O'Callaghan and Mansfield. *New Zealand Plant Protection 79* (2026) 17–32
https://doi.org/10.30843/nzpp.2026.79.11798
Published: 19/02/2026

PERSPECTIVE

Evaluating the field efficacy of microbial products – important technical matters to consider

Maureen O'Callaghan and Sarah Mansfield*

Bioeconomy Science Institute, Tuhiraki, 19 Ellesmere Junction Road, Lincoln 7608, New Zealand

**Corresponding author:* sarah.mansfield@agresearch.co.nz

(Original submission received 29 September 2025; accepted in revised form 8 February 2026)

Highlights

- Thorough field evaluation of local and imported microbial products under New Zealand growing conditions is needed to ensure growers can use these products with confidence.
- There are no uniform guidelines for evaluation of the diverse range and functions of microbial products containing live microorganisms, and trials must be designed to accommodate complex interactions among microorganisms, target pest and crop.
- Successful trial outcomes rely on: thorough assessment of prior knowledge of the target and microorganism; detailed pre-trial planning to minimise risk and ensure optimal trial design; well documented trial establishment and monitoring; appropriate statistical analysis of the data; and evaluation of trial outcomes.

Keywords Biofertilisers, biopesticides, biostimulants, live microorganisms, trial design

BACKGROUND

Microbial products such as biopesticides, biofertilisers and biostimulants are often suggested as the ideal solutions to address existing and anticipated gaps in plant health and nutrient management but, at the same time, they are often perceived as less efficacious and more expensive to develop and purchase than synthetic chemicals. Despite these issues, global interest and investment in microbial products is growing rapidly in response to: development of pesticide resistance; escalating costs for development and registration of new chemistries such that the cost of new chemical products is greater than development costs for microbial products; and concerns about the safety and environmental impact of synthetic pesticides and fertilisers. The current global biostimulants and biopesticides market is estimated at US$ 12.9B and is forecast to reach around US$ 29B by 2034 (Precedence Research 2025).

A plethora of biostimulants and biopesticide products (of varying quality) are now available on the global market. The most commonly available of these products can be grouped into three categories, those used for: (i) plant protection (biopesticides and microbial control agents for management of invertebrate pests and plant diseases, respectively); (ii) plant nutrition (biofertilisers); and (iii) enhanced plant growth and stress tolerance (biostimulants). These product categories are regulated differently in most countries, including New Zealand. Biopesticides need to be field tested in New Zealand to meet registration requirements for proof of efficacy and to support proposed label claims relating to target pest(s), crops, application rates and methods (Ministry for Primary Industries 2025). In addition, safety issues must be addressed including safe handling practices for biopesticides, toxicity to humans, vertebrates and other non-target species, as well as environmental and crop residues (Environmental Protection Authority 2025). The manufacturing and quality control process to ensure the product contains the stated levels of the active ingredient (i.e. microbial agent) must also be addressed during product registration but is not the focus of this paper. In contrast to biopesticides, biofertilisers and biostimulants are generally faster and less expensive to take to market because there has been less regulation and fewer data requirements with

©2026 New Zealand Plant Protection Society (Inc.) www.nzpps.org Refer to http://www.nzpps.org/terms_of_use.html

these products globally. This situation is changing and there is increasing pressure to scientifically validate biostimulants and biofertilisers. For example, a voluntary code of practice for the labelling and efficacy claims for biostimulants has recently been introduced by Fertilizer Australia (2025). In the case of biostimulants with claims of abiotic stress tolerance, it can be challenging to artificially create the appropriate field conditions required to detect treatment effects. In this case, field trials might be used to quantify general benefits of the biostimulant in terms of yield, while the specific biostimulant claim related to its mode of action could be demonstrated under controlled conditions (Ricci et al. 2019).

The range of microbial products available in New Zealand is small (nine microbial active ingredients registered as biopesticides, Holden 2025) relative to other regions but is steadily increasing. Some of these products were developed in New Zealand e.g. BioShield®, a biopesticide targeting the New Zealand grass grub (biostart 2025), and the microbial biocontrol agent AureoGold™ for management of a disease of kiwifruit caused by *Pseudomonas syringae* pv *actinidiae* (UPL 2025). Other microbial products have been developed elsewhere and imported into New Zealand e.g. the biostimulant LifeGard® (Arxada 2025), which enhances stress tolerance in a range of crops. Regardless of whether microbial products are developed in country or are imported from overseas, it is important that their efficacy is evaluated in New Zealand as performance of microbial products can be impacted by their interaction with local environmental conditions and/or agronomic practices. When a new microbial product enters the New Zealand market, growers need to receive relevant information regarding the product's specifications, application methods, etc, to ensure the product is used appropriately and the grower knows what to expect. Public perceptions that microbial products are less effective than traditional agrichemicals have arisen partly because growers do not always receive sufficient training or information to achieve the transition from traditional agrichemicals to novel microbial products (Parnell et al. 2016; Marrone 2023). Grower support is particularly important when a new product (microbial or chemical) is target-specific rather than broad spectrum in activity, such that the new product forms part of an integrated pest management (IPM) program that incorporates multiple control strategies. Another factor contributing to poor public perceptions of microbial products is the presence of products in the market that have not been scientifically validated, so called "snake oils". It only takes one bad experience to affect a grower's perception of all microbial products.

Due to the range of microbial products and their diverse modes of action, there are no uniform guidelines defining how they should be evaluated in the field. Some general methods to justify claims of biostimulants for regulatory authorities have been developed (Ricci et al. 2019) but they are not focused on evaluating agronomic benefits to farmers. The technical challenges associated with evaluation of field efficacy of microbial soil inoculants have been reviewed recently (O'Callaghan et al. 2022) and many of the same challenges apply to evaluation of biopesticides and biofertilisers applied by other methods. More recently Neuhoff et al. (2024) have proposed guidelines regarding field trial design and implementation of plant growth-promoting microorganisms (PGPM), including recommendations for the type and scope of data collection and evaluation. As part of their evaluation of product claims made for nano-fertilisers, useful guidelines for conducting robust field evaluation of prototype products have also been provided recently by Frank & Husted (2024), and these guidelines are applicable to design and execution of field evaluation of biofertilisers and biostimulants. Guidance and educational material for growers seeking to conduct research on their own farms is also available (Merfield & Johnson 2016; Sustainable Agriculture Research & Education 2017). Uptake of microbial products is often more successful when it occurs as part of an industry-wide change in crop management practices and accompanied by substantial investment in extension activities, including field trials. For example, the uptake of AureoGold™ by kiwifruit growers was supported by Zespri (Hoyte et al. 2018). This type of investment in extension ensures that growers understand the specific differences in usage between the new microbial product and the agrichemical product(s) they used previously, such as changes in application methods, action thresholds, or seasonal timing.

Early efficacy assessment of microorganisms occurs through laboratory bioassays, growth-chamber experiments and glasshouse pot trials (e.g. Akter et al. 2013). Laboratory and glasshouse trials allow pre-screening to identify promising candidate strains, are useful in characterising microbial modes of action and can provide early guidance on dose rates suitable for future field trials. Laboratory studies and pot trials often use artificial growth media and are run under highly controlled conditions, so they do not closely represent the field situation. Despite following good experimental practice (EPPO Bulletin 2021), it is common for microbial agents that look promising in laboratory and glasshouse pot trials to perform less well when applied in the field (Smyth et al. 2011; Bacillio et al. 2017). Field screening of potential microbial candidates sooner rather than later in the development process can be warranted to avoid investment into less promising candidates. Field trials are expensive however and evaluation of field efficacy takes place typically towards the end of the development process.

Robust experimental design is needed to quantify effects of microbial products, and the effects need to be evaluated over several seasons under a range of climatic/geographical locations and soil types. Effective field-trial design is more complex for 'live' microbial products (such as live bacteria or fungi) than for products with active ingredients derived from microbial sources. For products with live microorganisms, not only must the microorganism remain alive from production to field application, but the field environment (including the background microbiome) must support the natural biological interactions between the microorganism and target pest or plant to achieve efficacy. The greater the natural variability in the field environment, the more difficult it will be to achieve consistent product performance.

Field evaluation of microbial products that contain live microorganisms differs from evaluation of synthetic pesticides and fertilisers in several ways, as discussed throughout the text below. Maintaining the viability of the microbial agent before, during, and after application is crucial to optimising its efficacy so care must be taken at all stages of field trial execution, including careful attention to how and when the product is stored and applied. Live microorganisms may be more sensitive to other agricultural inputs than a chemical active, so compatibility with other components of tank mixes must be considered in advance. Similarly, microbial products may be more specific towards certain life stages of the pest or pathogen than a chemical insecticide, so thorough knowledge of the target's life cycle and careful timing of field application is essential. Choice of delivery system is also important; for example, efficacy may rely on early establishment of the microbial agent in the plant rhizosphere to prevent colonisation by pathogens, or it may be important to apply a biopesticide at a lower pest density than the standard practice, so pre-sampling of trial sites to determine pest densities may be required.

