Soil tests, vouchers, and the limits of site-specific fertilizer recommendations: Experimental evidence from Malawi

Thomas Assefa[*], Jonathan Atkinson[†], Hailemariam Ayalew[‡], Joachim De Weerdt[§],
Edwin Siyame[§], David J. Spielman[¶], Bjorn Van Campenhout[‖]

June 15, 2026

Latest version: https://github.com/bjvca/S2P/blob/master/S2PIEpaper.pdf

Abstract

Fertilizer subsidy programmes dominate agricultural spending in several African countries. Rising and volatile fertilizer prices have renewed concerns about the efficiency of such spending, prompting initiatives to tailor fertilizer recommendations to local soil conditions. We test this premise in Malawi, where the national subsidy programme distributes a standard package of urea and NPK fertilizer. In a cluster-randomized trial involving more than 2,000 households across 113 villages, we assign farmers to (i) a plot-level soil-test recommendation, (ii) the same recommendation combined with a voucher of subsidy-equivalent value, or (iii) a control group. The recommendation alone has no effect on fertilizer use, input choice, or maize yields. When paired with a voucher, fertilizer use and maize yields increase substantially, with yields rising by roughly one-third; at the maize prices prevailing in our study area, however, the value of this additional production falls short of the face value of the voucher, let alone the value of the voucher plus the cost of individualized soil test. For farmers themselves the package is nonetheless profitable, since the voucher largely displaces fertilizer they would have bought anyway and so lowers their own input spending, and because many households are net maize buyers, valuing the additional output at retail rather than farmgate prices narrows the shortfall. The yield gains do not reflect adoption of site-specific prescriptions: voucher recipients predominantly purchase urea and NPK in proportions closely matching the standard subsidy bundle, regardless of what their soil test recommends, and compliance with recommended alternatives such as potassium, lime, or calcium ammonium nitrate is negligible. Because the observed gains come almost entirely from nitrogen intensification while phosphorus, potassium, and lime gaps remain uncorrected, the measured yield response is a lower bound on what full-compliance fertilization could deliver. Relaxing financial constraints thus increases input use and productivity, but information alone does not redirect behaviour toward precision fertilization. Effective subsidy reform will require addressing broader supply- and demand-side constraints, including input availability, farmer familiarity with recommended products, and the practical implementation of site-specific recommendations, not only improving agronomic information.

keywords: input subsidies, fertilizer vouchers, smart subsidies, crowding out, soil test, agro-dealer supply, Malawi

JEL codes: O13, Q12, Q16, Q18, C93

1 Introduction

Public input subsidy programmes have long been the largest single line of agricultural-sector spending in much of Sub-Saharan Africa. Jayne et al. (2018) identified ten countries operating such programmes at scale

---

[*] Agrifood Innovation and Resilience Unit, IFPRI, Nairobi

[†] Solidaridad Southern Africa, Johannesburg, South Africa

[‡] Development Strategy and Governance Unit, IFPRI, Nairobi

[§] Development Strategy and Governance Division, IFPRI, Lilongwe

[¶] Agrifood Innovation and Resilience Unit, IFPRI; Washington, DC; Washington University in St. Louis, MO

[‖] Markets, Trade, and Institutions Unit, IFPRI, Leuven; corresponding author: b.vancampenhout@cgiar.org at the time of their review and estimated that the programmes absorbed roughly thirty percent of public agricultural expenditure in those settings. Two decades of farm-panel evidence document a recurring set of performance problems: low or negative benefit-cost ratios in several large programmes, distortion of commercial input markets through parallel public distribution channels, substantial crowding-out of commercial fertilizer purchases by subsidised receipts, regressive and politically captured targeting, fiscal entrenchment that makes the programmes difficult to wind down once established, and concerns about environmental externalities arising from inefficient fertilizer use (Holden, 2019; Mather and Jayne, 2018; Basurto, Dupas, and Robinson, 2020; Galloway et al., 2008). Recent increases in global fertilizer prices have amplified concerns about the efficiency and fiscal sustainability of such spending, strengthening the case for improving the return to each unit of subsidized fertilizer (Hebebrand et al., 2026). This combination of problems has motivated proposals to redesign public fertilizer support, or to repurpose part of this spending toward public investments with higher returns (Hill and Resnick, 2025).

