Silvio Knaebe, Pierre Mack[1] and Holger Bargen

Eurofins Agroscience Services Ecotox GmbH Eutinger Str. 24, 75223 Niefern-Oeschelbronn, Germany
[1] E-mail contact: Pierre.Mack@as.eurofinseu.com

First Practical Experiences with Equivalence Testing in Pesticide Risk Assessment for Bees

Introduction

Equivalence testing was first introduced in the context of pesticide risk assessment in the new draft guidance document for the risk assessment of pollinators [1]. The methodology is already used in clinical trials to compare different formulations of the same group. It is a statistical method that aims to prove the safety of a substance by demonstrating that any potential effects are smaller than the predefined acceptable threshold of 10% (Δ).

Equivalence testing contrasts with traditional difference testing, which aims to detect statistically significant differences between a treatment and a control group, often leading to inconclusive results regarding safety. In contrast, equivalence testing directly addresses the safety of a substance by aiming to prove the absence of relevant effects. Additionally, it allows researchers to directly control the risk of failing to detect a high-risk substance (false-positive rate). This error is of primary concern to risk assessors. The outcome of an equivalence test directly informs the conclusion regarding high or low risk, unlike difference testing which may lead to inconclusive results.

A further key advantage according to EFSA is that the study design can adapt to the level of concern associated with the plant protection products (PPP). No specific replication level is mandated; power analysis can inform the applicant about the likelihood of proving safety given specific effect size assumptions. Studies of high-concern PPPs may require more resources (e.g., replications) than those of low-concern PPPs.

Methods

Objective: Test equivalence testing effectiveness using a mock dataset based on realistic field studies for a lower-concern PPP.

Dataset: Colony strength data for 96 hives (48 control, 48 treated) across 6 field pairs (16 hives/pair). 6 assessments (1 pre-application, 5 post-application).

Analyses:

- **Approach 1 (TOST - Two One Sided T-tests [2]):** Equivalence tested for each assessment independently (6 tests, Figure 1). Check for all possible combinations of field pairs if 2 (15 combinations), 3 (20 combinations), 4 (15 combinations) or 5 (5 combinations) field pairs of the 6 pairs are used (Figure 2). Log-transformed data, fixed effects (treatment), random effects (field pair, field*treatment interaction). Significance α = 0.2 (one-sided). *High risk if any single assessment fails equivalence or power is not sufficient..*
- **Approach 2 (Mixed Model if TOST fails as a follow up):** Equivalence tested across all assessments combined (1 test, Figure 3). Check for all possible combinations of field pairs if 2, 3, 4 or 5 field pairs are used (Figure 4). Assessment added as fixed effect; interaction tested. Compared minimal GLM, simple mixed model (random field/region), and complex mixed model (hive nested, AR(1) covariance).
- **Sensitivity Analysis:** Both approaches repeated using subsets of data (5, 4, 3, and 2 field pairs) to assess impact of replication.

Results

- **Individual Assessment (TOST Approach):**
  - Using data from all 6 field pairs, safety **could** be demonstrated when testing each assessment time point independently (See Figure 1).
  - Success rate decreased rapidly with fewer field pairs. No safety could be found for any combinations of 2 field pairs (See Figure 2).

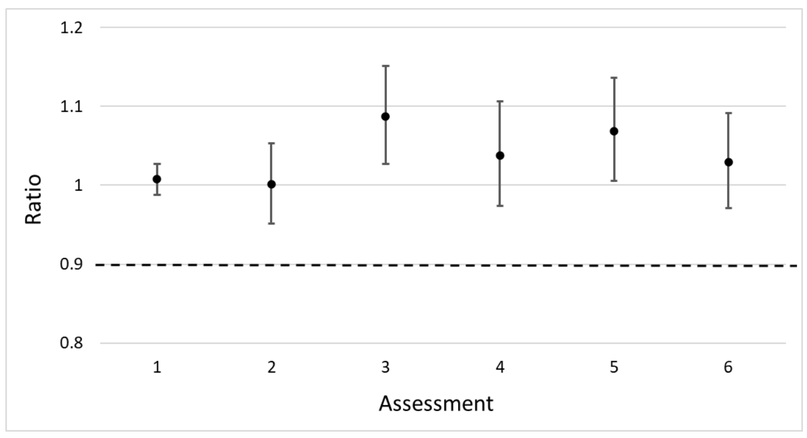

Figure 1: Equivalence test for each assessment, 6 field pairs using two one-sided t-tests (TOST) approach

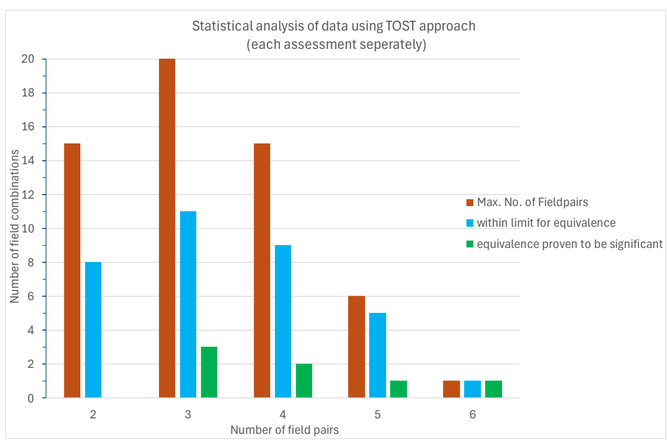

Figure 2: No of successful combinations using TOST approach depending on number of field-pairs

Overall Assessment (Mixed Model Approach):

- Analyzing the average effect across all assessments using a mixed model, the PPP **was found to be safe** (equivalence demonstrated) with 6 field pairs (See Figure 3).
- This approach showed potential to prove equivalence even with fewer field pairs (only proofed not safe with one set of 2 pairs) compared to the TOST method (See Figure 4).
- Simpler mixed models showed better convergence properties.

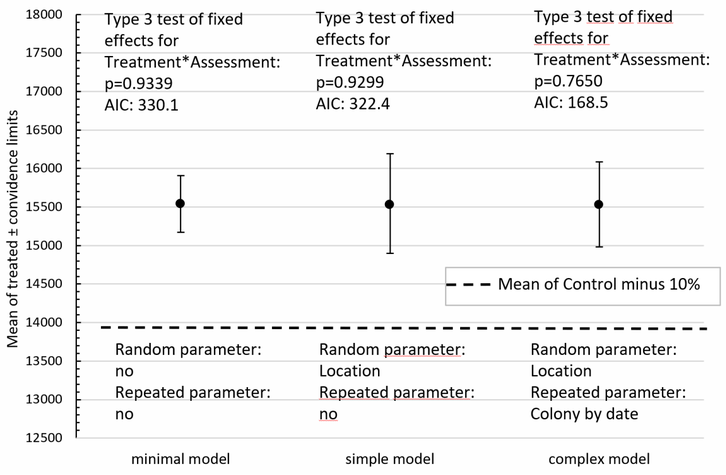

Figure 3: Equivalence test across all assessments, 6 field pairs using mixed model approach

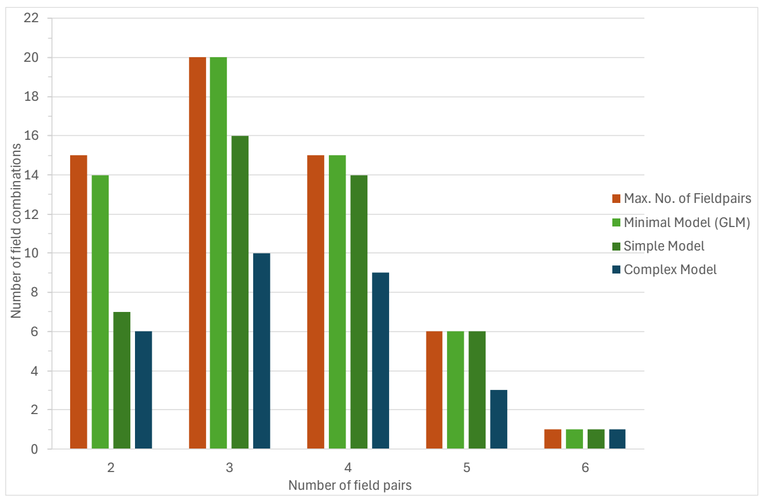

Figure 4: Number of successful combinations using mixed model approach depending on number of field-pairs

Conclusions

Method Validation: Equivalence testing, as proposed by EFSA , appears viable for bee field studies, effectively handling high natural colony variability.

Statistical Power: Sufficient statistical power to demonstrate safety (equivalence) for a low-concern PPP was achievable, even with high seasonal variation in hive development.

Efficiency Potential: The mixed model approach, analyzing effects across time, successfully demonstrated safety and required fewer field pairs than the independent assessment (TOST) approach which must be done as a first step of the data analysis in any case.

Resource Optimization: Findings suggest the number of field pairs needed for robust assessment might be lower than initially anticipated by EFSA guidance, potentially allowing for more resource-efficient study designs, especially for lower-concern PPPs.

References

[1] EFSA 2023. Revised guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA Journal 2023;21(5):7989

[2] Donald J. Schuirmann: 1987: *A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability*. In: *Journal of Pharmacokinetics and Biopharmaceutics*. 15. Jahrgang, Nr. 6, ISSN 0090-466X, S. 657–680, doi:10.1007/BF01068419