Field testing of microbial products may be conducted individually or collaboratively by a variety of interested parties ranging from academic or applied researchers, industry (including developers, marketers and distributors of products who are very experienced with evaluation of synthetic pesticides and fertilisers), grower levy bodies, independent contractors, through to growers. Whoever conducts the field trials, it is important to consider using methodology that minimises the risk of unintended bias, such as blind treatment application or using anonymised treatment codes during monitoring and analysis. At certain stages of development and commercialisation, microbial product developers may want to consider the merits of enabling independent field evaluation of their products. Sometimes product efficacy information presented on company websites is inconsistent with results from independent field evaluations. In a recent example, independent field evaluation of several novel gene-edited nitrogen-fixing bacterial inoculants (Franzen et al. 2023; Woodward et al. 2025) differed from information provided by producers and distributors of inoculant products. Published independent evaluation can help overcome perceptions of bias, while increasing confidence that the product will deliver as expected.

Here we highlight key challenges and provide suggestions for rigorous field evaluation, with a view to ensuring growers can have confidence in microbial products that have been field tested in New Zealand before entering the market. Specifically, we focus on microbial products that contain live microorganisms to provide context for our commentary and recommendations.

TECHNICAL MATTERS TO CONSIDER

Maximising the potential for success

The definition of success in evaluating efficacy of microbial products will look different for various products and different stakeholders. In general terms, success will be the demonstration of clear and measurable benefits and may also include demonstration of practical "usability" and potentially provide data to allow assessment of economic viability. For biopesticides, success will be the demonstration of effective disease suppression/pest control through measuring repeatable and statistically significant reductions in incidence/severity of plant disease or invertebrate pest populations, compared with untreated and/or conventional treatments (see 'Pre-trial preparation' for discussion of controls). To achieve registration in New Zealand, efficacy data are an essential part of the application. Depending on the trial's purpose (see 'Field trial categories'), practitioners may also be looking to detect improved plant growth and yield, increased marketable crop quality or improved uniformity. Evaluation of biostimulants might also involve detecting positive effects on crop vigour or resilience against abiotic stresses such as drought, salinity, and temperature extremes. Success may also include consistent treatment effects across environmental gradients, geographical locations, and seasons.

Another component to field trial success involves having confidence that the trial results are reliable, even if the desired effect from the product being tested is not demonstrated. Learning from trials that appear to have 'failed' is essential, particularly in the early stages of product development. This learning process includes identifying factors (environmental or experimental) that contributed to the failure, so that either adjustments can be made to protocols for future trials, or a decision can be made with confidence that the product is not suitable for further development.

Any field trial, irrespective of purpose, passes through four key phases (Figure 1). The first two phases (assessment of prior knowledge and pre-trial preparation) culminate in a trial plan, associated protocols for data collection in the field, and a plan for subsequent data analysis. Investment in the first two phases will pay dividends by increasing the likelihood of trial success through thorough planning, including consideration of potential risks and their mitigation. Trial execution is based on the pre-determined protocols, which must be communicated effectively to the team responsible for trial execution, and field practitioners need to be encouraged to record supporting observations, particularly if something unexpected occurs. Such observations may provide clues to important mechanisms or factors that may affect interpretation of trial results. Finally, post-trial analysis and evaluation should go beyond the planned statistical analysis, to include wider consideration of practical problems encountered, general observations, and implications for adoption by end-users. Such evaluation is critical to design of future trials as product development continues (orange feedback loop in Figure 1) as well as preparation of registration packages and extension materials if the product is moving towards commercial production and adoption.

Field trial categories

Before drilling down further into the planning, execution and evaluation of field trials, it is useful to consider the different categories of field trials, as these reflect the progress of a

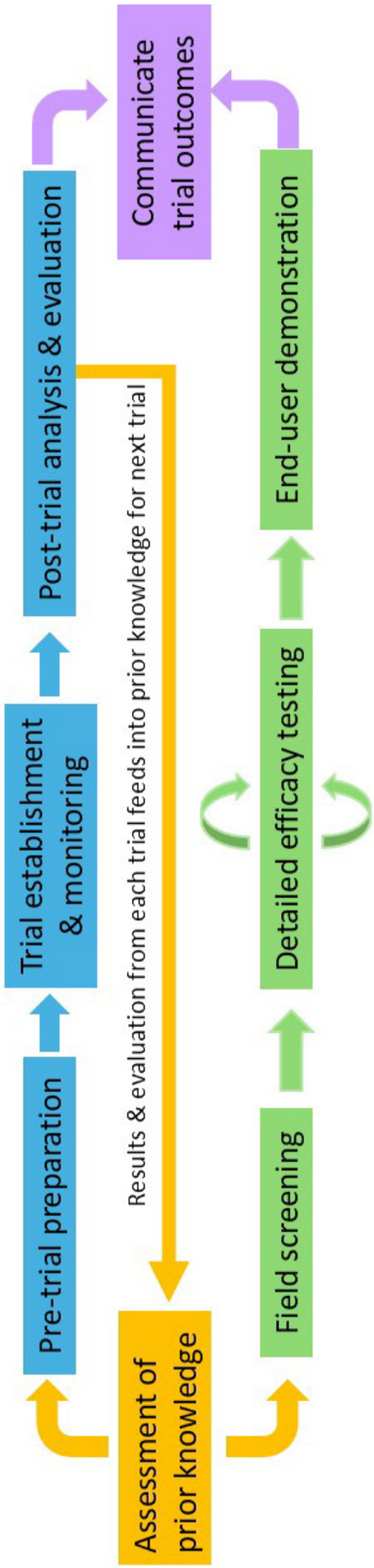

Figure 1. Schematic showing how information/data flows through the different phases of an individual field trial (blue boxes and arrows) and through the different categories of field trials (green boxes and arrows) as development of a new microbial product advances. Information from each completed trial is fed back into the pool of prior knowledge for future use (orange box and arrows). Trial outcomes need to be shared with stakeholders (purple box and arrows) during product development and when the product is made available for commercial use.

potential microbial product towards commercial production and adoption. There are essentially three types of field trials: 1) initial (small plot) screening; 2) more detailed efficacy trials; and 3) end-user demonstrations (Figure 1). Each category of trial may be conducted by researchers, manufacturers, independent contractors, industry bodies, or some combination of these. It is preferable to have appropriate industry stakeholders involved earlier rather than later, if possible, so that trial results can be evaluated from the end-user's perspective.

Initial field screening trials aim to demonstrate that the candidate microorganism(s) can function in a field environment and typically use either a prototype formulation or microorganisms that are unformulated. These trials usually comprise replicated small plots under more realistic growing conditions and will often be located on research farms to meet regulatory requirements for unregistered or novel products. At this stage, the primary goals are to demonstrate proof-of-concept by building on positive laboratory and pot trial outcomes and/or to evaluate several microbial agents for key characteristics in the field. For example, virulence among candidate biopesticide species or strains may need to be compared under more natural conditions than those available in a laboratory or glasshouse (e.g., Narciso et al. 2021, Bacilio et al. 2017). Candidates that show significant activity during screening field trials will progress to the next trial category: detailed efficacy testing.

Once a microbial product begins detailed efficacy testing, it is expected that multiple trials will be needed (some in sequence, some in parallel, see feedback loops in Figure 1). These trials will evaluate key factors impacting on efficacy including application methods and rates, different formulations, potential adjuvants, different crops, etc. Trials should be conducted on commercial farms, subject to regulatory approval, to ensure any issues or conflicts with general agronomic practices are detected and addressed. If suitable commercial farms are not available or regulatory requirements are still too stringent for a wholly commercial setting, research stations/farms can be used instead. Trialling microbial products under commercial field conditions is critical to understand how the new product will fit with current commercial inputs or practices such as crop rotation, and the impact of climatic and environmental factors such as soil type and water quality on performance. These trials provide insights into potential variability in product performance, and where grower adaptation and education may be needed to encourage adoption.

Detailed efficacy trials require larger plot sizes than initial screening trials – this will increase natural variability and provide more realistic outcomes. Replication and suitable controls remain essential to detect treatment effects. Ideally, plots will be large enough and the sampling regime appropriate to allow for commercial extrapolation, i.e. collection of yield data needed to support either cost-benefit analysis or calculation of the expected return on investment from the farmer's perspective. If there are regulatory or manufacturing limitations on the quantity of microbial product available, plot sizes and/or application methods may still be constrained at this stage. The results from each efficacy trial will inform design of subsequent trials. It may

be tempting to try and incorporate multiple experimental factors in a single efficacy trial. Overloading a single trial with too many treatment variables is often detrimental to achieving clear outcomes due to low statistical power (see 'Pre-trial preparation'). It is also important during these trials to widen the geographic/climatic scope of the trial sites so that efficacy data is available from a wide range of environmental conditions. The goal of detailed efficacy testing is to evaluate a (near) final product and associated recommendations for use. For biopesticides, these trials should generate the data needed for formal registration of a first-generation product. It is likely to take several rounds of trials to complete efficacy testing and progress to the last trial category, end-user demonstrations.