A central concern in this debate is that uniform input bundles may result in substantial misallocation of fertilizer across heterogeneous soils and production environments (African Union, 2024). In Malawi, the Farm Input Subsidy Programme (FISP) — called the Affordable Inputs Programme (AIP) at the time of our study[1] — distributes a standard 100 kg bundle of nitrogen-phosphorus-potassium (NPK) basal blend and urea topdressing to beneficiary households, regardless of plot-level soil conditions (Benson et al., 2024). While such programmes can raise input use and production, their cost-effectiveness depends on whether the fertilizer applied is well matched to local nutrient requirements. This concern has become more salient in a high-price environment, where inefficiencies in input use translate directly into higher fiscal costs per unit of output. One proposed response is to complement or replace uniform bundles with site-specific fertilizer recommendations derived from soil tests, with the aim of improving the productivity and efficiency of fertilizer use by tailoring inputs to local conditions.

Despite its prominence in policy discussions, there is limited causal evidence on whether site-specific recommendations meaningfully alter fertilizer use patterns in smallholder settings, particularly in environments where standard input bundles are well established. In particular, it remains unclear whether improved information can redirect farmers toward more efficient input combinations, or whether existing market conditions, input availability, and farmer preferences limit responsiveness to such recommendations. We examine this question experimentally.

At a basic agronomic level, there is a clear mismatch between blanket fertilizer recommendations and heterogeneous soil conditions. Blanket fertilizer recommendations are poorly matched to local soil conditions, often omit limiting micronutrients, and ignore the substantial within-field heterogeneity in soil chemistry and crop response (Johnston and Bruulsema, 2014). However, translating this technical case into effective farmer responses has proven more difficult. Direct experimental tests of soil-test recommendations without accompanying input or financial support have produced a mixed pattern. Ayalew, Chamberlin, and Newman (2022) find that soil-test recommendations alone raise fertilizer use and yields in Ethiopia, closing the macronutrient gap by roughly 19 kg/ha. Beg, Islam, and Rahman (2024) find little evidence that plot-specific recommendations affect input use or yields in Bangladesh. Most closely related to our setting, Nyondo et al. (2025) report a cluster RCT in central Malawi in which soil-test-based extension advice raises adoption of complementary soil-management practices (composting, residue incorporation, and timely planting) but does not move the use of commercial inorganic fertilizer (such as Urea, DAP or NPK), a pattern that is consistent with the null recommendation-only result we document below. Berazneva et al. (2023) document low farmer willingness to pay for soil-test information in Malawi, suggesting that even well-designed information products would scale poorly through private channels. A digital-advice literature on Nigeria and elsewhere reports heterogeneous behavioural responses to personalised site-specific soil nutrient management advice. (Arouna et al., 2021; Oyinbo et al., 2022; Maertens et al., 2023).

One possible explanation is that the technical limitations run deeper than delivery or uptake alone. Some have questioned whether soil-test recommendations can deliver on their own terms, even before any delivery or financing issue arises. Schut and Giller (2020) argue that soil tests cannot in practice deliver what they promise: within-field heterogeneity together with measurement error in the test itself leave the residual signal

---

[1] The programme has undergone several name changes over time; Benson et al. (2024) trace Malawi's input subsidy programmes back to the early 1950s. The large-scale subsidy programme implemented from 2005/06 was called the Farm Input Subsidy Programme (FISP), but was rebranded as the Affordable Inputs Programme (AIP) in the 2020/21 agricultural season. Following a change in government in late 2025, it was renamed back to FISP. During our study period, it operated as AIP.

too weak to support genuinely field-specific recommendations.

