ALLIANCE FOR RESPONSIBLE AQUACULTURE OUTCOME EVALUATION

July 2026

www.fishwelfareinitiative.org

Jenifer-Justine Kirsch
Director of Programs

Field Execution: Chaitanya Akula, Ravi Teja Asleti, Nagaraju Pulicherla

Alliance For Responsible Aquaculture Outcome Evaluation

Study lead: Jennifer Kirsch
Study execution: Chaitanya Akula, Ravi Teja Asleti, Nagaraju Pulicherla
Analysis assistance: Dr. Adam Holt & Norma Rocio Forero Muñoz

Executive Summary

- We ran a field-based randomized controlled trial (RCT) to evaluate if our main program, the Alliance for Responsible Aquaculture (ARA), is achieving its expected outcomes.
- The study comprised 53 farms: 25 treatment and 28 control.
- Data collection took place between February and May 2026 and the analysis was completed in June 2026 by two analysts, both of whom were blind to the treatment and control assignment. We also completed an analysis with Claude, prompted and supervised by the study lead (who was *not* blind to control/treatment assignment).
- All three analyses confirmed our hypothesis that farms where farmers received water quality readings and corrective actions experienced more frequent water quality improvements (82% on Day 3) than farms that did not receive this assistance (17% on Day 3).
- These results suggest that the ARA achieves its main outcome, overall strengthening the program's evidence-base.
- The main uncertainty about the results stems from a concern that farmers' self-initiated (and not FWI-prescribed) actions may drive some of the improvements.
- Limitations of this study include seasonal variations, previous treatment exposure across all farms, and limited geographic representation of the results.

Contents

1. Background

The current core program of Fish Welfare Initiative (FWI) is the Alliance for Responsible Aquaculture (ARA). By joining the ARA, farmers commit to reduced stocking densities and to maintaining water quality in required ranges. ARA field teams collect water quality data, including DO, pH, and ammonia, from member farms approximately once a month. If measurements reveal that key water quality parameters are out-of-range and fish may be exposed to unhealthy conditions, staff provide the farmers with recommendations for corrective actions. Corrective actions are practical steps such as turning on aerators, adjusting feeding, or adding farm treatments to improve water quality. Fig. 1 shows the theory of change for this program.

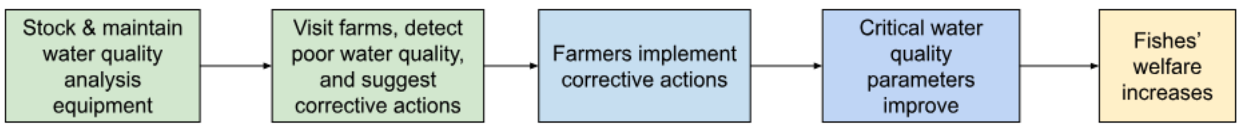

Figure 1. The Alliance for Responsible Aquaculture's theory of change.

The ARA has been operating since July 2021 and as of July 2026, 168 farms actively participate in the program. Our monitoring shows that the first parts of our theory of change work as intended.[1] Farms are visited at scheduled times, we are able to take follow-ups 95% of the time (sometimes rain and other external factors prevent field visits), farmers implement the recommended corrective actions 90% of the time, and targeted water quality parameters improve 87% of the time.[2] We previously had less certainty about whether water quality improves as a result of our corrective action. It's possible that in the absence of FWI-recommended corrective actions, external factors cause water quality to improve. Self-initiated actions taken by farmers could cause improvements that would have happened even without our intervention. Similarly, water quality is inherently dynamic and changes over time—with weather being an important factor—so it's possible that improvements in water quality that we currently assign to an effect of our intervention happen naturally. To investigate this assumption, we ran an evaluation of the ARA's primary outcome between February and May 2026. This report describes the study protocol, results, and what these results mean for our work.

2. Study Overview

We ran a field-based randomized controlled trial to evaluate whether the ARA achieves its intended outcome of improving water quality on fish farms. Our hypothesis was:

[1] See our latest M&E Report.
[2] To explore these numbers in real-time, visit our public ARA data dashboard.

Farms that experience out-of-range water quality instances and are provided with water quality readings and corrective actions will have more frequent water quality improvements (back into required ranges) than farms that do not receive water quality readings or corrective actions.

In other words, we explored whether providing farmers with water quality readings and corrective actions results in improved water quality. The protocol closely followed regular ARA operations to evaluate the effectiveness of the current program.

2.1 Overview Of The Study Design

The study was completed between February 4 and May 27, 2026 in one of our programming regions, Eluru, India. FWI staff visited treatment and control farms every two weeks to collect DO, pH, and ammonia measurements. These parameters were then compared against our required water quality ranges (Table 1). For treatment farms, if one or more parameter(s) fell outside the required range, FWI staff informed the farmer about their water quality values and requested the implementation of corrective actions (e.g., turning on aerators, stopping feeding for one day) to bring the poor parameter(s) back into the required ranges (Fig. 3). Control farmers did not receive information about their water quality or corrective actions. FWI staff then visited control and treatment farms two and three days after the initial out-of-range event to take another water quality sample. The primary analysis focused on Day 3 follow-ups because this is when most follow-ups in the ARA program are usually taken and it is thus most representative of the program. Day 2 follow-ups were analyzed as an additional metric to determine whether improvements are already visible then. The study's target for statistical significance was 16 out-of-range events for treatment and control, respectively.

Table 1. The required water quality ranges farmers commit to when joining the ARA.

Water Quality Parameter	Required Range
Dissolved oxygen (Morning) Dissolved oxygen (Evening)	3 - 5 mg/L 8 - 12 mg/L[3]
pH	6.5 - 8.5
Ammonia	≤ 0.05 mg/L

[3] Note that we have since adjusted our required ranges for DO. We removed the upper bound as a recent literature review suggested there is no evidence that higher DO compromises fish welfare. We ran an analysis to test whether these range changes affect the study results and interpretation and came to the conclusion that they do not.

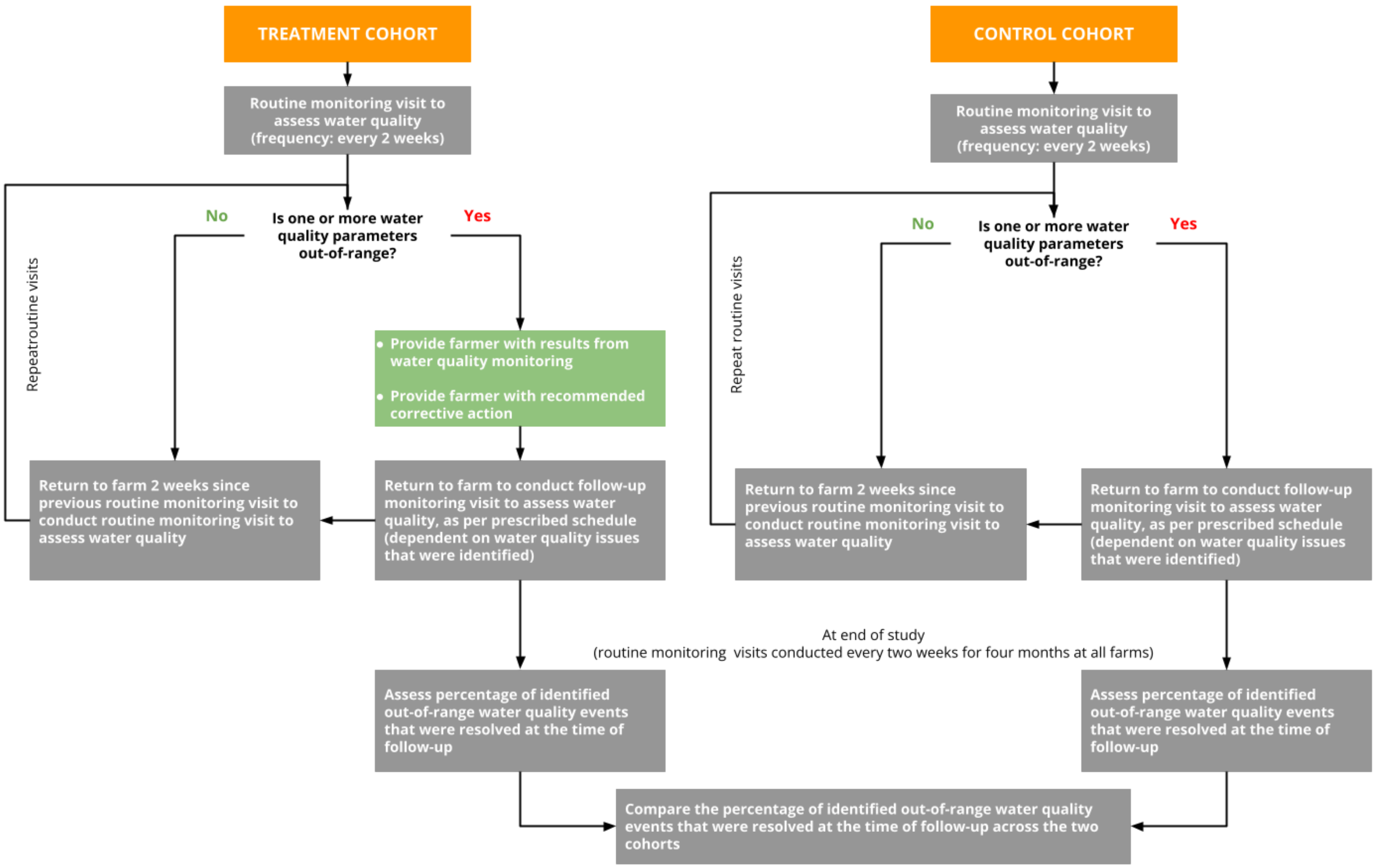

Figure 3. Overview of the study design. The key difference between the two cohorts was that control farms were not informed of the findings from routine monitoring visits, and no corrective actions were recommended to farmers. In the event that a routine monitoring visit at a control farm identified one or more water quality parameters being out-of-range, field teams revisited that farm to assess if water quality has resolved on its own and in the absence of corrective actions.