When a product is deployed in an end-user demonstration trial, it should be very close to commercial release, if not already available. These trials are typically large in scale and unreplicated, although similar trial designs may be implemented at multiple sites to provide geographic representation. The value of these trials lies in assisting uptake, development of extension materials, and building farmer confidence in the new product, while demonstrating product performance and agronomic compatibility under typical growing conditions alongside commonly used inputs, such as other pesticides, nutrients etc. Large-scale demonstration trials are particularly important when regulatory or manufacturing constraints have limited the scale of previous efficacy trials (as discussed above). This is also an opportunity to collect additional data such as more realistic yield measurements, particularly if part of the trial site remains untreated, although growers may be unwilling to do this if substantial yield losses are likely. Alternatively, a demonstration trial can provide a large-scale comparison with existing products already in use.

Regardless of whoever is running efficacy trials, in the following sections we review key points to consider during each field trial phase, while identifying similarities and differences between the trial categories. Examples drawn primarily from New Zealand field trials are used to illustrate the discussion, supplemented with examples from other jurisdictions where relevant.

Assessment of prior knowledge

The primary purpose of this phase is to collate and understand what is already known about the numerous factors that can affect field efficacy of the microorganism(s) to be tested and therefore affect the likelihood of a successful field trial. This prior knowledge may come from scientific literature, laboratory experiments, pot trials, and results from previous field trials. The latter will become increasingly important as more field trials are completed, but it may be necessary to return to laboratory experiments and/or pot trials if there are unexpected findings from field trials that need to be (re)-investigated under controlled conditions. Topics to review during this phase are:

- *Biology of the target microorganism(s)* including its mode of action, dispersal mechanisms, reproduction, and environmental persistence.
- *For biopesticides, biology of the target pests/pathogens* including life cycles, seasonality, generation time, natural enemies (particularly those that might affect field trials), damage thresholds and/or population density, geographic range and habitat/climatic preferences. For example, there are seasonal differences in the mobility of adult black beetle (Mansfield et al. 2016) and their physiology that are linked to seasonal differences in efficacy of both a conventional insecticide and a prototype biopesticide (Mansfield et al. 2020, also see Box 1).
- *Biology and agronomy of the target crop(s)* where the product is likely to be used including seasonal timing of planting/harvesting, crop stage affected by target pest/pathogen, crop stage for product application, economic thresholds, soil types, climatic preferences, irrigation, other pests/pathogens likely to be present, and current usage of chemical pesticides or fertilisers that may interfere with proposed field trials.
- *Environmental factors* including temperature, rainfall, sunlight, seasonal cycles, soil moisture, soil type, soil pH, local topography, and how these may interact with the target microorganisms, pests/pathogens or crops. Are there any regions where poor climate/habitat matching may affect product performance? For example, pot trials indicated that irrigation and/or rainfall were likely to influence dispersal of the bio-insecticidal bacterium *Serratia entomophila* into the soil profile after application. Field trials across sites that happened to experience different levels of natural rainfall confirmed this hypothesis (Wright et al. 2017).
- Ideally, trials will be located at sites with a fairly homogeneous distribution of *soil properties* and a known history of crops and management including fertilisation and crop protection. Where known gradients in soil properties or topographical factors cannot be avoided, appropriate experimental design must be used.
- How do each of the factors listed above contribute to *natural background variability* within the agroecosystem? What impact will natural variability have on statistical power?
- No one has full knowledge of all potential factors that affect field trials. When gaps are identified in prior knowledge, *what assumptions will be made* to cover these gaps? Is there potential for these assumptions to be tested as part of the proposed field trials?
- When a field trial is analysed and evaluated, any new knowledge gained should be incorporated to *ensure that the knowledge base remains current*. This should include explicit review of any necessary assumptions so that these can be changed or confirmed, depending on trial findings.

Pre-trial preparation

Variability at field scale necessitates multi-year, multi-site field data to obtain adequate statistical significance for separation of treatment effects and ensure robust analysis (Leggett et al. 2015). It is this natural variability that makes investment in pre-trial preparation worthwhile. In this phase, the overarching purpose of the field trial should be defined and hypotheses specified, taking into account the intended

Box 1. Potential for a biopesticide bait to control black beetle, *Heteronychus arator* (Coleoptera: Scarabaeidae) (Mansfield et al. 2020)

Aim: Development and evaluation of a prototype biopesticide bait for management of black beetle under field conditions in New Zealand.

Issues addressed by trial design: (1) Need for regulatory compliance during first screening trial including containment of the microbial active and treated insects; (2) naturally low pest density affects statistical power; and (3) timing and scale of efficacy trial to accommodate seasonal differences in biology as well as the high likelihood of strong edge effects due to beetle mobility.

Trial design: Two screening trials replicated across three sites in both spring and autumn using small, contained arenas (320 mm in diameter, n = 6 arenas/treatment) arranged in a randomised block design. These trials included a negative (untreated) control and a positive (chemical insecticide) control alongside the biopesticide bait containing *Yersinia entomophaga*. Adult *H. arator* were field collected and added to the arenas (n = 10/arena) to ensure sufficient beetles were exposed to each treatment to detect any effects. The larger scale efficacy trial was conducted in spring using open plots (40 x 40 m) with only a negative (untreated) control and the biopesticide bait. This trial was replicated across three sites with 3 plots/treatment at each site. Beetle numbers and mortality were monitored using pitfall traps followed by laboratory incubation. The following summer, soil samples were taken in each plot to measure larval density.

Conclusion: The biopesticide bait caused substantial mortality compared with both negative and positive controls in the spring screening trial but mortality was much lower during the autumn trial for all treatments. This was attributed to differences in beetle physiology between newly emerged adults (autumn) and older, reproductive adults (spring). The larger scale efficacy trial demonstrated a 2x reduction in beetle numbers and significant increases in beetle mortality from the biopesticide treatment. This led to a significant decrease in larval numbers the following summer.

use of the product, if this is known. It is important at this stage to be certain that the unit of replication (for a field trial this is often one plot) is appropriate relative to the hypothesis, and how the results of the specified trial will relate to the wider context of the work. For example, will this trial alone be sufficient to generalise conclusions to all situations, or is this one of a series of trials needed to establish a broader knowledge base? When these are clearly defined, decisions can be made to address the following points:

- *Ensure regulatory requirements are met when applicable*. This issue is particularly relevant if crop destruction is needed post-trial, compliance to achieve complete destruction must be negotiated with the landowner and compensation for crop losses may be necessary. Early field trials (such as screening trials) may be best placed on research sites where regulatory compliance is easier to achieve and crop destruction or poor performance of the product(s) being tested will have less commercial impact.
- *Assess potential trial sites for suitability*. Pilot sampling may be needed before site selection can be confirmed. Site selection should be considered in terms of both the individual trial and the collective data package. Pre-trial assessment of field conditions is essential to determine if, for example, the target pest is present at sufficient density (above the economic threshold) to allow detection of treatment effects. Pre-trial sampling is a common requirement for grass grub (Townsend et al. 2004) and porina field trials (Hurst et al. 2019) because these pests do not necessarily reach damaging levels every year (Jackson et al. 2012). When there is a target for minimum pest density to ensure any treatment effects are likely to be detected, and pest density is likely to be highly variable, it may be feasible to augment the population by adding a specified number of lab-reared or field collected individuals. Augmentation was used in mini-plot field trials with adult black beetle because this pest is highly mobile and occurs naturally at low densities (Mansfield et al. 2020, see Box 1 for more detail).
- *Properly characterise soil fertility when evaluating biofertilisers*. Even synthetic fertilisers do not always deliver expected treatment effects, and the most efficacious microbial product will not give a response if the constraint it is addressing is not present at the field site. It is often assumed that biofertilisers will be more effective under low input conditions, but effects of low or high soil nutrient contents remain inconclusive and likely differ between microbial species with differing modes of action (Schmidt & Gaudin 2018; Lopes et al. 2021). Some studies report no interactions between microbial inoculation and fertilisation or even improved nutrient utilisation by plants if fertiliser is applied with a microbial product, or crops are inoculated in soils with high nutrient contents (Galindo et al. 2022).
- *Address possible confounding factors*. Irrespective of the trial's purpose, the pre-trial survey needs to determine if there are any pre-existing gradients in terrain, soil condition, or crop status that must be addressed in the trial design so that treatment effects can be detected reliably and to ensure effective statistical analysis of the data generated.

- *Include negative and positive controls*. Inclusion of an untreated (negative) control in the trial design is an obvious necessity but also requires some thought. Negative controls may include not only wholly untreated plots but also plots sprayed only with adjuvants, or plots that are drilled without any inoculation. Inclusion of a "formulation blank" control (application of product containing all formulation ingredients but without active microorganism(s)) is desirable but not always possible when the full details of a commercial formulation are not disclosed (Bashan et al. 2020). Failure to include this control treatment can result in "unfair" comparisons between an unformulated prototype microorganism against commercial products with formulations optimised for extended field persistence, adherence to foliage, UV protection, etc. Positive controls will usually include a currently used product that targets the same problem as the novel treatment. Most biopesticide field trials aim to determine the efficacy of prototypes or commercial products against an untreated control and the current industry standard treatment, usually a synthetic pesticide, but sometimes a commercially available biological product (e.g. Hurst et al. 2019, Narcisco et al. 2021, Peiris et al. 2021). If specific effects are attributed to the microbial product, e.g. improved nitrogen supply or nutrient use efficiency, a mineral nitrogen application could be included as a positive control – possibly at lower and recommended rates – allowing the calculation of nitrogen fertiliser equivalence (Neuhoff et al. 2024).
- *Choose an appropriate spatial layout of trial plots*. Potential designs include simple random, randomised complete block (a common design choice, it is important to ensure blocks are aligned to a pre-existing gradient, such as a sloping paddock), Latin square (addresses two orthogonal gradients – less common in field trials), or split-plot (allows paired comparisons of adjacent treated/untreated plots, e.g. Shi et al. 2023 (see Box 2 for example of use of split-plot design), and Hett et al. 2023 (see Box 3 for split- plot design used over multiple location and seasons). Pre-trial site assessment and/or sampling is an essential step for choosing the most appropriate spatial layout that should be considered alongside any prior knowledge regarding the spatial distribution of important experimental factors. Quinn & Keough (2002) reviewed different trial designs and when these should be used.
- *Use appropriately sized plots*. Plot size can impact on the results of field trials, due to the potential for edge effects (e.g. due to pest/pathogen mobility or micro-climate effects at plot edges). Plots must be sufficiently large to allow measurement of treatments on disease severity, insect damage or plant growth. Plot size should also allow for repeated assessments and collection of plant material if required, as seen in a mycoherbicide field trial (Bourdôt et al. 2006). If there is potential for accidental drift or cross-contamination among plots, buffer zones may be required to maintain independence between treatments. For example, Shi et al. (2023; Box 2) used 2 m buffer zones between the main plots treated with different rhizobia strains. Analysis of multiple field trials measuring the impact of a phosphorus-solubilizing inoculant on maize yield found that yield increases and variance differed with plot size, with higher yields measured in large plots, with less variability (Legget et al. 2015).

Box 2. Use of split-plot design in evaluation of rhizobial inoculants for white clover (Shi et al. 2023)

Aim: To identify *Rhizobium leguminosarum* bv. *trifolii* isolates suitable for use in New Zealand pastures.

Issues addressed by trial design: New Zealand pasture soils contain background populations of naturalised rhizobia that vary significantly in both size and nitrogen-fixing ability and compete with new inoculants for clover nodule occupancy. The spatial distribution of these background rhizobia populations is highly variable between sites and even within a paddock.

Trial design: Three field trials were conducted over two years. In two trials, three rhizobia strains (applied as seed treatments) were compared for their ability to enhance clover growth against uninoculated control plants, and in the third trial, five rhizobia isolates were compared against an uninoculated control and the currently used commercial inoculant strain, also applied as a seed treatment. Uninoculated control seed received the same formulation minus the inoculant strain. Pre-sampling of the trial sites showed that background populations of rhizobia of various sizes and levels of effectiveness at N fixation were present at all sites. To address this variability, a split-plot design was used with rhizobia isolates (New Zealand isolates or commercial control strain) forming the main plots and inoculation (plus or minus) forming the sub-plots. Sub-plots were established as two parallel rows to allow direct pair-wise comparisons between each rhizobia treatment and the uninoculated control. Six replicates were established at each trial site, with the main plots arranged in a randomised complete block design. Above-ground biomass was sampled through growing seasons.

Conclusion: High variability in clover biomass was detected in all three trials. The impact of heterogeneity among treatments was somewhat mitigated by using the pairwise line sowing strategy and statistically significant responses to inoculation were detected (increased biomass and nitrogen fixation) in some treatments and at some sampling timepoints. Given the high level of variability in both the background rhizobia populations and yield responses, the authors recommended more replicated blocks would be needed to more reliably assess the performance of rhizobia isolates in future field studies.

Box 3. Comparing effects of microbial consortia against fertilisation regimes on maize growth (Hett et al. 2023)

Aim: to test hypotheses that: Maize inoculated with novel microbial consortia composed of collection of beneficial microorganisms, or a commercial microbial product, outperform non-inoculated plants in vegetative growth and grain yield; Effects of inoculation on crop performance decreases with increasing fertilisation; Inoculation leads to higher crop nitrogen and phosphorus contents.

Issues addressed by trial design: Measurement of significant, agronomically relevant and reproducible benefits from inoculation of maize across sites, seasons and fertiliser inputs.

Trial design: Trials were run in two consecutive years at different locations, resulting in three trial environments. Four microconsortia treatments were evaluated in factorial combinations with three fertilisation levels (unfertilised; 110 and 200 kg nitrogen per hectare). Viability of the consortia was confirmed before application. Split-plot experiments were performed over two consecutive seasons at three locations. Climatic conditions were well monitored. Growth and nutrient content measured at stem elongation and final yield measurements were made at crop harvest.

Conclusion: Inoculation increased early growth parameters, with crop growth, shoot nitrogen and phosphorus uptake all significantly increased in some treatments, but only in one season. The authors attributed this difference to "distinct weather and pre-crop effects", with previous cultivation of legumes leaving a supply of nitrogen at one site, which presumably provided less scope for further growth improvement from inoculation. An interaction between inoculation and fertiliser application was not detected, so the assumption that inoculants are more effective in low input systems was not confirmed. In this study, effects of fertiliser application on crop growth and nutrition were more consistent than the impact of the selected consortia in both years.

- *Choose an appropriate application method and rates for the microorganism and any positive controls.* Efficacy of a microbial product can be influenced by application method so when being compared with currently registered products, they should be delivered by the same mechanism to ensure a 'like for like' comparison (any machinery used to apply treatments should be cleaned thoroughly between each application to avoid cross-contamination). For foliar applications, factors such as spray volume, droplet size, and coverage play a crucial role in the distribution and effectiveness of the biopesticide on the target pests (Chandler et al. 2008). These key variables may be different for microbial products compared with traditional agrichemicals. Ensuring that the application method is properly calibrated is important to achieve adequate coverage and contact with the target pests/pathogens. When a product is in the 'detailed efficacy' phase, application methods may be tested in targeted trials if there are multiple options available for the microorganism under study or if different methods are needed for different crops. For example, autumn-sown arable crops are protected from grass grub with neonicotinoid seed treatments, so seed treatment with the bacterium *S. entomophila* was an appropriate comparison in the field (Mansfield et al. 2017). In contrast, established pasture is often drilled with insecticide granules for grass grub control, so comparison of drilled diazinon and Bioshield™ granules containing *S. entomophila* was appropriate for that situation (Zydenbos et al. 2016).
- *Determine the variables to be measured and associated methods to gather these data*. Variables may include infection and/or mortality rates, crop damage and/or yield, crop establishment, crop nutrition, environmental persistence of the active microorganism, and climatic or soil variables that may affect product performance. Selected variables should address the hypothesis being tested either directly or indirectly, as well as data required for product registration, such as crop safety. For example, mortality from a biopesticide is a direct measurement of treatment efficacy whereas crop yield from a plot treated with a biopesticide is an indirect measurement of treatment efficacy. Including both direct and indirect variables is likely to provide more information, particularly if trial outcomes are not what was expected.
- *Generate a sampling strategy*. Include the number of samples to be taken in each plot and the sampling plan. Samples are often taken randomly within plots, but systematic sampling may be more appropriate to ensure adequate plot coverage, particularly when the organism or variable being sampled has a patchy spatial distribution. Southwood & Henderson (2000) discussed sampling strategies in relation to the target species' distribution in the field. Some horticultural pests, such as whitefly, can pose challenges for sampling design in field and glasshouse trials. A key decision is whether to count every individual on a plant or count the number of infested plants instead (Naranjo et al. 1997; Spinner et al. 2011).
- *Collect growth data*. Phenological stages of crop growth are often monitored in evaluation of plant growth-promoting microorganisms. Non-destructive methods such as plant height are easily measured and may give early indications of beneficial effects, but for destructive sampling, more plant material will be collected so larger plots may be needed to avoid resampling the same plants/area within a plot. Higher plant biomass and

nutrient accumulation at a given phenological stage are also useful measures of plant growth promotion (Neuhoff et al. 2024) but ultimately crop yield and quality will be of most interest to growers. Fresh and/or dry matter yield need to be collected from a representative (e.g. plot) area, and additional crop-specific measures can be made e.g. ear density, grains per ear etc for winter wheat.