But even technically sound recommendations must be interpretable and actionable at the farm level. Evidence from at-scale soil-information programmes (India's Soil Health Card scheme in particular) points to a complementary design-related constraint: recommendations may be difficult for farmers to interpret, eroding the marginal value of the underlying soil data (Gars et al., 2025). This suggests a separate constraint: the message has to be translatable into something a farmer will act on. Chandrasekhar et al. (2022) show that a tangible, simple, transparent tool (a colour-coded measuring spoon calibrated to the recommended fertilizer dose) diffuses through smallholder networks far better than the quantitative recommendation it operationalises, because farmers share what they can credibly vouch for. In the same spirit, Islam and Beg (2021) show in Bangladesh that basic rule-of-thumb instructions tied to a leaf colour chart improve urea management, reducing use by 8 percent without compromising yield, suggesting that simplification rather than precision is often the actionable margin. The implication is that the form in which recommendations reach farmers - whether a printed prescription, an extension-worker conversation, or a physical artifact - is itself a core design feature, not merely an implementation detail.

Against this background, more encouraging results emerge when the recommendation is paired with cash, a voucher, or a discount. Harou et al. (2022) find in Tanzania that recommendation-plus-voucher bundles raise fertilizer use and direct purchases toward the recommended product while neither component alone does; the mechanism runs through input *choice* rather than total quantity, with ammonium-sulfate uptake rising from 4 percent in the voucher-only arm to 50 percent in the bundled arm despite 92 percent of plots being sulfur-deficient. Corral et al. (2020) find that recommendation-plus-grant raises yields in Mexico by twelve to seventeen percent while the recommendation alone produces only a modest adoption shift. Closer to our setting, Aggarwal et al. (2024) show in Malawi that cash transfers paired with near-village input fairs raise productive investment substantially above cash alone, with most fair purchases going to fertilizer, consistent with a complementarity between cash and last-mile access even when no agronomic targeting is attached. The conceptual common ground in this strand is that information and finance behave as complements: profitable technologies may not be adopted when households face binding credit, liquidity, or confidence constraints, a pattern familiar from the broader technology-adoption literature (Suri, 2011; Jack, 2013; Croppenstedt, Demeke, and Meschi, 2003; Duflo, Kremer, and Robinson, 2011; Beaman et al., 2013; Van Campenhout et al., 2025).

Our design contrasts Malawi's fixed-input subsidy model with a value-matched voucher that relaxes those constraints. That contrast between Malawi's fertilizer subsidy programme and a value-matched voucher also connects to the broader transfer-design literature on whether public support should be delivered in restricted in-kind form or as a more fungible instrument. In practice, Malawi implements a quasi in-kind subsidy programme: it fixes the products, the package size, the channel, the co-payment, and the bundle composition, leaving the beneficiary essentially no margin to substitute. Comparative evidence on in-kind versus cash or voucher transfers in low-income settings is mixed and tends to find that in-kind delivery binds beneficiary choice only when the targeted good is one beneficiaries would not have bought otherwise (Aggarwal et al., 2024). In the food-assistance literature, recipients are typically inframarginal: in-kind food transfers move consumption little differently from cash of equal value (Hoynes and Schanzenbach, 2009; Cunha, 2014), and where modalities do diverge it is usually on dietary composition rather than total consumption (Hidrobo et al., 2014; Aker, 2017). The Malawi fertilizer subsidy literature has documented partial crowd-out of commercial purchases by in-kind receipts (Ricker-Gilbert, Jayne, and Chirwa, 2011) but has not, to our knowledge, directly tested whether the composition restrictions embedded in the in-kind bundle bind at equal fiscal cost.