2.2 Farm Recruitment and Group Assignment

The study included 53 farms, randomly assigned to treatment (25 farms) and control (28 farms). Notably all participating farms were enrolled in the ARA when we added them to this study. This implies some previous exposure to the ARA intervention, including water quality readings and corrective actions. We considered adding a true control (i.e., non-ARA farmers) but were not able to motivate farmers to allow us access to their farm without sharing water quality measurements. The trust that we had built with ARA farmers over the years of their enrollment in our program allowed us to treat them as control farms without them interfering with the study design. While this was the only feasible way to conduct this study, we acknowledge that our control group is not a true control.

After the treatment and control group assignment, we checked the groups for balance on farm size, stocking density, fish weight and historical frequency of water quality issues, and we found them to be well balanced. Fig. 2 displays a map of the study area and participating farms.

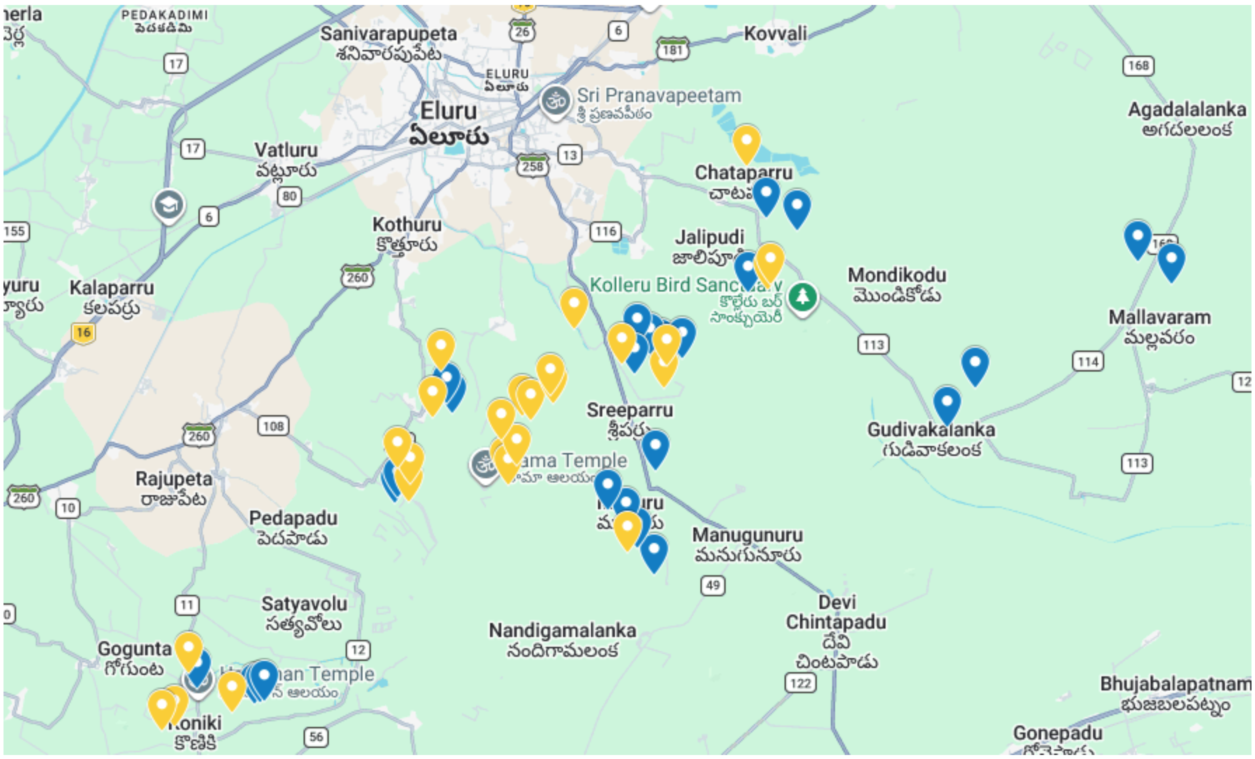

Figure 2. Overview of the participating treatment (blue) and control farms (yellow).

2.3 Data Quality Control and Assurance

The study protocol included processes to ensure high quality of data and reduce bias in the results. We checked for staff compliance with corrective action SOPs, farmer compliance with corrective actions and staff compliance with control protocols (to catch accidental treatment of control farms).

2.4 Data Analysis

The analysis was completed by three different entities to increase scientific rigor. Farm identifiers and control/treatment assignment were blinded for the two external analysts' primary analysis to reduce confirmation bias. Later analysis, including data on corrective actions, required unblinding the dataset to reveal the relevant columns. The analyses were completed by

- Adam Holt, Assistant Professor of Geophysics at the University of Miami
- Norma Rocio Forero Muñoz, PhD candidate in computational ecology at Université de Montréal
- Claude, prompted and supervised by the FWI study lead and Director of Programs, Jennifer Kirsch

For more details on the study methodology and analysis, see the full study protocol. Also note that we had run this study in the second half of 2025 but decided to repeat it due to external factors and protocol issues that resulted in data unfit for analysis.[4]

3. Results

3.1 Data Quality Control

Maintaining data quality was critical to ensuring the study results are trustworthy, particularly because the study required field staff to visit control farms without sharing water quality readings or corrective actions. This could have been difficult to maintain consistently in practice. We monitored quality across three areas.

3.1.1 Staff Compliance With Control Farm Protocols

The most important quality control concern was ensuring that control farmers were never told their water quality readings or given corrective action recommendations. To verify this, a senior team member shadowed field staff during eight control farm visits: two regular monitoring visits and two follow-up visits per field associate. All eight visits passed. This gives us confidence that the separation between treatment and control was maintained throughout the study.

3.1.2 Staff Compliance With Corrective Action Protocols

We also checked whether field staff were recommending the right corrective actions to treatment farmers according to our standard procedures. This was assessed through a survey where we asked staff what exactly they tell farmers when giving a corrective action. The initial surveys suggested low compliance, with 53% in January and 33% in March. We found these results suspicious given that staff had been trained on and used the corrective action SOPs before the outcome evaluation. Upon further investigation, we found that a translation and communication error in the survey led to these low compliance scores. Associates had misunderstood the questions and thus responded with SOPs on *when* to suggest corrective actions rather than *what* to suggest. We rectified this error and reran the survey in May. Staff achieved 100% compliance in the final May survey.

[4] For more details, see this blog post.

3.1.3 Farmer Compliance With Corrective Actions

We verified whether treatment farmers actually implemented the corrective actions recommended to them through two mechanisms: self-reported information collected at every follow-up visit, and direct shadowing visits where a team member observed the farmer implementing the action. Self-reported information was only assessed for events where farmers implemented corrective actions because in the four cases where they didn't, we did not receive data about the actions that we could assess. We reviewed the self-reported corrective action information twice during the study period and compliance with our SOPs was at 100% both times. The first review noted incomplete data entries for some visits but we don't expect these incomplete data entries to significantly change the overall compliance percentage. We also completed five shadowing visits, covering actions including stopping feeding, applying zeolite, and adding jaggery. All five passed. Finally, the results for farmer compliance, while not fully complete, pointed toward data integrity.

Apart from the communication challenge in the first staff surveys, quality control was completed at target levels across all three areas, and no visits were flagged for data exclusion due to protocol violations.

To view the raw data for each of these, please contact Jennifer.

3.2 Dataset Overview and Baseline Analysis

The three analyses looked at the same data but ended up applying different methodologies: Norma used detailed statistical models that investigated confounding variables, Adam focused on an expanded binary comparison of groups, and the Claude analysis, having the detailed study background, went one layer deeper and applied various tests and additional cross-comparisons. The following section describes the aggregate results, pulling the relevant analysis from each of the three completed. All individual analyses are accessible for further detail:

- Adam's analysis
- Norma's analysis
- Claude analysis

The study dataset[5] included 998 monitoring visits (initial and follow-ups) split between the two cohorts. We visited every farm in the morning and evening and most data analysis was completed in the unit of farm-day, combining morning and evening measurement results into one daily unit. Table 2 shows the breakdown. We exceeded our statistical target of 16 events for the control and treatment group, respectively, by almost double.

[5] The full dataset can be found here.

Table 2. Overview of the data collected for this study. OOR means "out-of-range". All monitoring visits include a morning and evening visit per regular ARA protocols. One control Day 2 follow-up was not collected due to staff getting sick.

	Control	Treatment	Total
Monitoring visits	266	233	499
routine (non-follow-up)	207	177	384
follow-up (Day 2)	29	28	57
follow-up (Day 3)	30	28	58
Farms (baseline)	28	25	53
OOR events	30	28	58
Farms with OOR events	15	17	32

Participating farms in this study were randomly assigned to treatment and control groups and then checked for balance on potential confounders. The analysis revealed that the groups were well balanced at baseline: The two groups had similar water quality parameters, suggesting that any effect observed from the treatment was likely not due to inherent differences between the groups (Fig. 4). This result was further supported by Levene's and Hedges' tests.

Levene's Test

Levene's test checks whether two groups have a similar spread of values, not just similar averages where $p > 0.05$ means no evidence the spreads differ. It asks: if the two groups truly had equal spread, how often would a spread difference this big appear just by chance? For this data, p clears this threshold for every water quality parameter with the smallest p-value obtained for evening DO (0.088).

Hedges' g

Hedges' g is a measure of how far apart two group averages are relative to natural variation between farms: the largest value, ammonia's 0.24, suggests that those group means differ by about a quarter of a standard deviation; the other parameters are near zero. Notably, Hedges' g is an effect size, not a test. As a result, there is no p-value, just a rough scale on which 0.2 counts as small.

Fig. 4 includes the obtained Levene's and Hedges' g scores. To summarize, none of the parameters shows a meaningful difference. The groups start out comparable, and the differences in later sections are not a baseline artifact.