- *Collect secondary data*. Field trials also provide the opportunity to collect additional data on microbial products. For example, crop health can be monitored to confirm the absence of phytotoxic effects under standard growing conditions (a common regulatory requirement for new pesticides) and potential beneficial effects such as greening and growth promotion. Non-target effects, if any, on beneficial species such as pollinators and natural enemies of other pests may be monitored to enhance data from laboratory-based toxicity tests.
- *Monitor persistence*. If needed for registration, persistence of the microorganisms in the environment can be monitored, which will also indicate if repeated applications may be needed to ensure efficacy. This process relies on availability of precise methods to differentiate the applied strain(s) from the naturally occurring resident microbial communities in soil or on plants and to measure viability. Tracking the fate of the applied microorganisms in the environment post-application may help in functional explanation of field trial results.
- *Consider the impact of other crop management practices on trial outcomes in advance of trial execution*. It may be necessary to omit application of fungicides or insecticides, except where these may be used as positive controls. For example, in the first year of field trials using *S. entomophila* as a seed-coat treatment, fungicide seed treatment was omitted due to concerns about compatibility. Once laboratory tests indicated no adverse effects, a standard fungicide seed treatment was incorporated (Mansfield et al. 2017). Where necessary, e.g. during evaluation of stress-protective effects of biostimulants, growing conditions that differ from standard practice may need to be imposed. For example, in evaluation of drought tolerance, rain shelters can be used, or plant performance can be compared under irrigated vs rain-fed conditions. The experimental protocol should include information and dates of all crop management practices.
- *Use appropriate statistical tests*. Proposed statistical analysis must account for the factors listed above, particularly the number of treatments and replicates, the variables measured, and whether any of the variables will be measured repeatedly over the course of the trial. While analysis of variance (ANOVA) is widely used, for many biological variables other generalised linear models will be more appropriate to the characteristics of the data. Quinn & Keough (2002) provide a good introduction to generalised linear models and other statistical analyses for biological data. Consultation with a statistician is highly recommended during the pre-trial phase to ensure the overall trial design is aligned with the trial aims and hypotheses, that the correct error term is used for your statistical model, and that the data collected can be analysed effectively with no risk of pseudo-replication.
- *Assess statistical power* (the likelihood of detecting a specified effect size), considering natural variability, the number of treatments, and the resources available, to determine the number of replicates. High natural variability will reduce statistical power, meaning more replicates will be needed to detect treatment effects. For example, to detect a 10% increase in yield from fertiliser trials with two treatments requires 9-28 replicates per treatment, depending on the level of natural variability (Johnstone & Sinclair 1991). In practice, when there are often constraints on funding and labour, it may be helpful to reduce the number of field treatments through prior laboratory experiments and/or pot trials. For example, a field trial used for initial screening should focus on a single high application rate for each novel microorganism (or strain) being tested, using the most effective application method identified from prior knowledge.
- *Increase statistical power in later trials.* Once a product/candidate microorganism has passed the field-screening stage, and more detailed efficacy evaluation is underway, greater statistical power is needed to distinguish between treatments because the effect size will usually be smaller. At this stage, it is important to avoid testing too many factors in a single trial, especially when those factors are likely to interact, because multi-way factorial field trials are very difficult to design and interpret effectively. It may be helpful to use prior knowledge to prioritise which factors are tested first (starting with the factor(s) that are expected to have the largest effect size) and conduct further trials sequentially to determine the role of each factor tested. It may also be helpful, if funds and resources permit, to increase plot size during this phase. Larger plots do not necessarily mean fewer replicates but can improve statistical power when natural variability is likely to be high and effect sizes are small. We recommend choosing the simplest experimental design that will meet the needs of the project.
- *Review data and trial strategy regularly*. As more field trials are completed, it is important to review the overall scope of the dataset to ensure that trials are conducted consistently across the full geographic range where the product is likely to be used, and to gather data across multiple years (e.g. Leggett et al. 2015). If the product is likely to be used during more than one season, then trials need to cover all seasons when the product may be used. For example, if the target pest is active from spring to autumn, would the product be recommended throughout that time or only during specific seasons?
- *Collect additional data to support product adoption*. If the main purpose of the trial is demonstration of product efficacy to end users, then it is important to consider what information, or data will be most useful

to support adoption. Are there additional data that will support development of user guidelines for the product or to support marketing? This may not be the same type of data that was collected during the detailed efficacy trials, particularly when specific data are needed for product registration. Consultation with industry partners is important for demonstration trials. For example, growers are often more interested in crop yield or economic gains than pest numbers. In this situation, fewer replicates (if any) and very large plot sizes are often more useful so that treatment effects are readily visible on-site. This will limit the potential for statistical analysis, although comparison between demonstration sites may be feasible if multiple sites can be established simultaneously.

- *Assess potential risks involved* in running a successful trial and *identify mitigation strategies* where feasible (see 'Minimising risks' section for more information about risk mitigation).

When pre-trial preparation is completed, a trial plan should be ready that incorporates all necessary protocols, equipment, and labour, as well as an approximate timeline for trial execution and an initial plan for post-trial analysis and evaluation. Execution of a field trial does not always go to plan, but investment in this pre-trial stage will help to mitigate risks and to salvage results when disaster strikes.

Trial establishment and monitoring

This is the crucial phase when the field trial is established, treatments applied, and the data gathered to determine treatment effects. The effort required to organise the logistics of staff, equipment, transport, and appropriate product storage, etc should not be underestimated at this stage. During this phase, the biggest challenge will be unexpected events/interactions that may change trial outcomes or cause substantial losses from damage to the crop itself, associated pests/pathogens, etc (see 'Minimising risks' section). There is almost always something unexpected during trial execution – documentation and, if necessary, adaptation of the trial plan will in most cases ensure that the trial is completed. Key issues to consider include:

- *Consider potential issues with water quality* if used as part of the application method. Some water supplies are chlorinated which can reduce survival of microorganisms.
- *Calibrate equipment before treatment*. This step is crucial to ensure the correct dose is applied; it may also be worth checking this throughout the application process to confirm consistency of dose.
- *Ensure any potential incompatibility* has been identified with other inputs used as part of standard agronomic practice during trial planning, and measures to mitigate them should be used during application of microbial products.
- *Confirm viability of microorganisms* throughout trial establishment by collecting samples of the microbial product pre-, during, and post-application, providing useful information on environmental stability of the product and confirming planned application rates.
- *Document any changes to the trial layout*. When the trial is established, it is essential that any changes made to the plot layout, due to unexpected problems, are documented thoroughly and the trial plan updated to reflect these changes. For example, if plot size needs to be adjusted during treatment application, or a mistake is made with plot layout, these need to be recorded so that the subsequent data analysis is adjusted to reflect reality in the field.
- *Document any changes to the monitoring protocol*. Changes need to be documented in the trial plan, and the reasons for changes noted. A common example is changes to planned sampling dates due to bad weather.
- *Collect additional data in the field or from other sources if something unusual occurs* (e.g. localised weather events) *or is observed* (e.g. unexpected signs of phytotoxicity). Field practitioners should be encouraged to report such events and respond appropriately. This may be as simple as taking photos on-site or collecting weather observations from a local weather station, but in some cases, may require more substantial data collection. Unwanted non-target effects such as phytotoxicity are a case in point because the cause will need to be identified and addressed as part of continued product development.
- *Maintain communication with the farm owner or manager*, in case agronomic factors mean there are unplanned impacts on the trial site, and to ensure the trial site is maintained appropriately. Common examples include outbreaks of pests or pathogens unrelated to the trial that require control measures or sudden weather events that require harvest dates to be changed at short notice.
- *Monitor any regulatory conditions associated with the trial to ensure compliance*, including clear demarcation and robust labelling of the trial area and individual plots, as well as crop destruction and appropriate disposal of crop residues, if this is required.
- *Collect, process and store field samples and data appropriately for future analysis*, including any processing of field collected samples that is done under laboratory conditions. This includes analysis of soil or foliage samples to measure persistence of microbial treatments or observed mortality of field collected insects held under controlled conditions.
- *Record any new information or observations* that are revealed during end-user demonstrations, which are usually the largest-scale field trials. The aims of the demonstration should be set during pre-trial preparation including any data to be collected, however data collection may be less quantitative and more observational than for screening or detailed efficacy trials. Instead, the focus may be more on stakeholder feedback from site visits, or informal comparisons of yield data from the demonstration site in discussion with the farm owner.

Successful (more or less) trial execution marks the transition to the final phase where results are analysed and evaluated in relation to the wider biological and agronomic context for the work.

Post-trial analysis and evaluation

Once all trial data have been collated, checked and visualised, consultation with a statistician is recommended to determine if the proposed analysis from the trial plan is still suitable. At this point, adjustments may be needed if unexpected events have affected trial results. For example, if a single replicate has been lost or damaged, the experimental design will be slightly unbalanced, but usually this can be addressed with minimal impact during analysis (see 'Minimising risks' section). In addition to testing the original hypotheses, this is the opportunity to look for unexpected interactions (which may be positive or negative) among experimental factors and the environment – the points listed below are intended to stimulate discussion while working through analysis of the results.