This paper reports on a cluster-randomized intervention in Malawi, a setting where FISP (then called AIP) distributes a 100-kilogram fertilizer bundle (one bag of NPK basal blend and one bag of urea top-dressing) at heavily subsidised prices to a rationed share of households each season. Onto this institutional backdrop we experimentally introduce two policy alternatives. The first is a soil-test-based, site-specific fertilizer recommendation, delivered to farmers by private-sector extension agents for a farmer-selected plot, with no fertilizer transfer attached. The second is the same recommendation paired with a 170,000 Malawi Kwacha (MWK) voucher (approximately USD 100 at prevailing exchange rates) redeemable for fertilizer at the nearest Farmers World shop, indicated on the voucher itself. The face value of the voucher is set to match the implicit subsidy in the FISP package, but the voucher is strictly more fungible along the product margin: any fertilizer product, any package size, no co-payment obligation. The design functions as a revealed-preference test of whether the composition restrictions embedded in FISP bind: at equal fiscal cost, do households choose a different bundle when allowed to?

The trial enrolled 18 households in each of 113 villages drawn from four districts in central Malawi (Kasungu, Dowa, Ntchisi, Mchinji), for a total sample of 2,034 households. The intervention (soil sampling, soil-test results, the accompanying recommendation, and (in the recommendation-plus-voucher arm) the voucher) was delivered to sampled households in October and November 2024, ahead of the 2024/25 main planting season. Endline data were collected in June 2025, after the 2024/25 main-season harvest. The endline measures fertilizer use (total and by product), nitrogen, phosphorus, potassium and sulfur application, maize yield, value of production, profit, FISP receipt, and a panel of complementary soil-management practices. Fertilizer-use data are collected on two plots per household (the soil-test plot and a second plot drawn at random from the household's other plots), which lets us separate household-level substitution against FISP from within-household reallocation across plots. Voucher redemption is captured directly through agro-dealer scans of a unique QR code on each voucher.

This paper contributes to three strands of literature. First, it provides experimental evidence on whether soil-test-based recommendations can alter fertilizer use under a large-scale subsidy regime. Second, it tests whether relaxing financial constraints through a flexible voucher induces adoption of site-specific fertilizer prescriptions. Third, it informs ongoing debates on the redesign of fertilizer subsidy programs by distinguishing between the effects of information, liquidity, and product choice constraints.

Four sets of findings emerge. First, recommendation-only assignment produces no detectable change in any outcome: fertilizer use, nutrient application, maize yield, returns, or FISP receipt. Information alone does not move behaviour in this setting. This null replicates the modal finding in the comparable literature (Harou et al., 2022; Beg, Islam, and Rahman, 2024; Corral et al., 2020).

Second, on the soil-test plot the voucher raises total fertilizer use by about 29 to 31 kilograms and raises maize yield by about 19 percent in total production terms.

Third, the composition of the voucher-induced additional fertilizer reveals that the soil-test recommendation does little of the work. The voucher-induced increment is overwhelmingly 18 kg of urea and 19 kg of NPK 23:10:5, a 50/50 split that exactly mirrors the FISP bundle the voucher is partially displacing. Where the recommendation differs from the modal Malawi maize prescription, farmer behaviour does not move. Across the products that differentiate the soil-test recommendation from the standard Malawi maize bundle, compliance is at or near zero, and exact-bundle compliance is not only near zero but is significantly lower in the voucher arm than in the soil-test-only arm: the voucher induces application of familiar but non-recommended products, which mechanically pushes farmers further from full compliance. The yield gain in the voucher arm is therefore generated by familiar fertilizer at familiar proportions, financed by a commercial-channel transfer of FISP-equivalent value; the soil test does not affect direct farmer behaviour in this setting. This finding suggests that liquidity constraints alone are insufficient to induce precision fertilizer adoption when recommended products differ substantially from farmers' established fertilizer portfolios.

Fourth, the voucher raises farmer profits through cost savings rather than value creation, and does not pass a simple cost–benefit test at farmgate prices. Recommendation-plus-voucher households see large private profit gains driven by reduced own input spending as voucher fertilizer displaces commercial purchases, but the induced production increase is too small to justify the voucher cost when valued at median maize prices, with break-even only at substantially higher retail-equivalent prices.

The decomposition of the voucher arm tightens the interpretation of the recommendation-plus-transfer literature. Field studies of recommendation-paired transfers, beginning with Harou et al. (2022), have read the bundle's productivity gains as evidence that information "becomes actionable" once finance is attached. In our setting it does not become actionable on the margins where it would have to: the recommended topdressing product is bought less, not more; the prescribed potassium products, lime, and MAP are not bought at all; and the within-arm product composition is statistically indistinguishable from a scaled-down FISP bundle. The voucher's productivity effect is generated by additional familiar fertilizer, not by precision-recommended fertilizer, so a recommendation-paired voucher in this setting is observationally close to a generic untargeted fertilizer transfer of equal value.

Reform proposals that present recommendation-paired vouchers as a smart, soil-aware alternative to FISP rely on the assumption that farmers given a recommendation and the money to act on it will apply the recommended inputs. In our data they do not. We cannot, from this experiment alone, pinpoint which constraint is doing the work: agro-dealer product depth, the form in which the recommendation is communicated, farmer trust in a prescription that diverges from the modal Malawi maize package, risk preferences over an unfamiliar product mix, or some combination of these. What we can say is that the recommendation-paired voucher does not, in the supply, information, and policy environment we observe, translate into adoption of the prescribed products. A reform that swaps FISP for a recommendation-paired voucher without first addressing the constraints that prevented recommendation-following in our setting will not deliver precision fertilization. It will deliver FISP through a different channel, at higher administrative cost. The case for routing fiscal support through commercial channels can still be made on targeting (Basurto, Dupas, and Robinson, 2020) and market-building grounds, but it cannot be made on the agronomic-precision grounds that animate the smart-subsidy framing.

The remainder of the paper is organised as follows. Section 2 states the two research questions and lays out the theory of change linking information, finance, and the agro-dealer supply chain to fertilizer choice and yield. Section 3 describes the intervention and the assignment. Section 4 presents the empirical specifications. Section 5 situates the trial in Malawi's fertilizer policy and agro-dealer landscape. Section 6 describes the study sample and reports baseline balance. Section 7 reports the main results, covering implementation, recommendation compliance, the voucher-induced product mix, fertilizer and nutrient use, FISP substitution, plot specificity, maize production and yield, and economic returns. Section 8 concludes.

2 Research Questions and Conceptual Framework

The paper asks two questions. First, does providing a site-specific soil-test recommendation change input use, FISP displacement, product- and nutrient-level compliance, maize production, and economic returns? Second, does pairing that recommendation with a voucher of fixed monetary value, redeemable at the nearest Farmers World shop for any fertilizer, change those outcomes further? The two questions imply three comparisons, all reported throughout: the recommendation alone versus control, the incremental effect of adding the voucher, and the combined package versus control. The first two identify behavioural margins; the third gives the per-beneficiary yield, production-value, and profit impact of the recommendation-plus-voucher package. A key contribution of the study is to distinguish between adoption of fertilizer in general and adoption of the specific fertilizer products prescribed by soil-test recommendations. This distinction is important because increased fertilizer use does not necessarily imply adoption of precision nutrient management practices.

The intervention we designed to answer these questions relaxes several constraints. A soil test can improve the agronomic content of the recommendation, but that recommendation affects production only if it is delivered, understood, affordable, translated into feasible product choices, and implemented on the plot. We organize the analysis along this chain, but distinguish three groups of links that play different roles in the design: implementation preconditions that had to be executed for the trial to be informative, behavioural mechanisms that connect assignment to behaviour, and the causal chain from behaviour to production and returns.

Implementation preconditions. