EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

17 September 2021
EMA/CVMP/ERA/87473/2021
Committee for Medicinal Products for Veterinary Use (CVMP)

# Reflection paper on higher tier testing to investigate the effects of parasiticidal veterinary medicinal products on dung fauna

| | |
|---|---|
| Draft agreed by ERAWP | July 2021 |
| Adopted by CVMP | 9 September 2021 |

| | |
|---|---|
| **Keywords** | **Environmental risk assessment, dung fauna, phase II ERA, veterinary medicinal product, parasiticides** |

**Official address** Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
**Address for visits and deliveries** Refer to www.ema.europa.eu/how-to-find-us
**Send us a question** Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

# Reflection paper on higher tier testing to investigate the effects of parasiticidal veterinary medicinal products on dung fauna

## Table of contents

# 1. Introduction

VICH guideline (GL) 38 ("Guideline on environmental impact assessment for veterinary medicinal products phase II [CVMP/VICH/790/03-FINAL]) requires laboratory-based effect studies ("tier A assessment") on dung fly and dung beetle larvae for endo/ectoparasiticides used for pasture treatments. Yet, no specific guidelines on dung fly or dung beetle larvae studies are listed, as no harmonised OECD documents were available at the time when VICH GL 38 came into effect in 2005. The CVMP "Guideline on environmental impact assessment for veterinary medicinal products in support of the VICH guidelines GL6 and GL38" (CVMP TGD; EMA/CVMP/ERA/418282/2005) was developed to give further technical support to the implementation of VICH GL6 and GL38 on the environmental risk assessment (ERA) of veterinary medicinal products (VMPs), where additional regulatory guidance was deemed necessary for the ERA of VMPs. However, the focus was on exposure assessment and the CVMP TGD does not include any reference on how to proceed if the initial tier A risk assessment indicates a risk to dung flies or beetles.

Since VICH GL 38 came into effect in 2005, the OECD has published two relevant guidelines (OECD guidance document 122 and OECD test guideline 228) for ecotoxicity testing of substances to dung fauna.

Dung, especially from larger mammals, makes up a complex and highly dynamic ecosystem within a small environmental scale. Odour from excreted dung almost instantly attracts flies, which feed, mate, and lay eggs on the dung, leading to a new generation of flies emerging within a few weeks. Fly numbers on the dung rapidly decline after a few hours when crust formation on the dung pat reduces the scent. After the flies, dung-feeding beetles arrive at the pat, with the colonisation peak typically finishing by the end of the first week. In contrast to flies, the development time of beetles may take weeks to months. Parasitic wasps and predatory beetles arrive concurrently with their prey (i.e. flies and beetles), and may either lay eggs or feed on the immature insects developing in the dung pat. In less than three weeks after the pat is dropped, the colonisation of dung is almost finished. After this time, tunnelling and feeding activities by insects and oligochaetes as well as the penetration of vegetation accelerate dung pat degradation. With this, access is provided to soil-dwelling organisms like earthworms and bacteria, which complete the breakdown of dung into parts that are finally incorporated into the soil matrix. From time of deposition to total degradation, a dung pat may contain several dozen species of dung insects exceeding thousands of individuals. Dung pats have also been shown to affect spatial earthworm distribution, with markedly more earthworms aggregating beneath pats (Bacher *et al.* 2018).

The species inhabiting these complex and dynamic ecosystems may be put under threat from numerous natural stressors as well as changes in agricultural practices, e.g. a decline in free-range animals in regions with intensively reared animals. Furthermore, the use of parasiticides is one of the well-documented threats to dung fauna, as the high effectiveness against invertebrate parasites in pasture animals also results in high toxicity to non-target invertebrates such as dung insects. The observed population decline of many dung-dwelling species, some of them being red-list species, may not be a result from a single stressor, but rather a combination of stressors and may differ regionally or locally.

This reflection paper solely addresses the concern and potential risk associated to the use of parasiticides, as it reflects on key elements relevant for investigating the environmental impact of parasiticides on dung fauna monitored in higher tier field studies, in situations where the initial tier A risk assessment indicates an unacceptable risk to dung flies or beetles (Adler *et al.* 2013). In such situations, the question of fundamental concern is whether the often very strong effects seen in experimental laboratory studies at realistic exposure concentrations are likely to have an impact on

insect populations, community interactions and the process of dung decomposition under realistic large scale field conditions. The scale, the required investment(s) and the scientific requirements outlined in this reflection paper focus on the situation where limited information about the active substance is available and where robust new knowledge is necessary in order to elucidate a potential environmental risk. It is hence not designed as a "one-way-suits-all-purposes" approach. For some existing substances, useful information may be retrieved from numerous field studies available in the public domain, which can help in optimising the design of a planned field study or may be submitted with the dossier if the publication contains a sufficient amount of data and sufficient details on the design and conduct of the study to allow a full and independent assessment. The CVMP "Reflection paper on the interpretation of Article 18(7) of Regulation (EU) 2019/6" (EMA/CVMP/ERA/622045/2020) describes in which cases an ERA for generic VMPs is to be provided under Regulation (EU) 2019/6[1]. If it comes to the exceptional situation that it is considered necessary to develop a higher tier study similar to the one described below when an ERA for a generic VMP needs to be provided, it is advised that the marketing authorisation holders of these similar VMPs join efforts towards the development of such studies due to their complexity as well as the required resources.

Chapter 2 addresses some of the challenges, pitfalls, limitations and possibilities that need to be considered when planning, executing and reporting a higher tier field study investigating the effects of parasiticidal veterinary medicinal products on dung fauna.

# 2. Discussion

Wall and Beynon (2012) reviewed the information on large scale studies on the ecological impact of parasiticides and concluded that "[...] [t]he extent to which chemical residues may have any sustained ecological impact will depend on both a range of farm management factors, such as the temporal and spatial patterns of chemical use, the number of animals treated and the choice of active ingredient, and a range of insect-related factors, such as abundance, population dynamics and dispersal rates. However, they also demonstrate that considerable uncertainty remains about the likely extent of such effects and that current data are insufficient to support firm conclusions regarding sustained pasture-level effects. More large-scale, long term field experiments are required, particularly in relation to insect dispersal and functional interactions within the dung insect community".

Furthermore, other spatial and temporal factors such as the local weather conditions, frequency and seasonal timing of treatments throughout the year as well as species life cycles may exert confounding effects on the toxicity of the VMP under investigation. All of these factors influence the natural variation and temporal as well as spatial fluctuations and are hence likely to hamper the interpretation of results. It is therefore of importance to design field studies that are, on the one hand, as realistic as possible and, on the other hand, so robust, standardised and reproducible that the results can be used in a wider context and interpreted in a straightforward and transparent fashion.

It is not within the scope of this reflection paper to outline specific study protocols or guidelines for field sampling, as examples of such can be found in various publicly available documents and papers. For instance, publications by Römbke *et al.* (2010), Jochmann *et al.* (2011) and Adler *et al.* (2016) may provide relevant information in this context.

[1] Regulation (EU) 2019/6 of the European Parliament and of the Council of 11 December 2018 on veterinary medicinal products and repealing Directive 2001/82/EC. OJ L 4, 7.1.2019, p. 43–167.

## 2.1. Overall considerations

When designing, executing and reporting a large scale higher tier field study, a number of issues need to be considered prior to its initiation. These include, but are not limited to aspects such as:

- Selection of protection goals
- Selection of assessment endpoints and acceptable outcomes
- Evaluation of natural variation and minimum detectable difference
- Design of study and associated spatial and temporal sampling strategy

Although the various aspects listed above are unique, they are interlinked in the sense that, for example, the selection of protection goals drives the design of the study, while the natural variation in assessment endpoints together with the available resources influence the potential of concluding on the impact of these protection goals. It is hence of paramount importance to ensure a feedback mechanism between the various aspects in order to end up with an optimal study design.

## 2.2. Selection of protection goals

Before initiating any field studies, the protection goals should be identified. Protection goals could span from an overall attainment of dung pat degradation, to the total abundance of dung beetles or the protection of individual red-listed species at field level. Relevant protection goals are identified below, with some of them being associated to ecosystem structure (biodiversity), while others are associated to ecosystem functioning (ecosystem services):

- **The populations of dung-dwelling beetles**
- **The populations of dung-dwelling flies**
- The populations of specified endangered dung fauna species
- The total abundance of dung fauna
- **The degradation of dung pats**
- The populations of soil dwelling fauna associated to dung pats

Not all protection goals may be equally important for all scenarios. In most cases, the three protection goals highlighted in bold would together be sufficient for a full protection of the communities of dung-associated species. However, study reports should include reflections and scientifically based judgements for the selection of protection goals relevant for the VMP in question. There is a link between structural changes and functional impairment, although these may not be correlated in a one-to-one fashion, as some tolerant species may compensate for the loss of more sensitive species. Nevertheless, over time, decreased dung beetle populations will ultimately result in a reduced dung degradation and dung burial (Jacobs and Scholtz 2015, Beynon *et al.* 2012). This process, often termed an ecosystem service, has been associated with large economic benefits and, for example, saves the UK cattle industry more than £367 million each year (Beynon *et al.* 2015).

For each relevant protection goal, one or more suitable assessment endpoint(s) need(s) to be identified (see below). The assessment endpoints and the design and outcome of the study should reflect the desired choice of protection goals. **Box 1** contains a list of recommended and suitable protection goals, which are considered sufficient for most situations.

## *2.3. Selection of assessment endpoints and acceptable outcomes*

When one or more protection goals has been identified, the next step would be to identify relevant assessment endpoints associated to each of these. If, for example, the population of dung beetles has been chosen as a relevant protection goal, the associated assessment endpoint would, for instance, be a taxonomic determination and quantification of beetles extracted from dung at species-, family- or sub-class level. If dung degradation would be the protection goal, changes in (dry or wet) biomass of dung pats over time could be a relevant assessment endpoint.

The ambition and objective when selecting the level of uncertainty may change according to the resources available. At the same time, the resources allocated must change according to the required certainty for establishing an acceptable impact. Acceptable effects and the uncertainty in establishing these would therefore have to be discussed and chosen upfront, as these feed into the design of the study. If only minor changes in e.g. abundance of dung beetles at species level is accepted, the study would have to be designed in a way that it enables a powerful statistical differentiation between non-exposed (control) and exposed populations even in case differences are minor.

The natural fluctuation (variation) observed for the chosen assessment endpoints will influence the outcome of the study. Natural fluctuations, due to confounding factors, affect the quantifiable level of difference and the statistical power of the study. In other words, the higher the variance, the lower the statistical power, which leads to an enhancement of the minimum detectable difference, i.e. with a high variance and a fixed set of replications, only very high effect differences between exposure and control populations can be statistically detected, which may be unsatisfactory from a regulatory point of view. Cases in which large effect differences (e.g. 50%) between exposure and control populations are not statistically significant may therefore not reflect a situation where there is no adverse impact on ecosystems, but rather a study with too low statistical power. **Box 1** contains recommendations regarding assessment endpoints and replication.

For substances with high intrinsic toxicity such as pesticides, biocides or parasiticides designed to kill not only target species, but also related non-target organisms, it is beyond any reasonable doubt that (temporary) impact will occur shortly after environmental release. Numerous published field studies have documented a significant impact of parasiticides in at least the first initial period after medication (Jacobs and Scholtz 2015). Others have shown long-lasting effects (Floate 2006, Nieman *et al.* 2018) for a period potentially longer than the active substance can be detected in the environment.

A long term and wide-ranging impact on dung fauna populations is likely to be concurrent to the period where animals excrete toxic manure. If this is rapidly reversed to a suitable non-toxic habitat for dung fauna, then the overall environmental impact at population level may not be dramatic and long-lasting. It must hence be discussed if a temporal window of effects can be accepted and, if so, how long such an acceptable time zone of effects (ATZE) would be. A similar approach is applied, for instance, in marine aquaculture, where competent authorities have identified so-called "allowable zones of effects" (AZE)[2] in a distance around fish net-pens, acknowledging that zero risk cannot be obtained for certain active substances or certain chemical use.

Such a time-window with acceptable effects depends on the life cycle of the species, as species with a long reproductive cycle (e.g. many dung beetles) with one or a few oviposition(s) would potentially be more sensitive to prolonged exposure to parasiticidal VMPs as their possibility of a successful breeding season is more narrow than species with a shorter life cycle and many ovipositions. Since resources rarely are unlimited, it is worthwhile considering when it is most optimal to collect samples. By

[2] AZEs are defined as the area (or volume) of sea bed or receiving water in which some exceedance of a relevant environmental quality standard (EQS) is allowed (by e.g. the Scottish Environmental Protection Agency [SEPA]).

reducing the sampling periods and instead increasing the number of replicates sampled in each period, the statistical power may be enhanced.

In summary, a suitable set of assessment endpoints could therefore be outlined as follows:

- The impact on beetles collected from dung pats at four time slots starting after an ATZE of 2 weeks. The last sampling would depend on metabolism (ADME) studies and the excretion profile, but could be aligned with the time until the concentration of the active substance in dung reaches the limit of detection.
- The impact on flies collected from dung pats sampled from animals at four time slots starting after an ATZE of 4 weeks. The last sampling would depend on ADME studies and the excretion profile, but could be aligned with the time until the concentration of the active substance in dung reaches the limit of detection.
- The impact on dung degradation, sampled at least at five sample slots covering the full degradation period or a maximum duration of 180 days.

The above is only a recommendation, and as part of the study planning phase, it is strongly advised to seek guidance from relevant competent authorities on this and other related issues, as it can help focus the sampling on endpoints and periods pertinent for decision-making. **Figure 1** and **Box 1** outline the sampling strategy for dung beetles and dung flies, which can be used as a starting point for such a discussion.

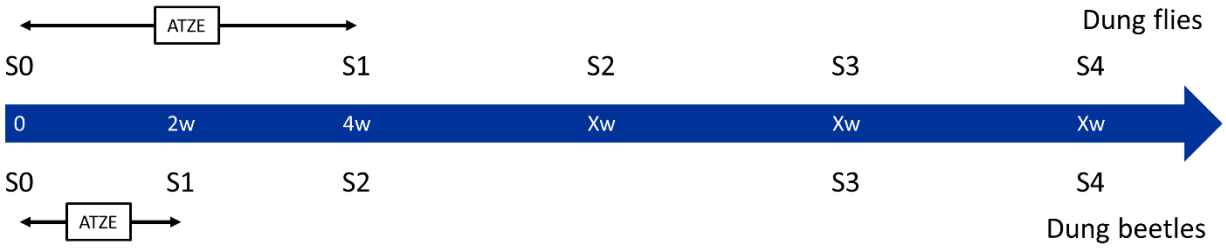


**Figure 1.** An example for a test and sampling strategy based upon two different ATZEs for dung flies and dung beetles, respectively. Xw indicates the time period in weeks after medication (day 0) and needs to be adjusted in accordance with ADME studies identifying the excretion profile and the duration with detectable amounts of active ingredient in the dung. S0 denotes the samples collected from animals before treatment (control). S1–S4 denotes the samples collected from animals after the treatment. All collected dung is left simultaneously as pats in the field for 7-10 days in order to be inhabited by fauna before being transferred to the laboratory for taxonomic identification.

## *2.4. Evaluation of natural variation and minimum detectable difference*

Natural variation over time and space is one of the key elements potentially hampering the conclusive outcome of field studies. Minimizing variation and influence of confounding effects is in fact one of the main motivations for developing and using standardised and simplified laboratory studies in ERA. However, in the field, spatial and temporal variation is a natural and unavoidable factor that needs to be considered in the design of a study and the evaluation of its results. Generally, the more detailed and specific a study aims to be, the more variance must be anticipated. It is expected that the variation in species abundance, diversity or richness monitored in dung pats at a narrow time slot will be larger than the same parameters collected over a longer period in a pitfall or bait trap (see below for more details).

One of the first steps in planning and designing a higher tier field study to investigate effects on dung fauna would hence be to collect reliable information about the natural variation for the assessment endpoints in question. This may change by the season and from region to region. Available information can be used to determine a suitable place and time of the year for conducting the experiment. If no such data is available, for instance from the publicly available scientific literature and/or experts with relevant experience, it can be useful to perform a pilot study prior to the full-scale full-factorial study.

The power of any (field) test is defined as 1-β. Power increases with increasing sample size and with decreasing variability and also depending on the value of α, i.e. the probability of making a type I error (false positive). The statistical power of a given study is inversely related to the probability of making a type II error (false negative), i.e. to conclude that there is no effect even though an effect is present[3]. That is, an effect has not been detected because of missing statistical significance.

Ecological experiments can be improved to increase the statistical power by selecting the sample sizes necessary to detect a given difference between treatments (control vs. treatment). It is generally recommended that the statistical power would be equal to at least 0.8, i.e. β should not exceed 0.2, but preferably, it should be 0.1.

Using statistical methods and information on the natural variation typically observed for the sampling or assessment endpoints, it is possible to predict the theoretical minimum detectable (significant) difference (MDD) between a control and exposed group with a given number of replicates prior to the actual performance of an experiment. Similarly, the theoretically minimum number of replicates (MNR) can be predicted, in order to statistically demonstrate a given significant difference between a control and exposed group. **Table 1** below includes a shortlist of the relationships between number of replicates, minimum detectable difference and the natural variance expressed as coefficient of variation (CV), i.e. the relative standard deviation or the ratio between the standard deviation and the mean (σ/μ) expressed as percentage. **Table 2** shows the minimum number of replicates in each of two groups (control vs. exposure) required to demonstrate a predefined difference between control and exposure groups, i.e. the impact. It is worth mentioning that these initial considerations are based on statistical predictions. The study-specific on-site variance will determine the actual statistical power of the study in question.

**Table 1.** Data extracted from Kraufvelin (1998) showing the theoretic relationships between CV and MDD using an α of 0.05 and a β of 0.8. n is the number of replicates in each treatment group. It is assumed that the study is only composed of a control and one exposure group (k = 2), e.g. control dung vs. dung from medicated animals at a given timepoint. If more exposure groups are included when keeping the total number of replicates in the study unchanged, then the MDD is increased accordingly. For instance, a scenario with a CV of 50, n = 4 and k = 2 has a MDD of 119%, whereas a scenario with CV of 50, n = 2 and k = 4 has an MDD of 186%.

| CV | MDD (%) | | | | | |
|---|---|---|---|---|---|---|
| | n = 2 | n = 3 | n = 4 | n = 5 | n = 6 | n = 7 |
| 10 | 54 | 30 | 24 | 20 | 18 | 16 |
| 20 | 107 | 61 | 47 | 40 | 36 | 33 |
| 30 | 161 | 91 | 71 | 61 | 54 | 49 |
| 40 | 215 | 121 | 95 | 81 | 72 | 65 |

[3] α is defined as the probability of making a type I error in hypothesis testing, i.e. the probability of incorrectly rejecting the null hypothesis when it is true (false positive); β is defined as the probability of making a type II error in hypothesis testing, i.e. the probability of incorrectly accepting the null hypothesis when it is false (false negative); 1-β defines the power of the test.

| | | | | | | |
|---|---|---|---|---|---|---|
| 50 | 268 | 152 | 119 | 101 | 90 | 82 |
| 100 | 536 | 304 | 237 | 202 | 179 | 163 |

**Table 2.** Data extracted from Kraufvelin (1998) showing the theoretic relationships between CV and MNR (in each group) required to confirm a predefined statistical difference (% effect) between a control and an exposed group using an α of 0.05 and a β of 0.8. It is assumed that the study is only composed of a control and one exposure group (k = 2), e.g. control dung vs. dung from medicated animals at a given timepoint.

| CV | MNR | | | |
|---|---|---|---|---|
| | 100% effect | 75% effect | 50% effect | 25% effect |
| 10 | 2 | 2 | 3 | 4 |
| 20 | 3 | 3 | 4 | 12 |
| 30 | 3 | 4 | 7 | 24 |
| 40 | 4 | 6 | 12 | 42 |
| 50 | 6 | 9 | 17 | 64 |
| 100 | 17 | 29 | 64 | 251 |

The natural variation in assessment endpoints, relevant in the context of higher tier field studies for assessing impact on dung fauna of parasiticides, may differ largely depending on the time of year, the geographic region and the monitoring endpoint or taxonomic resolution in data. Appendix A includes an example from a larger trans-national survey reported in Floate *et al.* (2016). Here, observations regarding CV in control pats were as follows:

- The CV in monitoring endpoints decrease in the order total abundance, species richness and species diversity;
- The average CV among abundance measurements for different taxonomic groups is 29.5% and spans from 13.2 to 60%;
- The average CV for abundance from the four locations in Canada (n = 10 from two years), France (n = 5), The Netherlands (n = 4) and Switzerland (n = 5) was 37.8, 20.7, 22.9 and 26.9%, respectively;
- Abundance data on family level (not shown in Appendix A) was somehow higher and ranged from 11.5 to 100% with a mean of 38% (Floate *et al.* 2016);
- The CV on the abundance at genus level is markedly higher reaching 300%, with a mean of almost 200% (data not shown);

The data extracted from Floate *et al.* 2016 summarised above and presented in Appendix A was based on dung fauna collected from dung pats collected from medicated animals and disposed in the fields. Floate (2007) reported the results from a three-year study using pitfall traps loaded with control cattle dung as bait as well as dung from animals medicated with various parasiticides. Insects captured comprised more than 300,000 individuals representing approximately 40 taxa. Cattle were excluded from the pasture until after the completion of the autumn trapping period, wherefore recruitment of insects came from the surrounding pasture area. The CV for abundance at individual species or family level in traps with control baits ranged from 6.7 to 50%, with a mean of 18.8% (data not shown),

which could indicate that the variation in counts from pitfall (bait) traps could be lower than when sampling directly from field-based pats.

To sum up the information in **Tables 1** and **2**, it demonstrates the importance of knowing the natural variation (CV) in the assessment endpoint and the benefit of a thorough planning when designing the field study. Pros and cons of including more treatment groups, i.e. dose-response strategy versus comparing one control and one exposure, must be discussed and sincerely considered, as the number of total replicates needed to demonstrate a predefined MDD increases rapidly with the number of treatment groups. Data shows, furthermore, that doubling the variance in data from CVs of 20 to 40% could greatly increase the minimum total number of replicates from 24 (2 x 12) to 84 (2 x 42) in cases where the aim is to show a statistically significant 50% negative effect of parasiticides on dung fauna compared to a control. Such a doubling in CV could, for example, occur by assessing abundance at species level versus at family level. Acknowledging the time and effort required for taxonomic evaluation of dung fauna communities, such a divergence in sampling size could require a substantial amount of resources. It is hence of paramount importance that theoretical considerations on the MNR and MDD forms the platform and starting point for the design of the field study, wherefore data and calculations should be included in the final study report. **Box 1** contains some general recommendations, which can form the basis of such a discussion.

When executing and performing the planned field study, the actual observed variance in samples may differ from the theoretical prediction and the final data and conclusion may hence be made with either higher or lower statistical power than predicted. As studies with an unsatisfying statistical power may be disregarded as non-conclusive, it is generally recommended to base the study on conservative estimates and predictions, i.e. assuming variance in the higher range of the reported observations, all while considering the available resources.

In summary, a suitable sampling strategy could therefore be outlined as follows (see **Box 1** for more details):

- Based upon a predicted CV of 25% and a target of detecting a 25% difference between control and treatment groups, it is recommended for each of the temporal sampling slots to sample a minimum of 2 x 20 replicates for evaluating the impact on dung beetles at family level[4].
- Based upon a predicted CV of 40% and a target of detecting a 50% difference between control and treatment groups, it is recommended for each of the temporal sampling slots to sample a minimum of 2 x 12 replicates for evaluating the impact on dung flies at family level.
- Less information regarding the CV for dung degradation is available, but it is generally anticipated to be lower than for species, wherefore a sampling of 2 x 10 replicates at each temporal sampling slot is considered sufficient.

The aim is not only to test whether two populations differ, but also to quantify how far apart the two means are. Therefore, consideration could be given to the determination of the confidence interval of the abundance with as much accuracy as possible. This could call for a larger sample size in the untreated group compared to the treated group, as variation in the former is expected to be higher.

The above is only a recommendation, and, as part of the planning phase, it is strongly advised to seek guidance from relevant competent authorities on this and other related issues, as it can help focus the sampling on endpoints and periods pertinent for decision-making.

[4] The variation for dung beetles at species level differs a lot. The indicated number of replicates may therefore be sufficient to distinguish differences at species level.

### *2.5. Design of study and associated spatial and temporal sampling strategy*

Even when considering the factors discussed above, there is a large degree of flexibility and possibilities for choosing the final design of the higher tier field study. These may, for example, be the following:

- Use of natural dung pats disposed on the field or constructed standardised pats collected from stabled animals;
- A full "Before-After-Control-Impact" (BACI)[5] test design or, for example, a "control-impact" design only;
- Placing of experimental dung pats in a (fenced) animal free area, within a field grassed by animals producing dung pats forming the recruitment potential for the experiment;
- Use of pitfall traps/bait traps alone or in combination with sampling of dung pats as a tool to monitor the overall presence and recruitment potential on a larger field or landscape level;
- Collection and monitoring of samples over a certain time period or only at (a) specific timepoint(s) having in mind potential time windows of acceptable effect identified by competent authorities;
- Performance of studies in only one representative region or in different regions acknowledging that the dung fauna community and the importance of soil fauna in dung degradation can be markedly different in different regions of Europe (e.g. the Mediterranean and the Atlantic regions);
- Decision of which target species to be included in the field study if the VMP is intended for use in different target animals;

Whatever the decisions finally made, they must consider the chosen protection goals, assessment endpoints and the necessary statistical power of the study. Furthermore, decisions must be transparent, scientifically justified and include reflections about any omission of e.g. husbandry and test regions.

Common for all of the possible test designs is a need of chemical verification of concentrations in dung of the active substance and relevant metabolites at any sampling point. When performing chemical analysis, special attention should be given to highly adsorptive substances, which may form non-extractable residues complicating the extraction procedure and hence the interpretation of results.

# 3. Conclusion

Very often, the tier A assessment of parasiticides result in a conclusion of unacceptable risk to dung fauna when performed according to the procedures outlined in relevant VICH and CVMP guidelines. In these cases, the VMP has to be evaluated in a higher tier assessment. Currently, no guideline on higher tier testing of parasiticides exists, wherefore applicants, often in association with competent authorities, have to design and execute higher tier studies on a case-by-case basis. This reflection paper addresses some of the challenges applicants will face and some of the points that need to be considered prior, while and after the performance of a field study.

[5] A "Before-After-Control-Impact" (BACI) design is suggested as being a statistically powerful experimental design in environmental impact studies. If the timing and location of the impact are known and adequate pre-study data are collected, the BACI design is considered optimal to help isolate the effect of the development from natural variability (Smokorowski and Randall 2017).

Generally, the process can be separated into four major steps described and discussed in detail in the sections above and summarised below:

1. Selection of protection goals.
2. Selection of assessment endpoints and acceptable outcomes with regard to magnitude of impact and potential time slots with acceptable effects.
3. Evaluation of natural variation and MDD.
4. Design of the study including a spatial and temporal sampling strategy.

Typically, both structural and functional protection goals need to be considered although it is generally accepted that, by protecting the structure of ecosystems, associated ecosystem services or functions will also be protected, provided the structural protection goals cover a sufficient taxonomic resolution. When selecting assessment points, a number of issues need to be taken into account, e.g. available scientific (taxonomic) expertise, available financial resources and time. Before designing and executing the study, the natural variation in the assessment endpoints needs to elucidated and determined in order to conclude if the available resources are sufficient to ensure a satisfactory statistical power in the study promoting solid conclusions. If no relevant data is available, a pilot study aiming at determining the natural variation(s) may be necessary. If a satisfactory statistical power, i.e. an adequate MDD, cannot be obtained, more resources must be allocated, or the protection goal(s) or the assessment endpoint(s) need(s) to be aligned accordingly. This could be done by reporting abundances at a lower taxonomic resolution, reducing sampling periods, increasing replication numbers, accept a higher level of effect as not being significantly different from control or otherwise reduce variation in data leading to a higher statistical power. When all of these desktop evaluations are made, the final design and practical execution of the study can be planned in details. It is not within the scope of this reflection paper to outline a specific mandatory study protocol or guidelines for field sampling, although **Box 1** includes the outline of a recommended study set up which could form the basis for the final design. It is strongly advised to seek guidance from relevant competent authorities, as it can help focus the sampling on endpoints and periods pertinent for decision-making.

Regarding decision-making, it can be considered to present data using, for instance, different regression models or more complex population models, in order to determine when residues in dung excreted from treated animals return to levels causing a small (e.g. 10 or 20%) impact on dung fauna.

**Box 1.** Outline of a draft design for a field study monitoring the effects of parasiticides on dung-dwelling organisms. The outline below is only a recommendation and hence not mandatory.

1. Use of ADME studies to determine the window of excretion in order to detect the peak of excretion and the timespan until no active ingredients can be analytically determined, i.e. < LOD.
2. A number of artificial dung pats (1 kg) produced on the basis of dung collected from medicated animals at day 0 (control) and at least four additional timeslots taking into consideration the ATZE for dung flies and dung beetles, respectively, and the period with detectable residues in dung (see **Figure 1** for more details). All constructed dung pats are placed in the field simultaneously on the same day.
3. The dung pats can be used directly or stored frozen (-18 °C) until use. All replicates should preferably be collected from separate animals. Alternatively, the pats must be collected from at least five treated animals and pseudo-replicates equally made from these. The concentration of the active ingredient must be verified in all dung pats by state of the art analytical means, with special attention being paid to the documentation of potential non-extractable residues.
4. Placement of constructed pats on top of a perforated container (to allow drainage) filled with soil and covered by wire mesh cages to exclude birds (but not insects). Pats are placed in a fenced part of a larger field, e.g. 2 m apart. Animals (preferably untreated) must grass on the remaining part of the field. Identification tags must be uniquely associated to each pat.
5. Pats used for the degradation study must be left in the field for a period covering the time until disappearance or a maximum of 180 days. During the monitoring period, the reduction in mass must be determined at up to five sampling time slots, e.g. 2, 4, 8, 16 and 26 weeks after placement in the field.
6. Pats used for taxonomic evaluations are left in the field for a period of 7-10 days to allow for insect colonization and oviposition. The containers and their associated pats must subsequently be tansferred to a laboratory and placed in insect emergence cages. Details on insect emergence cages can be found in publicly available scientific literature (e.g. in Floate *et al.* 2016).
7. Insects recovered from emergence cages are preserved (e.g. in 95% ethanol) and subsequently sorted, counted, and identified at the highest possible taxonomic resolution. For dung beetles this should preferably be done at least up to species level, while for dung flies and other species this can be done up to family level.
8. The number of replicates must reflect the protection goals, the assessment endpoints and the predicted natural variance in these. In many cases, a setup using 2 x 20 replicates for each dose (time after medication), acknowledging that these numbers in most cases would be sufficient to statistically distinguish deviations from controls of 25 and 50% for beetles and flies at family level, respectively. The specific statistical power would, however, be site- and study-specific.
9. The collected data are used to evaluate the potential risk of the parasiticidal product. Data for dung beetles must preferably be presented at both species and family level. Data on dung flies only at family level. Unacceptable risk can, for example, be identified as any statistically significant adverse effect on dung flies and beetles at family level in dung pats excreted 14 days or more after medication. Data analysis of variance should preferably demonstrate sufficient statistical power to identify a 25% deviation at family level between the control and exposed dung pats.

# 4. References


Adler N, Bachmann J, Römbke J. 2013. New test strategy for dung beetles during the authorization process of parasiticides. Integr Environ Assess Manag. 9(3):524–530. https://doi.org/10.1002/ieam.1399.

Adler N., Bachmann J, Blanckenhorn WU, Floate KD, Jensen J, Römbke J. 2016. Effects of ivermectin application on the diversity and function of dung and soil fauna: regulatory and scientific background information. Environ Toxicol Chem. 35(8):1914–1923. https://doi.org/10.1002/etc.3308.

Bacher MG, Fenton O, Bondi G, Creamer RE, Karmarkar M, Schmidt O. 2018. The impact of cattle dung pats on earthworm distribution in grazed pastures. BMC Ecol. 18:59. https://doi.org/10.1186/s12898-018-0216-6.

Beynon SA, Mann DJ, Slade EM, Lewis OT. 2012. Species-rich dung beetle communities buffer ecosystem services in perturbed agro-ecosystems. J Appl Ecol, 49: 1365-1372. https://doi.org/10.1111/j.1365-2664.2012.02210.x

Beynon SA, Wainwright WA, Christie M. 2015. The application of an ecosystem services framework to estimate the economic value of dung beetles to the U.K. cattle industry. Ecol Entomol, 40(S1; Suppl):124–135. https://doi.org/10.1111/een.12240.

Floate KD. 2006. Endectocide use in cattle and fecal residues: environmental effects in Canada. Can J Vet Res. 70(1):1–10.

Floate KD. 2007. Endectocide residues affect insect attraction to dung from treated cattle: implications for toxicity tests. Med Vet Entomol. 21(4):312–322. https://doi.org/10.1111/j.1365-2915.2007.00702.x.

Floate KD, Düring RA, Hanafi J, Jud P, Lahr J, Lumaret JP, Scheffczyk A, Tixier T, Wohde M, Römbke J, Sautot L, Blanckenhorn WU. 2016. Validation of a standard field test method in four countries to assess the toxicity of residues in dung of cattle treated with veterinary medical products. Environ Toxicol Chem. 35(8):1934–1946. https://doi.org/10.1002/etc.3154.

Jacobs CT, Scholtz CH. 2015. A review on the effect of macrocyclic lactones on dung-dwelling insects: Toxicity of macrocyclic lactones to dung beetles. Onderstepoort J Vet Res. 82(1):858. https://doi.org/10.4102/ojvr.v82i1.858.

Jochmann R, Blanckenhorn WU, Bussière L, Eirkson CE, Jensen J, Kryger U, Lahr J, Lumaret JP, Römbke J., Wardhaugh K, Floate KD. 2011. How to test nontarget effects of veterinary pharmaceutical residues in livestock dung in the field. Integr Envir Assess Management. 7(2):287–296. https://doi.org/10.1002/ieam.111.

Kraufvelin P. 1998. Model ecosystem replicability challenged by the "soft" reality of a hard bottom mesocosm. J Exp Mar Biol Ecol. 222(1–2):247–267. https://doi.org/10.1016/S0022-0981(97)00143-3.

Nieman CC, Floate KD, Düring RA, Heinrich AP, Young DK, Schaefer DM. 2018. Eprinomectin from a sustained release formulation adversely affected dung breeding insects. PLoS One. 13(8):e0201074. https://doi.org/10.1371/journal.pone.0201074.

Römbke J, Coors A, Fernández AA, Förster B, Fernández C, Jensen J, Lumaret JP, Cots MAP, Liebig M. 2010. Effects of the parasiticide ivermectin on the structure and function of dung and soil invertebrate communities in the field (Madrid, Spain). Appl Soil Ecol. 45(3):284–292. https://doi.org/10.1016/j.apsoil.2010.05.004.

Smokorowski KE, Randall RG. 2017. Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs. FACETS. 2(1):212–232. https://doi.org/10.1139/facets-2016-0058.

Wall R, Beynon S. 2012. Area-wide impact of macrocyclic lactone parasiticides in cattle dung. Med Vet Entomol. 26(1):1–8. https://doi.org/10.1111/j.1365-2915.2011.00984.x.

**Appendix A (data from Floate *et al.* [2016])**

| Location | Year | N | Group | Mean | Std | CV | Mean | Std | CV | Mean | Std | CV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Abundance | | | Richness | | | Diversity | | |
| Lethbridge, Ca | 2011 | 20 | All taxa | 24 | 5.3 | 22.1 | 5.8 | 0.5 | 8.6 | 4.3 | 0.4 | 9.3 |
| Lethbridge, Ca | 2012 | 10 | All taxa | 129.7 | 32.6 | 25.1 | 8.9 | 0.8 | 9.0 | 4.3 | 0.2 | 4.7 |
| Montpellier, F | 2011 | 10 | All taxa | 116.3 | 15.3 | 13.2 | 11.5 | 0.6 | 5.2 | 6.8 | 0.6 | 8.8 |
| Wageningen, NL | 2011 | 10 | All taxa | 107.2 | 17.3 | 16.1 | 12.9 | 1.3 | 10.1 | 5.7 | 0.6 | 10.5 |
| Zürich, CH | 2011 | 10 | All taxa | 162.3 | 27.7 | 17.1 | 16.7 | 1.4 | 8.4 | 6.1 | 0.6 | 9.8 |
| Lethbridge, Ca | 2011 | 20 | Coleoptera | 15.5 | 3.1 | 20.0 | 3.9 | 0.4 | 10.3 | 3.1 | 0.3 | 9.7 |
| Lethbridge, Ca | 2012 | 10 | Coleoptera | 68 | 21.7 | 31.9 | 3.3 | 0.5 | 15.2 | 1.9 | 0.1 | 5.3 |
| Montpellier, F | 2011 | 10 | Coleoptera | 13.4 | 3 | 22.4 | 2.9 | 0.3 | 10.3 | 1.4 | 0.1 | 7.1 |
| Wageningen, NL | 2011 | 10 | Coleoptera | 27.3 | 5.1 | 18.7 | 6.8 | 0.6 | 8.8 | 2.2 | 0.2 | 9.1 |
| Zürich, CH | 2011 | 10 | Coleoptera | 119.3 | 22.2 | 18.6 | 6.7 | 0.7 | 10.4 | 2.5 | 0.2 | 8.0 |
| Lethbridge, Can | 2011 | 20 | Diptera (Brachycera) | 5.8 | 2.5 | 43.1 | 1.2 | 0.2 | 16.7 | 1.3 | 0.1 | 7.7 |
| Lethbridge, Can | 2012 | 10 | Diptera (Brachycera) | 53.6 | 12.3 | 22.9 | 3 | 0.4 | 13.3 | 1.9 | 0.2 | 10.5 |
| Montpellier, F | 2011 | 10 | Diptera (Brachycera) | 61.2 | 8.7 | 14.2 | 4.6 | 0.3 | 6.5 | 2.4 | 0.1 | 4.2 |
| Wageningen, NL | 2011 | 10 | Diptera (Brachycera) | 75.5 | 15.5 | 20.5 | 4.4 | 0.5 | 11.4 | 2.2 | 0.1 | 4.5 |
| Zürich, CH | 2011 | 10 | Diptera (Brachycera) | 6.5 | 2.6 | 40.0 | 2.9 | 0.5 | 17.2 | 1.2 | 0.1 | 8.3 |
| Lethbridge, Can | 2011 | 20 | Diptera (Nematocera) | 1 | 0.6 | 60.0 | 0.4 | 0.1 | 25.0 | 1 | 0 | 0.0 |
| Lethbridge, Can | 2012 | 10 | Diptera (Nematocera) | 6.6 | 2.4 | 36.4 | 1.8 | 0.2 | 11.1 | 1.3 | 0.2 | 15.4 |
| Montpellier, F | 2011 | 10 | Diptera (Nematocera) | 39 | 10.9 | 27.9 | 3.2 | 0.4 | 12.5 | 1.8 | 0.1 | 5.6 |
| Wageningen, NL | 2011 | 10 | Diptera (Nematocera) | 4.4 | 1.6 | 36.4 | 1.7 | 0.5 | 29.4 | 1.2 | 0.1 | 8.3 |
| Zürich, CH | 2011 | 10 | Diptera (Nematocera) | 30.4 | 9.1 | 29.9 | 5.6 | 0.6 | 10.7 | 1.8 | 0.1 | 5.6 |
| Lethbridge, Can | 2011 | 20 | Hymenoptera | 2.1 | 1.2 | 57.1 | 0.4 | 0.1 | 25.0 | 1 | 0 | 0.0 |
| Lethbridge, Can | 2012 | 10 | Hymenoptera | 1.5 | 0.9 | 60.0 | 0.8 | 0.4 | 50.0 | 1.9 | 0.4 | 21.1 |
| Montpellier, F | 2011 | 10 | Hymenoptera | 2.7 | 0.7 | 25.9 | 0.8 | 0.1 | 12.5 | 1.1 | 0.02 | 1.8 |
| Wageningen, NL | 2011 | 10 | Hymenoptera | * | * | * | * | * | * | * | * | * |
| Zürich, CH | 2011 | 10 | Hymenoptera | 6.2 | 1.8 | 29.0 | 1.5 | 0.3 | 20.0 | 1.2 | 0.1 | 8.3 |
| **MIN** | | | | | | **13.2** | | | **5.2** | | | **0.0** |
| **MAX** | | | | | | **60.0** | | | **50.0** | | | **21.1** |
| **MEAN** | | | | | | **29.5** | | | **14.9** | | | **7.7** |
| **MEDIAN** | | | | | | **25.5** | | | **11.2** | | | **8.2** |

* No data available