- *Hypothesis testing* is just the start because statistical significance does not necessarily mean biological or economic significance. If a statistically significant result ($P < 0.05$) is found, it is important to look at the magnitude of the differences between treatments (compare means/medians) and the level of variability (how big is the standard error or confidence interval?) to place that significant result in an appropriate context. If it was a factorial trial design, are the interaction effects significant or not, and what does that mean biologically?
- If statistical analysis identifies factors or interactions with marginally significant results ($0.05 < P < 0.1$), again it is important to *compare treatment means and the level of variability* to determine if there is a biological trend in the data that warrants further investigation. This may be particularly useful for making progress with detailed efficacy trials, if no treatment stands out as more effective than the others.
- If a trial has no significant treatment effects, and assuming this is not due to unexpected events damaging the trial, a *post-hoc power analysis* may help to determine the likely detectable effect size. The outcome of power analysis can be used to improve the design of future trials.
- When the statistical analysis is completed, *the outcomes for each hypothesis should be reviewed* with respect to the overall purpose of the trial. For a screening trial, which novel treatments showed significant impacts and warrant further investigation? Which do not warrant further investment? For detailed efficacy trials, other commercial factors may need to be considered, such as cost of production, when deciding which treatments to investigate further or to recommend for adoption.
- *Did the outcomes reflect pre-trial expectations?* If not, why? Were there extenuating circumstances such as extreme weather events that need to be considered? What does review of prior knowledge indicate as possible explanatory factors for unexpected outcomes? Are adjustments needed to future trial design to incorporate factors identified in the current trial?
- *What new hypotheses have been generated* for testing in the next trial? Does the next round of field trials need to incorporate new experimental factors, variables, treatments? Is the product ready to be used in demonstration trials?
- Demonstration trials are less likely to be subject to formal statistical analysis, but if site or treatment comparisons were planned, *the analysis should be reviewed* considering the points raised above. Evaluation of field observations, yield, agronomic issues, stakeholder feedback, and any other findings that may impact on product performance or adoption are important to get the most out of end-user demonstrations.
- *Communication of trial results* (successful or otherwise) and their implications to industry stakeholders is an important part of trial evaluation because it provides an opportunity to get feedback from the intended end-users (see 'Communicating field trial outcomes').

Trial analysis and evaluation mark the end of one field trial and potentially, the beginning of planning for the next trial. The decision to establish a new trial to build on previous results will also depend on stakeholder needs, funding, and whether the target microbial product is still demonstrating suitable characteristics to warrant continued investment. In the following sections, we address risk management, and the role of communication with stakeholders and the wider scientific community, in relation to microbial product field trials. These two aspects are vital for successful field trial delivery so warrant more in-depth treatment here.

Minimising risks – what could possibly go wrong?

Field trials are always at risk from unexpected events, impacts, interactions, etc. Some risks can be mitigated with good trial design, planning and execution, or can be mitigated afterwards by appropriate adjustments to the analysis and evaluation of results, provided these adjustments do not introduce bias. Most field scientists at some point will have to write off a field trial due to circumstances beyond their control. Here, our aim is to outline options that can help to salvage some results when things go wrong or when it may be best to end the trial altogether and, if possible, learn from the experience.

- The first step is to establish how much of the trial site has been affected by the unexpected event (e.g. flooding, accidental grazing, unexpected locust plague). As noted above ('Post-trial analysis and evaluation'), if only a single replicate is affected within a well-replicated trial, the statistical analysis can be adjusted to mitigate this.
- If multiple replicates from one treatment or one block have been affected, it may be best to remove that treatment or block altogether from the analysis. Again, if there has been sufficient replication this mitigation should be feasible, unless it is the untreated controls that have been damaged (see next bullet point). Depending on the nature of the unexpected event and its potential impact on the data, it may be helpful to analyse the

whole dataset as originally planned and then analyse it without the affected treatment or block. Sometimes this type of comparison will give an indication of the severity of the unexpected event. It can then be decided whether the result was biologically significant and, therefore, exclusion from analysis is warranted, or if exclusion is not necessary. This does not mean, however, that outliers can be removed from the dataset just because they don't fit the overall trend. Extreme data points should always be checked for data entry errors but if the data are correct, and there has not been a known issue that impacted the trial, then the data reflect natural variability.

- If the untreated controls have been affected by contamination or some other unexpected event, then this will have a substantial impact on the field trial and almost certainly prevent effective statistical analysis. In this situation, the best outcome will be to learn more about the biology of the microbial agent, its interaction with the target pest/pathogen/plant, and the impact of environmental factors. For example, field trials to test the efficacy of a fungal pathogen against a scarab pest discovered the fungus was able to disperse well beyond the treated plots, including into the untreated controls (Villamizar et al. 2024). While this was disappointing in terms of the trial analysis, it provided useful biological information regarding the dispersal and environmental persistence of this pathogen to inform future work.
- During the trial or post-trial analysis, an unexpected factor is detected at the trial site. For example, there are non-uniform soil characteristics or crop history that were not recognised during the pre-trial survey. Consultation with a statistician is highly recommended in this situation because a possible mitigation is to measure this variable (this can be quantitative or qualitative) and incorporate it as a co-variate into statistical analysis. If measurement is not possible (therefore preventing mitigation) then this unexpected factor will increase natural variability within the trial and is likely to reduce statistical power.
- Pest /disease pressure proves to be too low to detect treatment effects. This should have been checked during the pre-trial survey but for some pests or pathogens, the population may not develop as expected during the trial. One possibility is to augment the pest/pathogen population to an appropriate level (only feasible if problem detected early and augmentation is permitted by farm owner). Are there other variables that may indicate treatment effects, e.g. plant/crop characteristics, or can persistence of the microbial product be measured? This would provide some information to improve product performance in the field environment, even if efficacy data against the target pest/pathogen will not be available.
- Pest pressure is higher than allowed for during pre-trial planning and too high for products to be effective at the rate(s) and schedule(s) being trialled. If pest/disease pressure is very high, an effective product should still cause a statistically significant reduction in pest numbers/disease incidence, even if that reduction is not sufficient to meet the desired threshold. This finding can be addressed as part of the planned data analysis and interpretation. If the product has no detectable impact on very high-pressure situations, it is unlikely to be suitable for commercial use, at least in the formulation tested during the trial. Again, such a finding can be addressed during data analysis.
- Number of samples and/or replicates were too low to achieve statistically significant results (despite apparent trends), either due to the original trial design or after a subset of replicates were removed after unexpected problems occurred. It is important to review all variables measured in this situation because some may be less affected and still provide useful data. It is worth considering retrospective power analysis (noted above in post-trial section) to improve future trial design.
- The timing of unexpected impacts relative to the duration of the trial is an important factor to consider. If the trial is damaged/destroyed soon after it began, is it possible to restart from the beginning? For example, if the crop fails to establish, can it be resown? On the other hand, if the trial is damaged close to the planned end date, can you analyse the data gathered up to that point, i.e. end the trial early? If the whole trial site is affected significantly and the impact occurs before data were collected but too late to re-establish the trial, then it will have to be written off.

Communicating field trial outcomes

In the early stages of product development, communication is likely to be constrained by IP considerations and/or commercial considerations. Ideally, a stakeholder advisory group or industry partner(s) will be able to provide feedback to inform early trial design, delivery and analysis, while maintaining confidentiality until product development is more advanced. Wider sharing of trial outcomes with stakeholders should take place alongside end-user demonstrations (if not before), so that development of extension materials, product labels and supporting documentation can incorporate feedback from the target market.

When communicating trial outcomes, disclosure of as much scientific and technical information as possible supports greater scrutiny and understanding of field efficacy (Bashan et al. 2016). For example, product information sometimes reports percentage increases in yield but does not provide data on actual yield, so it is not clear if crop growth has been constrained in any way. Similarly, the nature of the untreated control and positive controls is not always clear in product dossiers, and commercial sensitivity is often the reason cited to justify less than full disclosure of product information. Where it is possible to publish or share trial outcomes, it is helpful to clearly define conditions under which the trial was conducted – season, soil type, fertility, moisture, crop, etc. This information helps to define the conditions where consistent benefits can be expected. This information can also be helpful in informing expectations around anticipated efficacy levels, potential

variability in performance, and educating growers about optimal conditions for use. In addition to data on efficacy, where possible, it is useful to compare marketable yield with untreated controls and provide information that can be used in return on investment or cost-benefit analyses so that end-users can evaluate the suitability of the product for their situation.

Formal scientific communication is also important. Compared to publications reporting efficacy of microbial candidates and products from laboratory and glasshouse trials, there are relatively few published reports of field evaluation of microbial products. This is due, in part, to the lower quality of data often obtained from field trials and potential confounding factors compared with trials in controlled environments. Some journals, such as *New Zealand Plant Protection* offer a specific type of submission for applied practices, such as field trials. Registration of biopesticides requires provision of significant field efficacy data but this often seems to remain unpublished (often due to commercial confidentiality issues), even though there is substantial value in providing independently peer reviewed published data to regulatory authorities during registration of products. Some reluctance to publish could be attributed to commercial sensitivity but once IP protection is in place, it would be helpful to see more field trial outcomes published. In their new Biostimulants Code of Practice, Fertilizer Australia (2025) suggest that product efficacy claims be supported by reference to relevant, peer reviewed, published literature. Such publications will help to improve future field trial design and execution, and it is encouraging to see recent examples of independent field validation of emerging microbial products (e.g., Woodward et al. 2025).

Failed field experiments probably occur quite frequently but are rarely reported in the literature or product dossiers (Bacilio et al. 2017; Bashan et al. 2020). Reluctance to publish negative results is understandable but the sharing of all results is important and there is value in reporting negative results if possible. One way to publish negative results is to incorporate these alongside successful trials for the same microbial agent so that lessons learned from 'failed' trials can be shared. Negative results may indicate conditions under which microbial products are less efficacious, which is important information for eventual users of the products. Failure to publish non-significant or negative information can lead to overly optimistic impressions of product efficacy when studies are combined in meta-analyses and waste of limited funding on "reinventing the wheel". Evidence of success is sometimes presented as meta-analysis of large numbers of trials, often over multiple years and locations (e.g. Leggett et al. 2015; Schütz et al. 2018). While these analyses are useful summations of research, closer inspection of the data sometimes indicates variable levels of efficacy, for example gains reviewed for biological control of Take-all disease in wheat were reported to be around 3% but these ranged from -15% to +33% (Gupta et al. 2011). Also, not all trial designs may be comparable. Even for long used and well-studied beneficial microorganisms such as arbuscular mycorrhizal fungi, scientific opinion is divided as to whether AMF inoculants deliver consistent benefits in cropping situations (Ryan & Graham 2018; Zhang et al. 2019). This lack of consensus further highlights the importance of evaluating products imported from overseas under New Zealand growing and environmental conditions.

CONCLUSIONS

Historically, the lack of independent field validation of microbial products has undermined confidence in their field efficacy (Owen et al. 2015). Greater effort is needed to ensure early and independent scientific evaluation of both potential and current commercial products under real-world conditions, using robust experimental design. Multi-environment, multi-seasonal field trials focussed on agronomically relevant parameters (e.g., crop health and yield) are essential to provide urgently needed information on the reliability, effect size and utility of microbial products used as biopesticides, biofertilisers and biostimulants.

Field trials are expensive and need to be set up for success. It is not possible to develop uniform criteria to define efficacy levels and evaluate performance for all microbial products because of their many and various modes of action, as well as the complexity of potential biological and environmental interactions between the microbial agent, the target pest/disease, and the target crop. Instead, trial designs appropriate to the microorganism(s) being tested and the proposed research hypotheses need to be developed on a case-by-case basis. Investment in all aspects of trial planning is crucial to successful delivery in the field and increases the likelihood of generating useful insights, even if unexpected events occur during the trial itself. Well replicated trials with appropriate spatial layout will be more robust and have some capacity for risk mitigation, when this is needed.

Collaboration between researchers, industry stakeholders, and end users is essential for successful field trials because trial design needs to be informed by the wider agronomic context. Communication of trial outcomes is an important part of this collaboration and needs to be as transparent as possible, subject to IP constraints, including reporting of negative results as well as positive outcomes. Independent peer review of trial results will increase the confidence of end users and encourage adoption of novel microbial products in future.

ACKNOWLEDGEMENTS

Thanks to David Wright and Nicole Schon for reviewing an earlier version of this manuscript. We gratefully acknowledge our co-authors of publications cited in this paper. Preparation of this paper was partly funded by the Early-stage Product Development Flagship (AgResearch SIFF). The authors were also informed by discussions with practitioners in Platform 10, an international grower-driven collaboration that aims to accelerate development of biopesticides through rigorous field evaluation and supporting adoption Platform 10 | The Global Specialty Crop Biological Platform

REFERENCES

Akter Z, Weinmann M, Neumann G, Römheld V 2013. An *in-vitro* screening method to study the activity potential of biofertilizers based on *Trichoderma* and *Bacillus* sp. Journal of Plant Nutrition 36: 1439-1452. https://doi.org/10.1080/01904167.2013.793713

Arxada. 2025. Lifegard biostimulant label. Retrieved 8 September 2025 from https://arxada.co.nz/wp-content/uploads/2023/10/lifegard-label-1-1.pdf

Bacilio M, Moreno M, Lopez-Aguilar D, Bashan Y 2017. Scaling from growth chamber to the greenhouse to the field: Demonstration of diminishing effects of mitigation of salinity in peppers inoculated with plant growth-promoting bacterium and humic acids. Applied Soil Ecology 119: 327-338. https://doi.org/10.1016/j.apsoil.2017.07.002

Bashan Y, Kloepper J, De-Bashan L, Nannipieri P 2016. A need for disclosure of the identity of micro-organisms, constituents, and application methods when reporting tests with microbe-based or pesticide-based products. Biology & Fertility of Soils 52*:* 283-284. https://doi.org/10.1007/s00374-016-1091-y

Bashan Y, Prabhu S, De-Bashan L, Kloepper J 2020. Disclosure of exact protocols of fermentation, identity of microorganisms with consortia, formation of advanced consortia with microbe-based products. Biology & Fertility of Soils 56: 443-445. https://doi.org/10.1007/s00374-020-01464-x

biostart 2025. Product information on BioShield. Retrieved 9 September 2025 from https://biostart.co.nz/bioshield-grass-grub-liquid/

Bourdôt GW, Hurrell GA, Saville DJ, Leathwick DM 2006. Impacts of applied *Sclerotinia sclerotiorum* on the dynamics of a *Cirsium arvense* population. Weed Research 46: 61-72. https://doi.org/10.1111/j.1365-3180.2006.00481.x

Chandler D, Davidson G, Grant W, Greaves J, Tatchell G 2008. Microbial biopesticides for integrated crop management: an assessment of environmental and regulatory sustainability. Trends in Food Science & Technology 19: 275-283. https://doi.org/10.1016/j.tifs.2007.12.009

Environmental Protection Authority 2025. Approvals and group standards for substances. Retrieved 31 July 2025 from https://www.epa.govt.nz/hazardous-substances/substance-approvals-and-group-standards/

EPPO Bulletin 2021. Conduct and reporting of efficacy evaluation trials, including good experimental practice. PP 1/181 (5). EPPO Bulletin 2021. (Paris, France: EPPO), 1-13.

Fertilizer Australia 2025. Retrieved 9 September 2025 from https://fertilizer.org.au/biostimulant

Frank M, Husted S 2024. Is India's largest fertilizer manufacturer misleading farmers and society using dubious plant and soil science? Plant & Soil 496: 257-267. https://doi.org/10.1007/s11104-023-06191-4

Franzen D, Camberato J, Nafziger E, Kaiser D, Nelson K, Singh G, Ruiz-Diaz D, Lentz E, Steinke K, Grove J, Ritchie E, Bortolon L, Rosen C, Maharjan B, Thompson L 2023. Performance of selected commercially available asymbiotic N-fixing products in the North Central Region. North Dakota State University. https://www.ndsu.edu/agriculture/sites/default/files/2023-04/sf2080.pdf

Galindo F, Rodrigues W, Fernandes G, Boleta E, Jalal A, Rosa P, Buzetti S, Lavres J, Teixeira Filho M 2022. Enhancing agronomic efficiency and maize grain yield with *Azospirillum brasilense* inoculation under Brazilian savannah conditions. European Journal of Agronomy 134: 126471. https://doi.org/10.1016/j.eja.2022.126471

Gupta V, Rovira AD, Roget DK 2011. Principles and management of soil biological factors for sustainable rainfed farming systems. In: Tow P, Cooper I, Partridge I, Birch C Eds. Rainfed Farming Systems. Springer. pp. 149-184. https://doi.org/10.1007/978-1-4020-9132-2_6

Hett J, Döring TF, Bevivino A, Neuhoff D 2023. Impact of microbial consortia on organic maize in a temperate climate, varies with environment but not with fertilization. European Journal of Agronomy 144: 126743. https://doi.org/10.1016/j.eja.2023.126743

Holden P (Ed) 2025. New Zealand Novachem Agrichemical Manual 2024/2025. AgriMedia Ltd, Christchurch, 824 pp.

Hoyte S, Elmer P, Parry F, Phipps J, Spiers M, Taylor J. et al. 2018. Development of a new biocontrol product (AUREO® Gold) for control of *Pseudomonas syringae* pv. *actinidiae* in kiwifruit. Biological and Integrated Control of Plant Pathogens IOBC-WPRS Bulletin 133: 164-166.

Hurst MRH, Swaminathan J, Wright DA, Hardwick S, Ferguson CM, Beattie A, Richards NK, Harper L, Moss RA, Cave VM, van Koten C, McNeill MR 2019. Development of a *Yersinia entomophaga* bait for control of larvae of the porina moth (*Wiseana* spp.), a pest of New Zealand improved grassland systems. Pest Management Science 76: 350-359. https://doi.org/10.1002/ps.5521

Jackson TA, Townsend RJ, Dunbar JE, Ferguson CM, Marshall SDG, Zydenbos SM 2012. Anticipating the unexpected - managing pasture pest outbreaks after large-scale land conversion. Proceedings of the New Zealand Grassland Association 74: 153-158. https://doi.org/10.33584/jnzg.2012.74.2861

Johnstone PD, Sinclair AG 1991. Replication requirements in field experiments for comparing phosphatic fertilizers. Fertilizer Research 29: 329-333. https://doi.org/10.1007/BF01052402

Leggett M, Newlands N, Greenshields D, West L, Inmann S, Koivunen M 2015. Maize yield response to a phosphorous-solubilizing microbial inoculant in field trials. Journal of Agricultural Science 153: 1464-1478. https://doi.org/10.1017/S0021859614001166

Lopes M, Dias-Filho M, Gurgel E 2021. Successful plant growth-promoting microbes: inoculation methods and abiotic factors. Frontiers in Sustainable Food Systems 5: 606454. https://doi.org/10.3389/fsufs.2021.606454

Mansfield S, Chynoweth R, Hurst M, Noble A, Zydenbos S, O'Callaghan M 2017. Novel bacterial seed treatment protects wheat seedlings from insect damage. Crop & Pasture Science 68: 527-533. https://doi.org/10.1071/CP17176

Mansfield S, Gerard PJ, Hurst MRH, Townsend RJ, Wilson DJ, Van Koten C 2016. Dispersal of the invasive pasture pest *Heteronychus arator* into areas of low population density: effects of sex and season, and implications for pest management. Frontiers in Plant Science 7: 1278. https://doi.org/10.3389/fpls.2016.01278

Mansfield S, Wilson M, Gerard P, Wilson D, Swaminathan J, Wright D, van Koten C, Hurst M 2020. Potential for a biopesticide bait to control black beetle, *Heteronychus arator* (Coleoptera: Scarabaeidae). Pest Management Science 76: 4150-4158. https://doi.org/10.1002/ps.5973

Marrone PG 2023. Biopesticide commercialization in North America: state of the art and future opportunities. In: Koul O Ed. Development and Commercialization of Biopesticides. Academic Press, pp. 173-202. https://doi.org/10.1016/B978-0-323-95290-3.00017-0

Merfield C, Johnson M 2016. Understanding biostimulants, biofertilisers and on farm trials. Lincoln, New Zealand: The BHU Future Farming Centre 13. https://www.bhu.org.nz/wp-content/uploads/sites/155/ffc-files/soil-management/understanding-biostimulants-biofertilisers-and-on-farm-trials-2016-ffc-merfield-johnson.pdf

Ministry for Primary Industries 2025. Authorisation of agricultural compounds and veterinary medicines (ACVM) Retrieved 31 July 2025 from https://www.mpi.govt.nz/agriculture/agricultural-compounds-vet-medicines/authorisation-of-acvm/

Naranjo SE, Diehl JW, Ellsworth PC 1997. Sampling whiteflies in cotton: validation and analysis of enumerative and binomial plans. Environmental Entomology 26:777-788. https://doi.org/10.1093/ee/26.4.777

Narciso J, Ormskirk M, Jones S, Rolston P, Moran-Diez ME, Hurst M, Brookes J, Glare T 2021. Using multiple insecticidal microbial agents against diamondback moth larvae - does it increase toxicity? New Zealand Journal of Agricultural Research 64: 178-193. https://doi.org/10.1080/00288233.2019.1582074

Neuhoff D, Neumann G, Weinmann M 2024. Testing plant growth promoting microorganisms in the field - a proposal for standards. Frontiers in Plant Science 14:1324665. https://doi.org/10.3389/fpls.2023.1324665

O'Callaghan M, Ballard R, Wright D 2022. Soil microbial inoculants for sustainable agriculture: Limitations and opportunities. Soil Use & Management 38: 1340-1369. https://doi.org/10.1111/sum.12811

Owen D, Williams A, Griffith G, Withers P 2015. Use of commercial bio-inoculants to increase agricultural production through improved phosphorous acquisition. Applied Soil Ecology 86: 41-54. https://doi.org/10.1016/j.apsoil.2014.09.012

Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, DiLeo MV 2016. From the lab to the farm: An industrial perspective of plant beneficial microorganisms. Frontiers in Plant Science 7: 1110. https://doi.org/10.3389/fpls.2016.01110

Peiris P, Xu, C, Brown P, Li Y 2021. Assessing the efficacy of alternative chemical and organic products against *Meloidogyne* spp. in sweetpotato. Scientia Horticulturae 283: 110079. https://doi.org/10.1016/j.scienta.2021.110079

Precedence Research 2025. Biostimulants and biopesticides market size to hit USD 29.32 Bn by 2034. Retrieved 9 September 2025 from https://www.precedenceresearch.com/biostimulants-and-biopesticides-market

Quinn GP, Keough MJ 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK. 537 p. https://doi.org/10.1017/CBO9780511806384

Ricci M, Tilbury L, Daridon B, Sukalac K 2019. General principles to justify plant biostimulant claims. Frontiers in Plant Science 10: 494. https://doi.org/10.3389/fpls.2019.00494

Ryan MH, Graham J 2018. Little evidence that farmers should consider abundance or diversity of arbuscular mycorrhizal fungi when managing crops. New Phytologist 220: 1092-1107. https://doi.org/10.1111/nph.15308

Schmidt J, Gaudin C 2018. What is the agronomic potential of biofertilizers for maize? A meta-analysis. FEMS Microbiology Ecology 94: fiy094. https://doi.org/10.1093/femsec/fiy094

Schütz L, Gattinger A, Meier M, Müller A, Boller T, Mäder P, Mathimaran N 2018. Improving crop yield and nutrient use efficiency via biofertilization-a global meta-analysis. Frontiers in Plant Science 8: 2204. https://doi.org/10.3389/fpls.2017.02204

Shi S, Wakelin S, Gerard E, Young S, van Koten C, Caradus J, Griffiths A, Ballard R, O'Callaghan M 2023. Screening and field evaluation of white clover rhizobia for New Zealand pastures. Crop & Pasture Science 74: 1258-1271. https://doi.org/10.1071/CP22405

Smyth E, McCarthy J, Nevin R, Khan M, Dow J, O'Gara F, Doohan F 2011. *In vitro* analyses are not reliable predictors of the plant growth promotion capability of bacteria; a *Pseudomonas fluorescens* strain that promotes the growth and yield of wheat. Journal of Applied Microbiology 111: 683-692. https://doi.org/10.1111/j.1365-2672.2011.05079.x

Southwood TRE, Henderson PA 2000. Ecological methods. Blackwell Science, Oxford, UK. 575 p.

Spinner JE, Mansfield S, Pilkington LJ, Thomson P 2011. Sampling protocol to detect *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in mixed species populations in greenhouse vegetable crops. Australian Journal of Entomology 50: 276-280. https://doi.org/10.1111/j.1440-6055.2011.00819.x

Sustainable Agriculture Research & Education (SARE) 2017. SARE Technical Bulletin. How to conduct research on your farm or ranch. https://www.sare.org/wp-content/uploads/how-to-conduct-research-on-your-farm-or-ranch.pdf

Townsend RJ, Ferguson CM, Proffitt JR, Slay MWA, Swaminathan J, Day S, Gerard E, O'Callaghan M, Johnson VW, Jackson TA 2004. Establishment of *Serratia entomophila* after application of a new formulation for grass grub control. New Zealand Plant Protection 57: 310-313. https://doi.org/10.30843/nzpp.2004.57.6927

UPL 2025. AureoGold product information. Retrieved 9 September 2025 from https://www.uplcorp.com/nz/product-details/aureo-gold

Villamizar LF, Barrera GP, Luange A, Sagata K, Gende P, Chris S, Tsatsia H, Mudu F, Weston M, Van Koten C, Mansfield S, Jackson TA, Marshall SDG 2024. Characterization and screening of new *Metarhizium* isolates to control the coconut rhinoceros beetle in the Pacific islands. Fungal Biology 128: 2127-2138. https://doi.org/10.1016/j.funbio.2024.08.009

Woodward L, Sible C, Seebauer J, Below F 2025. Soil inoculation with nitrogen-fixing bacteria to supplement maize fertilizer need. Agronomy Journal 117: e21729. https://doi.org/10.1002/agj2.21729

Wright D, Zydenbos S, O'Callaghan M, Jackson T, Townsend R, van Koten C, Wessman P, Mansfield S 2017. Surface coating aids survival of *Serratia entomophila* (Enterobacteriaceae) in granules for surface application. Biocontrol Science & Technology 27: 1383-1399. https://doi.org/10.1080/09583157.2017.1402861

Zhang L, Wang D, Hu Q, Dai X, Xie Y, Li Q, Liu H, Guo J 2019. Consortium of plant growth-promoting rhizobacteria strains suppresses sweet pepper disease by altering the rhizosphere microbiota. Frontiers in Microbiology 10: 1-10. https://doi.org/10.3389/fmicb.2019.01668

Zydenbos SM, Townsend RJ, Lane PMS, Mansfield S, O'Callaghan M, van Koten C, Jackson T 2016. Effect of *Serratia entomophila* and diazinon applied with seed against grass grub populations on the North Island volcanic plateau. New Zealand Plant Protection 69: 86-93. https://doi.org/10.30843/nzpp.2016.69.5919