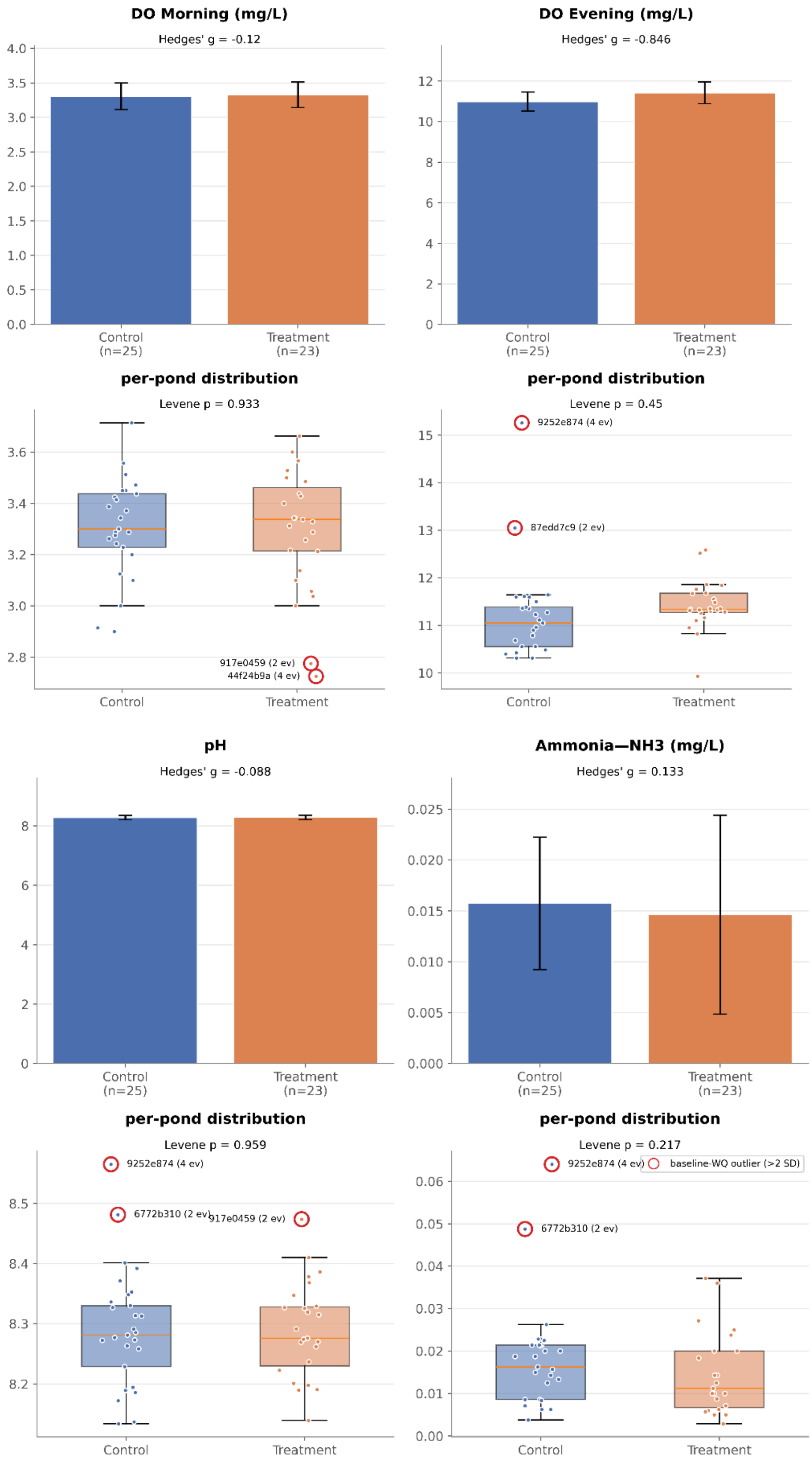
DO Morning (mg/L)
Hedges' g = -0.12
Control (n=25)
Treatment (n=23)
per-pond distribution
Levene p = 0.933
917e0459 (2 ev)
44f24b9a (4 ev)
Control
Treatment
DO Evening (mg/L)
Hedges' g = -0.846
per-pond distribution
Levene p = 0.45
9252e874 (4 ev)
87edd7c9 (2 ev)
pH
Hedges' g = -0.088
per-pond distribution
Levene p = 0.959
9252e874 (4 ev)
6772b310 (2 ev)
917e0459 (2 ev)
Ammonia—NH3 (mg/L)
Hedges' g = 0.133
per-pond distribution
Levene p = 0.217
9252e874 (4 ev)
baseline-WQ outlier (>2 SD)
6772b310 (2 ev)

Figure 4. The whisker plots show the per-farm baseline water quality with each point representing an individual farm and DO split into morning and evening. The top panels for each parameter show the group mean ± standard deviation. The bottom panels show the box (median, inter-quartile range). Outlier farms (marked in red) are excluded from the plots. See section 3.3.1 for the outlier definition.

3.3 Water Quality Problems Breakdown

The study defined a water quality issue as an event where one or more of the three key parameters (DO, pH, ammonia) are out-of-range. As such, the breakdown of individual parameter combinations did not matter for the analysis of the data. At the same time, it is useful to know the distribution of out-of-range events across parameters and groups to determine whether there is an inherent bias in the data.

Table 3 shows the frequency of each combination of water quality issues present at the time of the initial out-of-range event, by cohort. DO-only events were by far most common, followed by DO and pH. We found no statistical evidence that this distribution differs between treatment and control. Because several categories in Table 3 have very small counts (e.g., only one "Ammonia only" event overall), a standard chi-squared test would not be reliable, so we instead used a Monte Carlo permutation-based test: randomly reshuffling the cohort labels across the 58 events 100,000 times and recalculating a chi-squared statistic each time to build an empirical null distribution. A gap this large or larger between treatment and control arose by chance alone in 29% of the shuffles (χ^2 = 7.23, $p \approx 0.29$), indicating no evidence that the mix of out-of-range issue types differs meaningfully between the two cohorts.[6]

This supports the qualitative observation from Table 3: while the two cohorts weren't identical (for instance, only the treatment group had events with ammonia-only or pH-and-ammonia issues), these differences are well within what random chance would produce given the small counts involved, and do not represent a systematic imbalance introduced by randomization.

Table 3: Breakdown of the parameters that were out-of-range during each of the out-of-range events of the study.

WQ Issue Type	Control (n)	Treatment (n)	Total (n)	Control (%)	Treatment (%)
DO only	14	11	25	47%	39%
pH only	2	4	6	7%	14%
Ammonia only	0	1	1	0%	4%

[6] Note that this Monte Carlo permutation-based test has been run in response to feedback on this report and is thus not present in any of the three analysis reports linked in Section 3.

DO + pH	6	9	15	20%	32%
DO + Ammonia	3	1	4	10%	4%
pH + Ammonia	0	1	1	0%	4%
DO + pH + Ammonia	5	1	6	17%	4%
Total	30	28	58	100%	100%

3.3.1 Outlier Definition

Before analyzing outcomes, we checked whether any farm started the study with unusually high or low water quality compared to the others. Using only initial measurements (excluding follow-ups), we compared the farm's baseline value for each parameter (DO, pH, ammonia) against its own group's average, and flagged a farm if any one parameter sat more than 2 standard deviations from that average, indicating an unusually high or low starting value. Five farms met this on at least one parameter. Because the outlier definition is defined only from starting values, it is independent of the study's outcome. Where useful, results are reported both with all farms and with these five removed. Most sections use all farms, since excluding them does not meaningfully change the results. Separately, we checked whether farms with many out-of-range events drove the outcome and the farm-level analysis in 3.4.1 shows they did not.

3.4 Primary Outcome: Resolution at Day 3

Per the protocol, the data included two follow-ups for each water quality out-of-range event: one on Day 2 and one on Day 3 after the initial measurement. Out-of-range events were considered resolved when all parameters (DO, pH, ammonia) were in range during the follow-up, regardless of which parameter was out-of-range during the initial measurement. For example, if the initial measurement saw pH out-of-range and pH resolved by the follow-up but then DO was out-of-range, the event would not be considered "resolved" and would count toward unresolved events. This binary assessment of all parameters (resolved/unresolved) showed that 82% of treatment events resolved water quality issues by Day 3 whereas only 17% of control events resolved. This trend persisted even when removing outlier farms, as defined in section 3.3.1. Figures 5a and 5b display the resolution rate with and without outliers and across different parameters.

The analysis included all farms and events according to the "intention-to-treat" definition and regardless of whether a farmer implemented corrective actions or not. Intention-to-treat means that we counted all farms that received the intervention, whether or not the farmer acted on the advice. In the treatment group, there were four events where farmers didn't implement corrective actions. If we remove these, the treatment resolution rate jumps from 82% to 96%. The following sections use the definition of "intention-to-treat" and thus include treatment events where farmers didn't implement corrective actions because the implementation is an assumption in our theory of change and not something we actively influence.

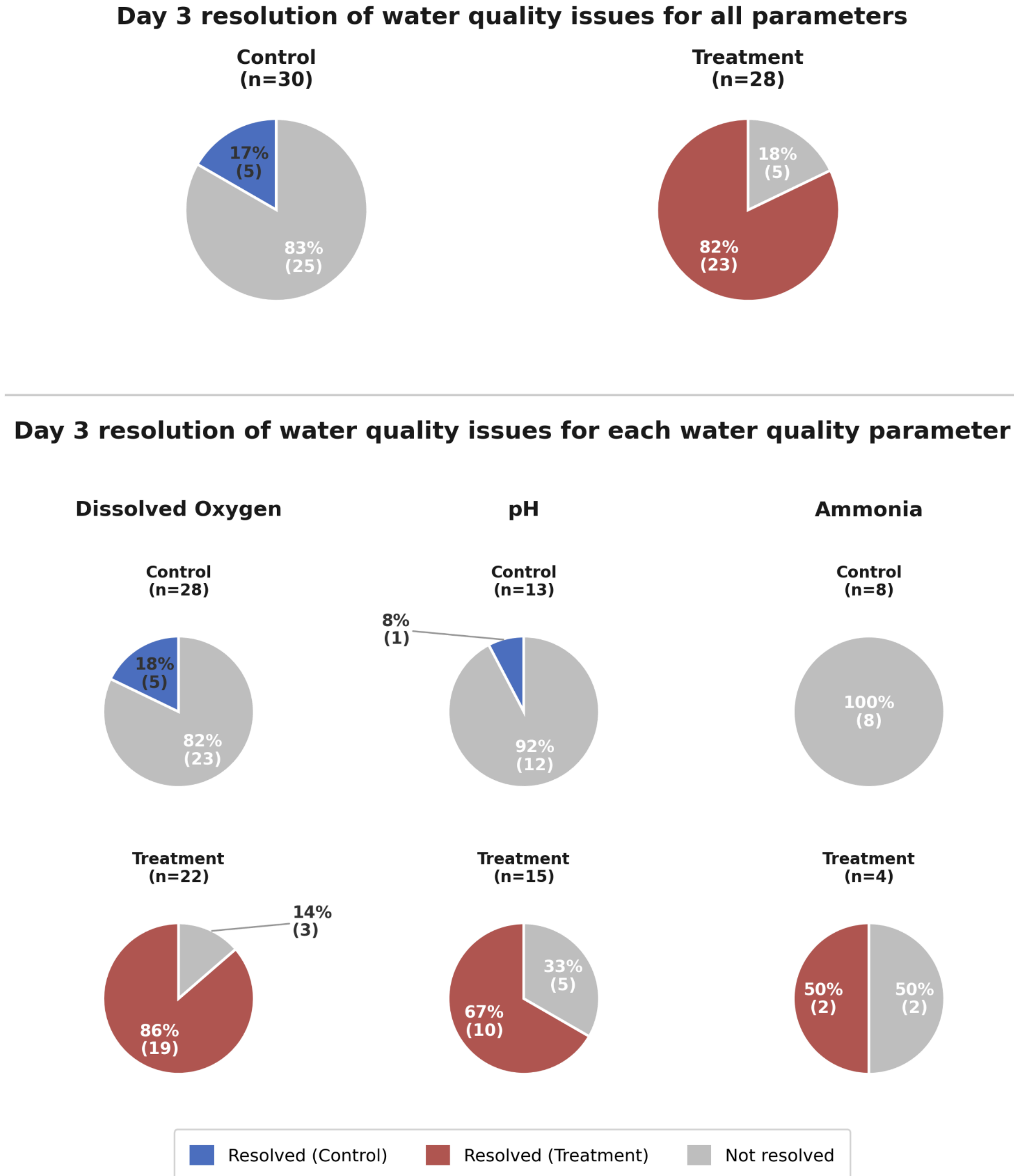

Figure 5a. Overview of Day-3 resolution across different breakdowns: The top two charts are the overall per group and the lower charts show the breakdown for each parameter individually.

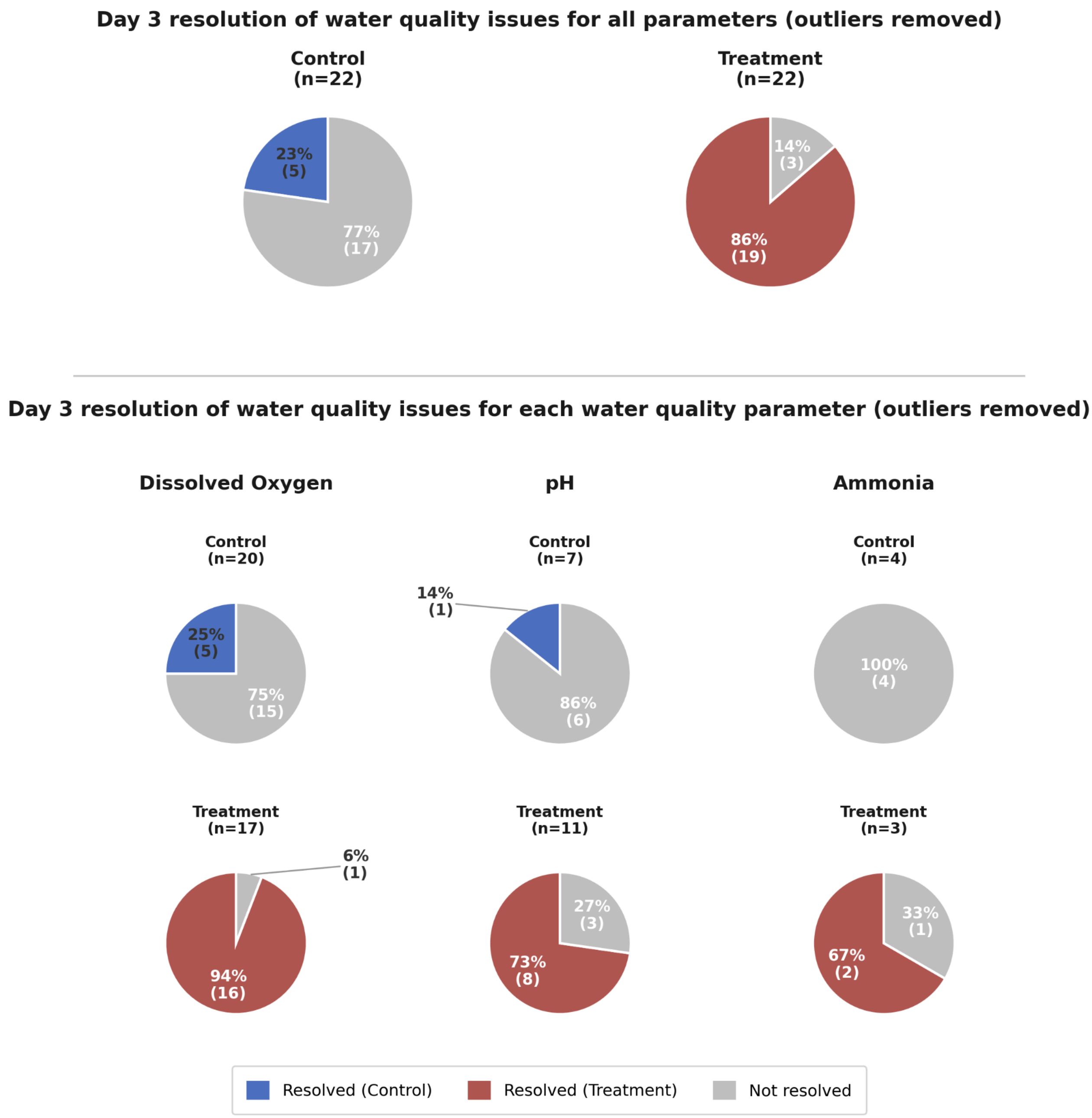

Figure 5b. Overview of Day-3 resolution across different breakdowns: The top two charts are the overall per group and the lower charts show the breakdown for each parameter individually. Outliers, as defined in section 3.3.1, were removed from this dataset.

3.4.1 Farm-Level Resolution Rate

The pie charts above count every out-of-range event once, meaning farms that experienced more events contribute more to the result. To check this isn't distorting the picture, Figure 6 takes a farm-level view instead, averaging each farm's own resolution rate so that every farm counts equally regardless of how many events it had. Most out-of-range events at control farms were not resolved, whereas most out-of-range events at treatment farms were resolved. The treatment-control gap holds whether we count events or farms, and removing the five outlier farms actually widens it slightly. Notably, resolution rate shows no relationship to how many events a farm had: a farm with four events was no more or less likely to resolve than one with a single event, which suggests the result is not being driven by a small number of particularly active or problematic farms.

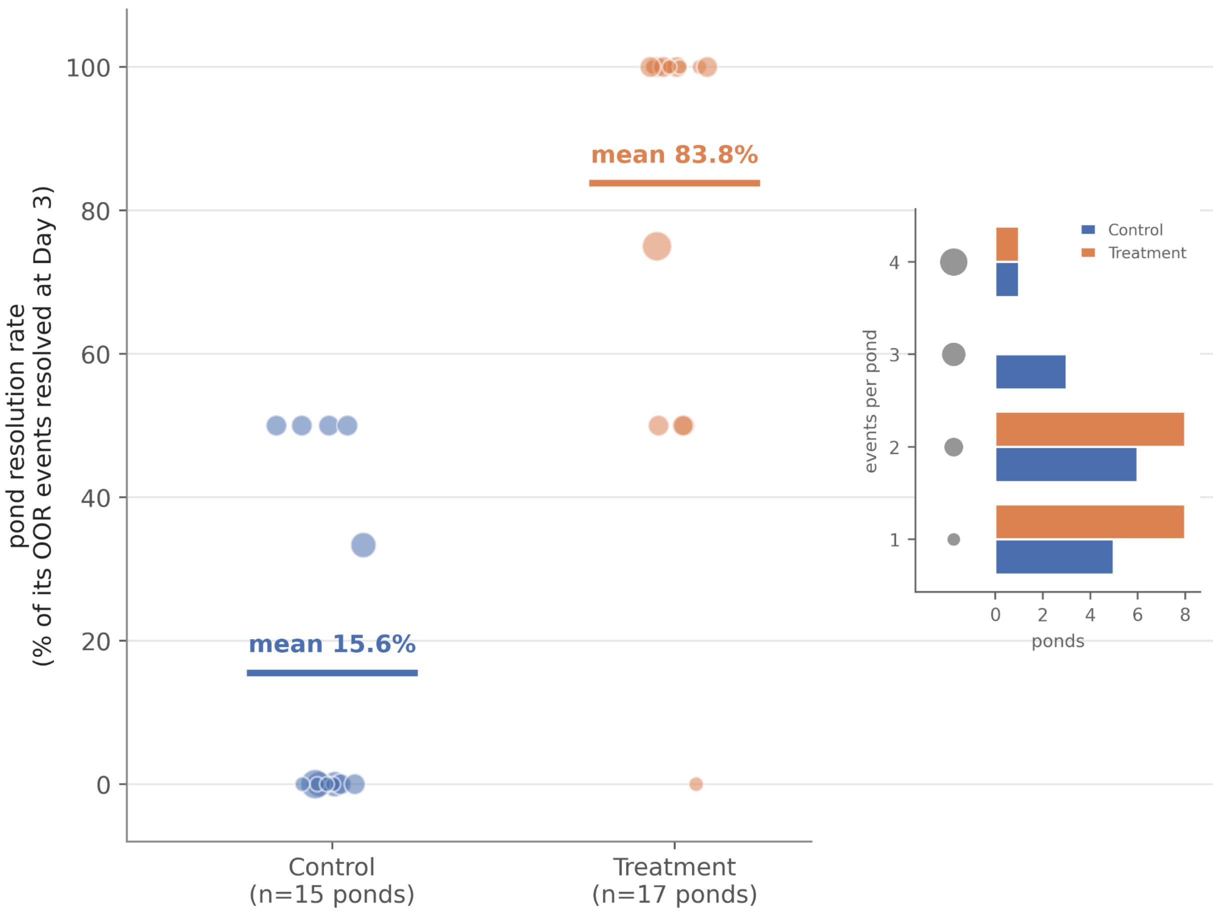

Figure 6. Each dot is one farm. Treatment farms almost universally resolved their water quality issues and control farms almost universally did not, regardless of how many issues they had.

