In the last few decades managed honey bees have faced a widespread decline (Brodschneider et al., 2016; Genersch et al., 2010; Jacques et al., 2017; Lee et al., 2015; Porrini et al., 2016; Steinbauer et al., 2014; VanEngelsdorp et al., 2011) raising concerns for the sustainability of the beekeeping sector (Potts et al., 2010). Recent research has shown the importance of considering the co-occurrence and interaction of different drivers/stressors when assessing the health of the honey bee colony (Doublet et al., 2015; Goulson et al., 2015; Sánchez-Bayo et al., 2016). However, laboratory and field studies highlight some limitations in understanding the combined influence of factors on honey bee colony health (Becher et al., 2013). Therefore, there is an expressed need for new methodological tools capable of analysing the impact of co-occurring factors on honey bee health. These tools if properly developed and tested could contribute to holistically assess the impacts caused by multiple stressors on honey bees (Goulson et al., 2015). Furthermore, they support stakeholders’ (beekeepers, risk managers etc.) decision-making (Rortais et al., 2017). Honey bee colony health is a complex, dynamic and multidimensional property. It is the result of the interaction between the colony demography, energetics, temporal and spatial pattern of environmental resources availability in the landscape. Honey bee
colony health is also influenced by the population dynamics, epidemiology of pests and diseases, the level of contamination in the environmental matrices, and also beekeeping practices. As an emerging property (EFSA AHAW Panel, 2016), honey bee colony health can be considered, from a statistical point of view, as a latent variable (Spearman, 1904) since it cannot be directly measured and it has to be estimated from one or more observable variables (called indicators). Structural Equation Modelling (SEM) (Bollen, 1989) is the most widely used methodology for the analysis of latent variables, as it assesses the relation between latent variables and their indicators to summarise them into a composite index. By applying SEM it is possible to: (i) take into account, as indicators, different types of variables (e.g. continuous, ordinal, nominal, dichotomic), (ii) analyse the causal relationships existing between variables (both latent and indicator variables), (iii) examine whether the relationship assumed a priori is consistent with empirical data and (iv) provide estimates of latent variables that can be used in predictive models. Initially used within the social sciences (Guo et al., 2008), SEM-based indexes are currently widely used in various research fields. Significant examples of application concern ecological (Arhonditsis et al., 2006; Iriondo et al., 2003; Villeneuve et al., 2018) and human health related issues (Boniface and Tefft, 1997; Cheung and Hong, 2017; Eldridge et al., 2017).

The aim of our paper was to develop and test a methodological framework based on SEM, for both, the definition and quantification of the health status of managed honey bee colonies (Health Status Index, HSI), taking into account the influence of biotic and abiotic drivers/stressors. The conceptual framework at the basis of the HSI was tested using a dataset generated through a semiformal Expert Knowledge Elicitation (EKE) procedure (EFSA, 2014, 2018).

### 2. Materials and methods

#### 2.1 Conceptual framework

We developed a hierarchical conceptual framework for the definition of the health status of a honey bee colony based on the work carried out by EFSA AHAW Panel (2016). Two overarching constructs, the “external drivers” influencing the honey bee colony and the “colony”, are used to define the conceptual framework. A set of attributes that are not directly measurable, (namely latent variables) defines each construct. Each attribute is composed by a set of indicators (i.e. variables) that could be directly measured. The selection of the most relevant indicators was performed taking into account their biological relevance in relation to honey bee colony health and the SEM requirements, (e.g. unidimensionality and internal consistency) (Simonetto, 2012). The external drivers construct is composed by three attributes: resource providing unit (RPU), environmental drivers (ENV) and beekeeping management practices (BMP). The colony construct is composed by five attributes: queen (QUE), in-hive products (IHP), contamination (CON), disease-infection-infestation (DII), demography (DEM) and behaviour (BEH). The conceptual framework adopted in this study and a detailed description of the attributes is presented in Table 1. A throughout description of the selected indicators can be found in Appendix A (Section A.1). We hypothesised that the external drivers construct has a direct influence on the colony construct. Then, the HSI is an opportune synthesis of the colony construct providing a quantification of the health status of a honey bee colony (Fig. 1).

#### Table 1

Hierarchical approach followed for the definition of the conceptual framework.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Attributes</th>
<th>Attributes description</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource providing unit (RPU)</td>
<td>Abiotic factors (i.e. land cover, land use, nectar and pollen availability) characterising the landscape within a 3 km radius around the hive, responsible for the provision and regulation of the resources for a honey bee colony</td>
<td>Land cover/use</td>
<td></td>
</tr>
<tr>
<td>Environmental drivers (ENV)</td>
<td>Abiotic factors related to weather and climate influencing the demography, the physiology and the behaviour of a honey bee colony</td>
<td>Climate type, Relative humidity, Total precipitation, Minimum temperature, Maximum temperature</td>
<td></td>
</tr>
<tr>
<td>Beekeeping management practices (BMP)</td>
<td>Factors related to beekeeper’s category (non-professional or professional), experience and production system adopted (e.g. conventional or organic)</td>
<td>Beekeeper category, Beekeeper experience, Organic treatment</td>
<td></td>
</tr>
<tr>
<td>Queen (QUE)</td>
<td>Factors related to the presence, age and performances of the honey bee queen</td>
<td>Presence, Potential fecundity, Youth, Natural queen replacement</td>
<td></td>
</tr>
<tr>
<td>In-hive products (IHP)</td>
<td>Factors referring to the amount of honey and bee bread present in the honey bee colony</td>
<td>Honey, Bee bread</td>
<td></td>
</tr>
<tr>
<td>Contamination (CON)</td>
<td>Factors referring to the level of contamination of honey, bee bread and wax in the honey bee colony</td>
<td>Honey quality, Bee bread quality, Wax quality</td>
<td></td>
</tr>
<tr>
<td>Disease-infection-infestation (DII)</td>
<td>Factors referring to the presence/abundance of parasites, viruses and bacteria affecting a honey bee colony and the occurrence of clinical signs attributable to a diseased colony</td>
<td>Varroa, American foulbrood, Nosema, Deformed Wing Virus (DWV), Other clinical signs</td>
<td></td>
</tr>
<tr>
<td>Demography (DEM)</td>
<td>Factors referring to the amount of brood, adults and dead bees in the colony</td>
<td>Dead bees, Brood, Adult</td>
<td></td>
</tr>
<tr>
<td>Behaviour (BEH)</td>
<td>Factors referring to the behavioural traits of the honey bee colony</td>
<td>Atypical behaviour, Foraging disruption</td>
<td></td>
</tr>
</tbody>
</table>
is a linear combination of the other attributes. \( \theta \) is the vector of path coefficient (also called loadings). The loading \( \lambda_{ij} \) represents the regression coefficient between the \( r \)-th indicator and the \( j \)-th latent variable. \( \varepsilon \) is the measurement error.

The HSI outer model (Fig. 3) is based on the indicators mentioned in Table 1. The HSI latent variable is considered a second-order construct since it involves more than one dimension (i.e. the six attributes referring to the colony construct) without any specific indicator. To model the second-order construct HSI we adopted the strategy proposed by Sanchez (2013) attributing to HSI the 19 indicators related to the colony construct.

2.3. Data generation

The generation of datasets to estimate the parameters of the HSI required the following: (i) the development of realistic scenarios, related to the external drivers construct and the indicators referring to the DII attribute, (ii) the estimation of uncertainty distributions of the other indicators for each scenario and (iii) the generation of observations based on random sampling from estimated probability distributions. A scenario is defined by the set of values assumed by the indicators referring to RPU, ENV, BMP and DII. The scenarios were used to provide the expert realistic contexts from which starting the EKE procedure. Firstly, we chose five geographical areas in Greece considered suitable for beekeeping activities. For each area, we defined the prevalent climate type according to Köppen-Geiger classification (Kottek et al., 2006), the season (month) and the prevalent land cover/use according to the first level of CORINE Land Cover classification (EEA, 2014). The availability of nectar and pollen in the RPU and the characteristics of the beekeeper (experience, category and adherence to organic production system) were selected to cover a wide range of beekeeping conditions. Finally, we defined the sanitary conditions of the colony (DII indicators). Seven scenarios representing realistic beekeeping conditions in different areas of Greece were developed (Appendix B, Section A.2). Two scenarios were related to agricultural areas (AGR1 and AGR2) while five scenarios were related to forests and semi-natural areas (FOR1, FOR2, FOR3, FOR4 and FOR5). Particular attention was paid to represent a gradient of possible beekeeping situations. Thus, scenarios were diversified by:

- The availability of pollen and nectar in the RPU which is high in FOR1, FOR3 and FOR5 and low in AGR1;
- The experience of beekeepers, considered high in AGR1, AGR2, FOR2 and FOR3 and low in FOR1 and FOR5;
- The sanitary conditions of the colony, which are considered good in FOR3 and poor in AGR2, FOR2 and FOR4.

Based on these seven scenarios, we generated a dataset through a semiformal EKE approach to eliciting probability distributions, based on the Sheffield method (EFSA, 2014, 2018). The EKE protocol, the specific questions to be elicited, their metrics, scales and units were carefully planned according to the conceptual framework presented in Table 1. The EKE procedure involved one expert (with knowledge on honey bees and beekeeping) and one elicitor (with expertise in performing EKE). A specific training was provided to the expert about the EKE methodology. The expert was asked to provide the uncertainty distributions related to the indicators of QUE, HIP, CON, DEM and BEH. The expert also provided the probability of occurrence of each scenario. To generate the simulated dataset, we used an iterative procedure. We extracted 1000 random numbers from a uniform distribution \([0, 1]\) and we used the inverse function of each probability distribution provided by the expert to obtain the values of the corresponding quantiles. Each
The HSI model was fitted to the data generated through the EKE procedure. The modelling approach was able to discriminate the health status of the honey bee colonies according to the characteristics of each scenario. To compare the estimations of the HSI scores among the seven scenarios, several approaches are available, among the simplest are the boxplot of the HSI distributions for each scenario (Fig. 4) and comparison of tertile of the HSI score distribution computed on the entire dataset (Fig. 5). In our simulation, the highest median value of the HSI distribution was reached by the FOR1 scenario, characterized by high RPU and low DII. On the other hand, the FOR4 scenario showed the lowest HSI median value in agreement with the medium RPU and the high DII. The comparison of the tertiles of the HSI scores distribution among the seven scenarios confirmed these results (Fig. 5). The 80% of the estimated HSI scores related to FOR1 were in the third tertile of the HSI score distribution, while the 70% of the estimated HSI scores related to FOR4 were in the first tertile. Furthermore, the HSI scores allow to order and compare honey bee colonies according to their estimated health status. Model outputs (in terms of sign and statistical significance of the coefficient estimates) are graphically showed in Fig. 6 (outer model results) and Fig. 7 (inner model results).
3.1.1. Outer model results

The loadings estimations (Fig. 6) confirmed the structure of the outer model defined. Loadings were consistent by sign and statistically significant for all indicators, excluding those associated to ENV. Furthermore, all the indicators were positively associated with their respective latent variables (green arrows), with the exception of the group of indicators referred to DII when directly associated with HSI. Temperature-related indicators also showed a negative association with ENV, but the estimations of their coefficients were not statistically significant (dashed arrows).

3.1.2. Inner model results

The sign (positive or negative) and the statistical significance of the estimates for the inner model are showed in Fig. 7. RPU positively influenced the scores of QUE, IHP and DEM, and negatively influenced the scores of CON and DII. The estimate of the path coefficient of RPU on BEH was not statistically significant. Estimates of BMP path coefficients were all negative, apart from that on QUE, which however was not statistically significant. ENV was positively related to IHP and DEM and negatively related to CON and DII. The estimates of path coefficients linking ENV to QUE and BEH were not statistically significant. All the latent variables of colony attributes were positively related to HSI, with the exception of DII that presented a negative (and significant) path coefficient estimation.
4. Conclusions

The main purpose of this study was to develop and test a Health Status Index (HSI) to quantify the health status of honey bee colonies considering the influence of multiple drivers/stressors. The results of our analysis, confirmed the validity of the proposed conceptual model. The selected attributes, with the exception of ENV, were positively and significantly related to the proposed sets of indicators. The HSI provided good capacity to integrate different sources of data, allowing for adaptability (i.e. capacity to include different attributes and/or indicators). The SEM approach used to develop the HSI allows accounting for emerging evidence related to honey bee health, by adding new attributes and/or indicators. Our work was mainly focussed on testing the capacity of the SEM approach to represent the relationships hypothesised in our conceptual framework. Calibration of the model against monitoring data will allow: (i) the holistic assessment of honey bee colony health status and the estimation of risks linked to colony losses, (ii) the identification of the key drivers influencing health and the assessment of both the intensity and significance of their causal relationship, (iii) the comparative assessment of the temporal and spatial variation of honey bee colony health, (iv) the analysis of risk scenarios for the implementation of knowledge-based risk mitigation measures. Once calibrated, the proposed HSI might represent a valuable support for management decisions in the beekeeping sector for relevant stakeholders (e.g. beekeepers, risk assessors, risk managers etc.) at different spatial (from local to area-wide beekeeping) and temporal scales (medium- to long-term management decisions).

Overall, future research will focus on the investigation of the potential causal relationship between honey bee colony health status (estimated with HSI), colony outputs in terms of provisioning services (honey productivity, production of nectar) and regulating services (i.e. pollination services).

Conflict of interests

None.

Acknowledgements

This research has been supported by “Fondazione Cariplo” and “Regione Lombardia” under the project: “La salute della persona: lo sviluppo e la valorizzazione della conoscenza per la prevenzione, la diagnosi precoce e le terapie personalizzate”. Grant Emblematici Maggiori 2015-1080. We kindly thank Davide Gibin and Linda Reilly for their manuscript revision and comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2019.01.024.

References