3.4.2 Follow-Up Timing Investigation

The above analyses are all for the follow-up on Day 3, our primary unit for analysis. We also collected follow-up measurements on Day 2 to compare the two with each other. The Day 2 and 3 follow-up data suggest that the improvement only happened on Day 3. There was barely any improvement on Day 2; both groups saw water quality resolve at 21-25%. The treatment group then increased from 25% at Day 2 to 82% on Day 3 while the control group dropped to 17% on Day 3.

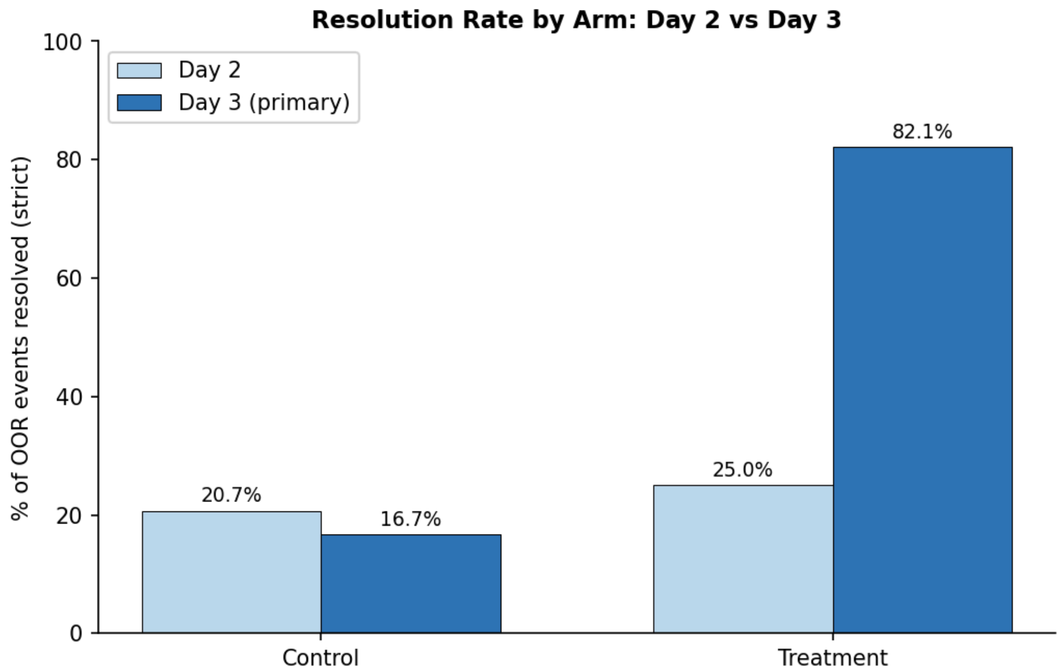

Figure 7. At Day 2, the difference between treatment and control is small and not statistically significant. The treatment effect only emerges at Day 3.

These results prompted us to further investigate the follow-up timing to ensure there is no timing difference that would cause the difference in resolution rate. This didn't seem to be the case. Day 3 visit times were essentially identical between the arms, a difference of about three minutes in both the morning and evening sessions (morning $p = 0.59$, evening $p = 0.65$) and the same balance held at Day 0. Visit time was also unrelated to whether an event was resolved, and all visits fell within the tight windows specified by the protocol.

The sudden spike in the resolution rate between Day 2 and 3 is an oddity we cannot fully explain and may hint at a study bias that we could not identify. One possible explanation may be that most corrective actions require time to take effect. For example, when a farmer stops feeding for one day to restore the pond's natural balance of algae, a change in plankton counts and subsequent improvement in DO is not immediate. Instead, the effect may require a few days to show. The study data were not able to confirm this hypothesis or show what happened after Day 3. We noted these considerations down for a future study focused on follow-up timing to investigate this further.

3.5 Statistical Tests

The three analyses included various statistical tests to evaluate the effectiveness of the intervention. The following sections describe the different tests and their results.

Table 4. Overview of the statistical tests run in the three analyses and their respective p-values.

Method	Guards against	p-value
Chi-squared test	Standard baseline	< 0.000001
Fisher's exact test	Small cell counts	< 0.000001
Farm-level permutation (20,000)	Clustering; no distributional assumptions	< 0.00005

3.5.1 Farm-Level Permutation Test

One analysis challenge was the small and unequal number of events. There were 58 events from only 32 farms, and the events within a farm were not independent. The farm-level permutation test permutes at the farm level, taking into account the clustering, and because it is non-parametric, it does not rely on the data following any particular distribution. With the test, we randomly shuffled the group labels across farms (not events) 20,000 times and asked how often chance alone would produce a gap as large as 82% vs 17%. The resulting p-value was < 0.00005, indicating that we would almost never see such a gap between treatment and control due to chance.

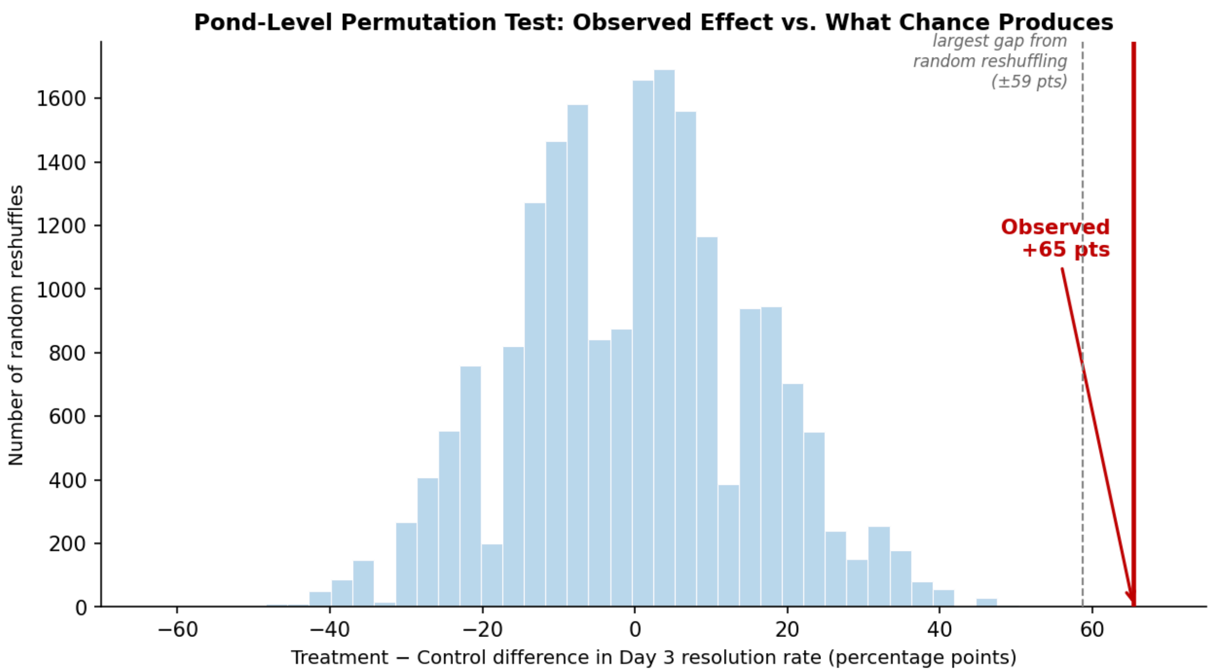

Figure 8. Results of the farm-level permutation test. The blue bars indicate the 20,000 random shuffles for farm labels. None reached the observed 65-point difference between treatment and control (red line).

3.5.2 Chi-Squared And Fisher's Exact Tests

Two simpler tests confirm the same result. The chi-squared test and Fisher's exact test both ask the same basic question: given the counts of resolved and unresolved events in each group, how likely is it that the gap we see happened by chance? The chi-squared test is the standard approach for this kind of comparison; Fisher's exact test is a more conservative version that is better suited when the numbers in some cells are small, as they are here (only 5 control events resolved). Both returned a p-value of less than 0.000001, meaning that if there were truly no difference between treatment and control, we would expect to see a gap this large less than once in a million random trials. The limitation of both tests is that they treat each of the 58 events as independent, when in reality events from the same farm may be related. The farm-level permutation test above corrects for this.

3.6 Relative Improvement of Water Quality

The statistical tests in the sections above asked whether a farm resolved, whereas this section asks how far water quality values moved. The measure is the out-of-range gap closed: the distance outside the in-range band at Day 0 minus the distance at Day 3 (positive = moved back toward the band). For instance, if DO was 2 mg/L below the range at Day 0 and only 0.5 mg/L below at Day 3, the gap closed by 1.5 mg/L. Each farm contributes one value (its mean across events), and each parameter keeps its own units. Since mg/L and pH points can't be pooled into one score without an arbitrary conversion, parameters were tested separately.

Using this gap definition, we calculated means as group averages of the per-farm values and compared them between treatment and control using Welch's t-test. Notably, outlier farms as defined in section 3.3.1 were removed from the analysis. A small p-value indicates the two groups improved by meaningfully different amounts, not just by chance. As a cross-check, we also ran the rank-based Mann-Whitney U test. This test is a method that ranks all values from lowest to highest across both groups and compares those ranks, making it less sensitive to any single extreme farm skewing the result. The test also gave a small p-value, suggesting the results are unlikely to be by chance.

The treatment group improved more on every parameter, while the control group's mean pH and ammonia drifted slightly further out-of-range over the follow-up period (Fig. 9). This pattern was robust to outlier removal (see section 3.3.1 for the definition of outliers). Ammonia was the one exception: under an outlier definition based on out-of-range events across any parameter, its result was borderline ($p = 0.066$), but under a definition based on ammonia-specific out-of-range events (only removing those farms that were outliers due to ammonia issues), the effect returned to significance ($p = 0.030$). Overall, these results suggest a consistent treatment benefit, though the ammonia finding warrants some caution given its sensitivity to the outlier definition.

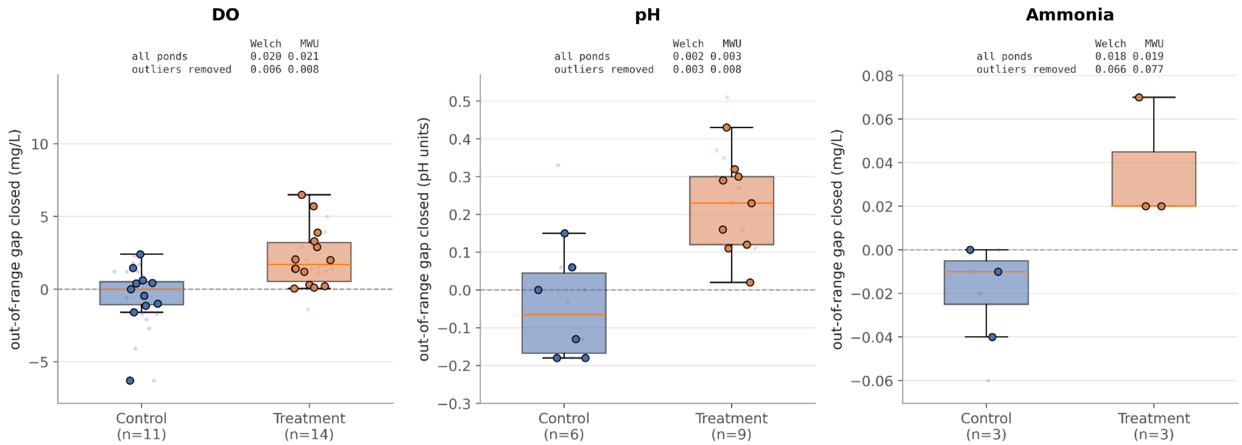

Figure 9. Magnitude of water quality improvement per farm from Day 0 to Day 3, by parameter. Each point is one farm's mean gap closed (the distance outside the acceptable range at the start minus the distance at Day 3). A negative number indicates a move away from the required range and a positive one a move toward the required range. Outliers were removed before plotting this figure.

3.7 Weather as a Potential Confounder

We expected weather to be among the primary natural drivers of water quality changes, and therefore checked whether seasonal conditions could explain or bias our results. We obtained hourly ERA5 data from Open-Meteo for the full study period and region (16.71°N, 81.10°E). This dataset is a globally standardized weather dataset that reconstructs historical conditions from satellite and weather station records. The study period fell in the dry summer season: daily mean temperatures averaged 29.4°C and daily highs averaged 36.3°C, peaking at 42°C in late April and May. Rainfall was sparse with only 69mm across the entire period and just 20 days recording more than 1mm. Solar radiation was consistently high. This confirms the protocol's summer characterization and means the study was conducted under relatively stable, predictable weather conditions rather than the more variable monsoon season.

For weather to bias the treatment-control comparison, the two groups would need to have experienced systematically different conditions during their out-of-range events. The data suggested that they did not differ much: mean temperature, maximum temperature, rainfall, and solar radiation were all statistically indistinguishable between treatment and control farms during their respective event windows (all $p > 0.06$). Pooled across both groups, none of the four weather variables predicted whether an event was resolved by Day 3 (all $p > 0.10$). A logistic regression adjusting for temperature and rainfall left the treatment effect essentially unchanged (the treatment odds ratio moved from 23.0 to 25.6) and neither weather variable was a significant predictor. This suggests that the large resolution gap between treatment and control is not a weather artifact. Weather conditions were balanced throughout the study and the treatment effect holds independently of them.

3.8 Self-initiated Actions (SIAs)

Farmers in both groups frequently took their own actions, not prompted by any recommendation from the study team, during out-of-range events. These self-initiated actions (SIAs) were recorded and warrant examination, both to characterize their frequency and types, and to assess whether they may account for some of the observed resolution difference between groups. Our SIA analyses, further described below, lack sufficient detail to draw a final conclusion on whether water quality improvements were driven by SIAs or corrective actions. While there is some indication, including a panel assessment by our welfare experts, suggesting that corrective actions are primarily responsible for water quality improvements, we cannot conclude this with certainty from the available data.

3.8.1 Frequency of Self-initiated Actions Compared to Baseline

Data from the regular ARA program (August 2025 – June 2026) provide a baseline against which study SIA rates can be interpreted. During routine visits, farmers recorded SIAs at approximately 31% of visits overall: 30% on visits where water quality was *in-range* and 40% on visits where at least one parameter was *out-of-range*. SIAs therefore appear to be part of routine farm management.

Within the study, SIA rates on routine visits (any initial, non-follow-up visit) were comparable across treatment and control farms and the ARA data. Table 5 shows a breakdown of the SIA rates across different scenarios and compared to the regular ARA numbers. Across both study groups, farmers applied SIAs in 36% of the measurements taken, similar to the 31% in the ARA dataset. The study's treatment and control groups exhibit similar rates for in-range visits and differ slightly in out-of-range visits with treatment farms applying SIAs more frequently (50%) than control farms (33%). The slightly higher rate in treatment farms on out-of-range visits could be because those farmers were informed of their water quality readings and given a corrective action on that day. They may have implemented further actions beyond the corrective actions. Control farmers, who didn't receive water quality readings or corrective actions, may have been less aware of the water quality issue and thus implemented fewer SIAs.

Table 5: Measurements where farmers reported taking self-initiated actions (SIAs) as a share of the total measurements taken (unit: farm-day). A limitation of this data is the difference in total measurements (n) across the baseline and study data.

	All days	In-range days	Out-of-range days	n (farm-day)
Routine ARA (Aug '25 - Jun '26)	31%	30%	40%	808
Study all (treatment and control)	36%	35%	41%	384
Study treatment	36%	33%	50%	177
Study control	37%	37%	33%	207

3.8.2 Self-initiated Actions Timing

Changing our view from measurement days to the full Day 0 to Day 3 event window, SIAs were recorded in 47% of control events and 61% of treatment events. This is higher than the 33% and 50% per-day figures in Table 5 because it counts an SIA taken anywhere across the three-day window, not only on the out-of-range day itself. This difference was not statistically significant (Fisher's exact $p = 0.31$). At the follow-up visits on Days 2 and 3, when most initial actions had already been taken, SIAs were recorded in 25% of control follow-up days and 23% of treatment follow-up days, reflecting broadly similar activity in both groups following water quality issues (Fig. 10).

One limitation of our data is that only 56% of the recorded self-initiated actions detailed a specific day of implementation. Many farmers did not remember exact days and instead gave us a wide range (e.g., in the last 0-7 days). The resulting limited data accuracy reduces our ability to interpret the impact of SIAs because we cannot determine whether there were differences in when exactly SIAs were implemented in relation to the out-of-range measurement.

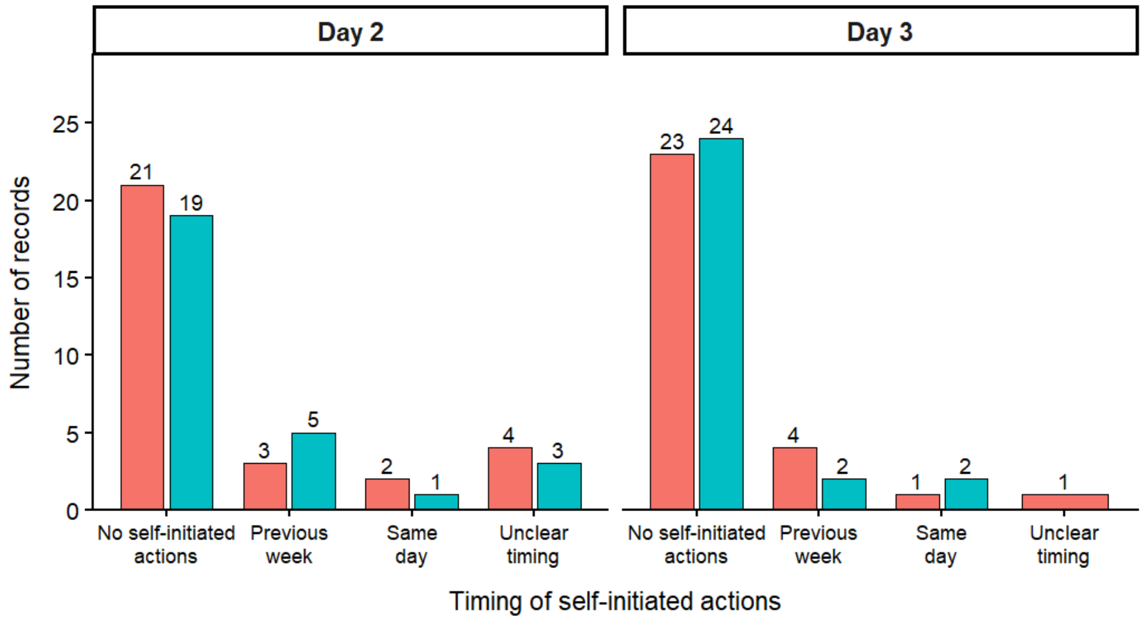

Figure 10: Distribution of self-initiated actions recorded in treatment (green) and control farms (red) across the Day 2 and Day 3 follow-up measurements. Bars indicate the number of follow-up observations in which no self-initiated action was reported, the action was implemented during the previous week, on the day of sampling, or the timing was unclear. Farmers may have implemented multiple SIAs and reported them separately on Day 2 and Day 3, leading to a difference in bars between these follow-up days.

3.8.3 Self-initiated Action Types

The types of SIAs recorded somewhat differed from FWI's implemented corrective actions. To analyze this data, we assessed corrective actions and SIAs at each event chain, including initial measurement and follow-ups on Day 2 and Day 3. If a farmer reported the same action twice, we did not double count it. If they implemented multiple actions, we listed all of them. The analysis suggested that “stop feeding” was the most common corrective action (58%) and the most common SIA in the control group. In the treatment group, “stop feeding”, water inflow, and water sanitizers were the most frequent SIAs. Beyond this, FWI corrective actions concentrated in a small set of targeted interventions, notably jaggery, present in 28% of corrective actions but entirely absent from SIAs. Farmer SIAs were spread more broadly across general management practices such as water inflow, water sanitizers, aeration, and probiotics. See Table 6 for a breakdown of the most common actions.

Table 6: Overview of the most common self-initiated actions (SIAs) and corrective actions (CAs), each counted once per event in which a farmer implemented it. A farmer could take more than one action per event, so a column's total may not equal the number of events and percentages are calculated within each column.

Action	FWI CA (n)	FWI CA (%)	SIA Treatment (n)	SIA Treatment (%)	SIA Control (n)	SIA Control (%)
Stop feeding	21	58%	5	17%	7	28%
Jaggery	10	28%	0	0%	0	0%
Zeolite	3	8%	1	3%	0	0%
Water inflow	0	0%	6	20%	3	12%
Water sanitizers	0	0%	5	17%	2	8%
Salt	0	0%	3	10%	1	4%
Lime	0	0%	3	10%	1	4%
Aeration (every day)	0	0%	1	3%	3	12%
Water probiotics	0	0%	2	7%	1	4%
Lice medicine	0	0%	1	3%	2	8%
Aeration (2 days)	1	3%	0	0%	1	4%
Other	1	3%	3	10%	4	16%
TOTAL	36	100%	30	100%	25	100%

The difference between the type of actions taken through SIAs and corrective actions presents a challenge when interpreting the results. Most treatment farmers who took an SIA chose an action that didn't overlap with the specific CA we gave them, making it harder to distinguish between corrective action and SIA impact. In many cases the different actions targeted different water quality parameters (e.g., adding jaggery for pH vs. turning on aerators for DO). We do not have the relevant data to conclusively determine whether SIAs or corrective actions improved water quality in these cases.

3.8.4 Association of Self-initiated Actions with Resolution

Several analyses examined whether SIAs were associated with water quality resolution, including a case-by-case comparison of each out-of-range event's SIA (if any) against the water quality parameter it targeted.

Within the treatment arm, events where the farmer's SIA specifically targeted the out-of-range parameter (e.g., stop feeding or aeration for a DO issue) resolved 100% of the time (10/10). Events with a "mixed" SIA, one that partially targeted the out-of-range parameter alongside an unrelated action (e.g., aeration alongside lice medicine), resolved 60% of the time (3/5). Events with no SIA at all resolved 73% of the time (8/11). Events where the farmer took a non-targeted routine action (such as water sanitizers) resolved 100% of the time (2/2). These sample sizes are very small and should be interpreted with care. They also cannot isolate what the SIA contributed: every treatment event received a corrective action, so these resolution rates reflect the corrective action and any SIA combined, not the SIA on its own. Within the treatment arm, adding a targeted SIA did not clearly beat taking no SIA at all and a gap that size is within what such small numbers can produce by chance. The cleaner test of whether SIAs work on their own is the control arm, where farmers received no corrective action.

The control group showed differing results. Events where the farmer took an SIA that targeted the water quality issue resolved 0% of the time (0/10), while events with no SIA resolved 19% of the time (3/16). Events with mixed SIAs saw water quality issues resolve 2 out of 3 times, though with only three events this figure is highly uncertain. One control event involved a purely non-WQ-targeted SIA and did not resolve. This generally lower rate follows the overall lower improvement rate of the control group compared to the treatment. The number of events is small and should thus also be interpreted with caution. It appears unusual that farmers' SIAs performed far better in treatment farms than in control farms. One possible explanation could be timing, as noted in 3.8.2. We were unable to obtain roughly half of the exact implementation dates for SIAs and it is possible control farmers' SIAs were applied too late in the window for us to detect their effect, while treatment farmers' corrective actions (and any accompanying SIA) were anchored to our fixed Day 0–1 timeline. We were not able to get certainty on this issue from the data we have.

The analysis on SIAs also included a mixed-effects logistic regression examining predictors of water quality resolution across all follow-up observations, with farm identity included as a random intercept, an adjustment that lets each farm have its own starting baseline, so the results aren't skewed by any one farm's unusual behavior. This logistic regression did not find a significant association between any SIA timing category and the probability of resolution (Fig. 11). Treatment status was a significant predictor of resolution (odds ratio = 4.08,[7] 95% confidence interval: 1.31–12.71; $p = 0.015$), as was Day 3 compared with Day 2 follow-up (odds ratio = 2.97, 95% confidence interval: 1.19–7.45; $p = 0.020$). This analysis did not identify evidence of an interaction between the treatment group and SIA timing ($p = 0.685$).

[7] Meaning treatment events were about four times as likely to resolve.

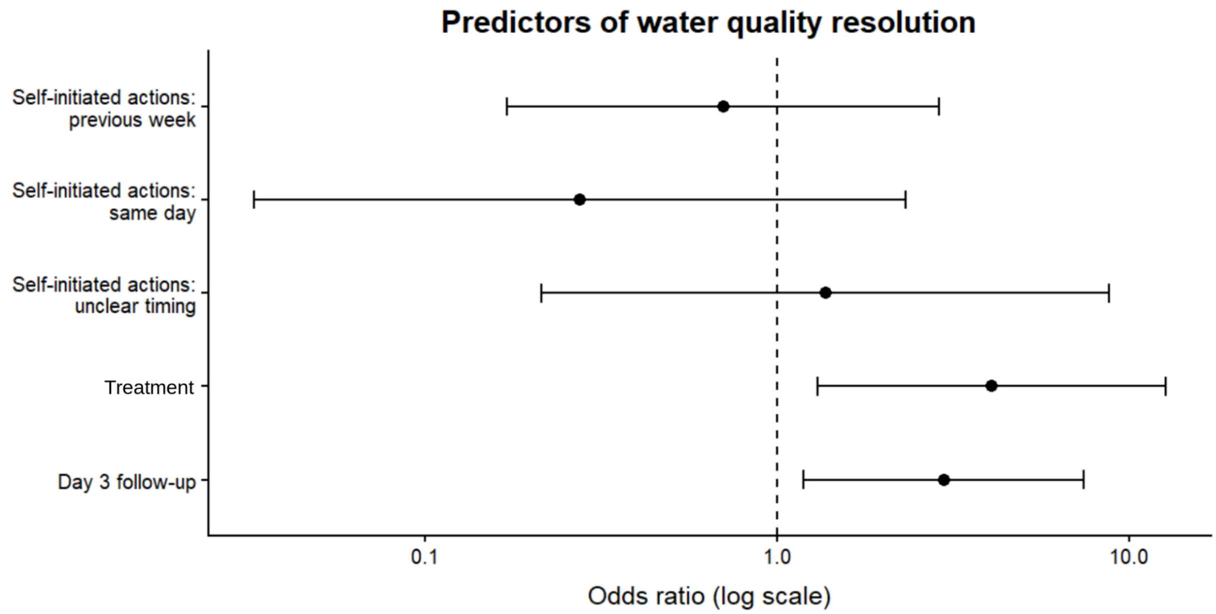

Figure 11: Mixed-effects logistic regression model evaluating predictors of water quality resolution. The points represent odds ratios, and the horizontal bars indicate 95% confidence intervals for each predictor. Farm identity was included as a random intercept to account for repeated follow-up observations within farms. Values to the right of the vertical dashed line (odds ratio = 1) indicate increased odds of water quality resolution, whereas values to the left indicate decreased odds.

Taken together, these analyses suggest that SIA timing alone does not explain resolution, and that the treatment effect holds after accounting for it. The case-by-case breakdown suggests the type of SIA, specifically whether it targets the same parameter as our corrective action, may matter more than SIA presence alone. The interpretation is subject to the limitations noted throughout the section, including small group sizes and the observational nature of SIA recording. Whether the pattern would hold in a study with larger numbers or more precise SIA documentation remains to be seen.

3.8.5 Self-initiated Actions Panel Assessment

We identified remaining uncertainty in cases where farmers implemented SIAs and staff-recommended corrective actions and it is mostly unclear which of these caused the ultimate welfare improvement. In regular ARA programming, we address this uncertainty through a panel of expert reviewers who decide whether they believe water quality has improved due to our recommended actions or other factors (e.g., weather or SIAs).[8] We applied this assessment method to the study data to evaluate whether our expert panel suggests a given event's resolved status was due to SIAs or corrective actions.

[8] For more details on our impact assessment method, see this blog post.

For the review, each panel expert received an anonymized spreadsheet with the 15 study events where the corrective action was fully implemented, the farmer also took an SIA, and the flagged water quality parameter was back in range by Day 3. Events where water quality did not resolve were excluded because there is no resolution to attribute. Reviewers independently selected one of four verdicts: "Corrective actions likely primary driver," "SIA likely primary driver," "Both contributed," or "Cannot determine" with a Low/Medium/High confidence and a brief written rationale. The full results from this review are available here.

Across the 45 individual votes (15 cases x 3 reviewers), the expert panel chose corrective actions as the primary driver 29 times (64%) and SIAs once (2%). Table 7 details these results further. Aggregating all reviews, 10 of 15 cases had a corrective action majority (6 unanimous; 4 by 2-1), one case had a "Both" majority and one case a "Cannot determine" majority. Three cases had no majority because for these each reviewer picked a different verdict. To account for how sure reviewers were, we also computed a confidence-weighted aggregate (weighting each vote Low = 1, Medium = 2, High = 3). On this basis the panel leaned strongly toward corrective actions, with a mean "lean" score of 0.86 out of 1.0 across the 15 cases and a mean attribution confidence of 88% (indicating how much of the panel, weighted by confidence, made a clear call instead of "cannot determine").

Table 7: Overview of expert panel's majority verdict on whether the corrective actions, self-initiated actions or both caused the improvement in water quality or whether they could not determine this from the available data. The last row shows the number of cases where the reviewers landed on a 3-way split without a majority consensus.

Majority verdict	Cases
Corrective action (unanimous, all reviewers agreed)	6
Corrective action (majority, 2 out of 3 reviewers agreed)	4
Both contributed (2 out of 3 reviewers)	1
Cannot determine (2 out of 3 reviewers)	1
No majority (3-way split)	3

A look at the reviewers' assessment rationale suggested that there was consistent reasoning about whether the corrective action and the SIA each had a chemically plausible route to fixing the flagged parameter (e.g., jaggery lowers pH; zeolite binds ammonia; water sanitizers don't affect DO), and whether the SIA's timing, where many were implemented days before the flagging visit, could realistically explain a Day-3 recovery. This rationale agreement for most cases may indicate that the near-unanimity on corrective action being the major driver is not an assessment bias but instead rooted in technical reasoning.

4. Limitations

While we try to mitigate bias and noise as much as possible, this study is field-based and thus comes with a number of limitations.

- **Previous treatment exposure**: The participating farms were all enrolled in the ARA prior to the study. It would have been extremely difficult, if not impossible, to recruit external farmers to participate in this study. Having selected ARA farmers presents the challenge of studying farms that potentially may have better water quality overall or otherwise modify their farm environment in ways that improve water quality differently than for non-ARA farms. E.g., if the ARA made them more aware of the importance of water quality and they now regularly apply probiotics, their water buffering capacity may simply be higher and out-of-range measurements may improve more quickly. Given that we use ARA farms for the control and treatment cohorts, some of this noise is accounted for, but the data cannot tell us how this study would fare with non-ARA farmers in one or both cohorts.
    - Relatedly, most farms had also participated in a previous iteration of this study. The farms that participated in the previous study were assigned to the opposite group. For example, a farm that was in the treatment group in last year's study was assigned to the control group in this study. This flipping of study groups could introduce some bias to the results but we expect this to not be more grave than the fact that all farms have already been part of the ARA before joining either of the studies.

- **Limited Geographic Representativeness of Farms**: Our Eluru ARA farms are likely *not* representative of other regions in Andhra Pradesh or India. However, it is unclear whether any farming region is. It appears that farming clusters adopt similar practices which are not common elsewhere and finding a truly representative region is thus difficult. The variability in farm characteristics (e.g., farming practices, ecological conditions, and farmer behavior patterns) has implications for how applicable the study results are to regions outside of Eluru. That being said, currently over ¾ of ARA farms are in Eluru and the study area is thus representative of the majority of farms in the program at this time.

- **Self-initiated actions**: We were unable to attribute the resolution of water quality issues solely to corrective actions. Farmers also initiated their own actions. We tracked these but were unable to determine their effect because we missed some implementation dates and because there was general uncertainty about what exactly drives water quality when multiple actions are implemented.

- **Parameter-Dependent Results**: DO makes up ~86% of out-of-range measurements (sometimes including other parameters out-of-range; DO-only was 43%) and we thus expect most corrective actions to be targeting DO. The study results are less descriptive of the effectiveness of pH and ammonia.

- **Limited Follow-Up Duration:** This study assesses only short-term (48–72h) changes in water quality parameters. Long-term effectiveness of corrective actions remains unmeasured. The study wasn't designed to evaluate the efficacy of SIAs outside the measurement days. It could be that self-initiated actions address water quality issues outside FWI's regular visit schedule.

- **Control Group Contamination Risk**: Despite best efforts to keep control farmers blind to water quality measurements and corrective actions, visiting the farm for two follow-ups on consecutive days may prompt the farmer to take actions to improve water quality when they otherwise wouldn't have. We expect to have caught most, if not all, of these by asking about SIAs but there is a chance that farmers weren't always honest about these.

- **Follow-up timing**: We used the follow-up on Day 3 for our primary analysis. However, during regular programming, the ARA's follow-ups may also be taken later, depending on when corrective actions are implemented. This limits the applicability of this study to everyday field conditions, which tend to be messier. We decided to standardize the follow-up timing because with the wide range of follow-up times that exists otherwise, running an RCT would have been impossible.

- **ARA changes**: With the recent adjustments to the ARA water quality ranges, we slightly changed the DO range to remove the upper bound and set the evening limit lower. The study is thus not fully representative of this adjusted program. Still, the headline results remain: When rerunning the analysis with the new ranges, the control and treatment gap remains large, with 30% and 88% of resolved out-of-range events, respectively.[9]

- **Seasonality**: The study was conducted during summer and pre-monsoon with little rain and high temperatures. The weather changes throughout the year, leading to different water quality parameters becoming more or less problematic throughout the months. We cannot fully extrapolate the study results beyond the assessed time period and further research is needed to verify the treatment effect across seasons.

5. Discussion

The purpose of this field-based randomized controlled trial was to provide evidence that FWI's core program, the ARA, is achieving its intended outcome of improving water quality at fish farms. Specifically, we tested if the provision of water quality testing and corrective actions to farmers results in improved water quality. The study's primary finding was that

[9] See this report for further details.

water quality improved at treatment farms for 23 of 28 (82%) out-of-range events, compared to 5 of 30 (17%) out-of-range events at control farms. This supports that the ARA is achieving its intended outcome: giving farmers water quality readings and corrective actions has the intended effect of improving water quality.

The treatment effect seems to require three days to show. During the first follow-up, on the second day after the initial poor water quality reading, treatment and control farms showed similar resolution rates (25% of out-of-range events improved at treatment farms, compared to 21% at control farms). By the second follow-up, on the third day, treatment farms' resolution rate increased dramatically to 82%, whereas the control farms' resolution rate dropped to 17%. It is unclear whether improvements last beyond the three days as the study did not test this. However, as the three-day follow-up is what most regular ARA measurements see, this treatment/control difference may suggest the effectiveness of our intervention.

The treatment effect holds despite various confounding factors, including weather. However, it is unclear how much farmers' SIAs drive water quality improvements. The study data did not provide sufficient insights to confidently discern whether our corrective actions or farmers' SIAs were ultimately responsible for water quality improvements. We have some indication that at least the combination of water quality readings, corrective actions, and SIAs works better than only SIAs but there is no absolute certainty.

Farmers reported implementing SIAs at a high frequency: 47% of the 30 out-of-range events at control farms had a SIA of some kind, and 61% of the 28 out-of-range events at treatment farms had a SIA separate from the corrective action that we recommended. There was considerable uncertainty around when SIAs were conducted, in relation to study visits (either the routine visits for testing water quality, or the follow-up visits). This uncertainty in timing is important, particularly given that our study highlights the importance of follow-up time in detecting improvements after an action (i.e., different resolution rates on Day 2 versus Day 3 follow-up). The study wasn't designed to assess the efficacy of SIAs outside of the ARA's follow-up time window (i.e., three days after the initial poor water quality reading). This introduces uncertainty as to whether SIAs contribute to improved water quality. It is conceivable that some of the SIAs from control farmers actually improved their water quality issues, just that we didn't detect improvements due to time window constraints. If this were true, the difference in resolution rates that we detected between the two cohorts (82% for treatment farms, 17% for control farms) may be less, thereby reducing the effectiveness of the intervention.

A noted limitation of the ARA is the low frequency of water quality measurements at each farm. At a given farm, the program only measures water quality and provides corrective action recommendations once a month. Water quality is inherently dynamic, but we are only able to intervene at each farm once a month. The study shows that a high percentage of farmers are conducting SIAs without FWI's intervention. This suggests that farmers may take actions more frequently than the ARA program does. If this were true, while the findings from our study may be solid in relation to study design, it is unclear whether farmers improve water quality themselves outside FWI monitoring visits. We had not conducted a needs assessment before starting the program in 2021, and thus we don't fully understand existing practices and the frequency of water quality issues without FWI intervention. The question of what specifically drives resolution, the water quality reading, the corrective-action recommendation, the specific action taken, or the dosage, cannot be fully resolved from this data alone and is a topic for future investigation.

While the high rate of SIAs recorded as part of this study presents questions around the true effectiveness of the program, we believe that cases where SIAs may have contributed to resolving the water quality issue do not invalidate the hypothesis tested in this study. Within the confines of the study—most specifically, retesting water quality three days after detecting an out-of-range event (the process followed as part of the ARA)—there is a much higher resolution rate in the treatment group (82% of OOR events resolved by Day 3 vs 17% for the control). This suggests that the program is having an effect. Paired with our evidence that the chosen water quality ranges improve fish welfare[10], this increases the evidence base that the ARA program is improving the lives of fishes.

6. Acknowledgments

This was one of the biggest field-based studies that FWI has ever undertaken. Conducting a randomized controlled trial in field conditions is challenging and we are deeply grateful to everyone who supported us in this endeavor. This study would not have been possible without the tireless efforts of our ARA field team: Ravi Teja, Nagaraju, and Chaitanya.

We are grateful for the inputs of Paul Monaghan (R&D department, FWI), Tom Billington (The Mission Motor), Lia Boldt (M&E Expert), and Daniela Lopez Cajiao (Impact Evaluation Specialist) in designing this study and Adam and Norma for completing the data analysis.

We also thank all other supporters who make this work possible. Your support is deeply appreciated.

[10] See this blog post for more details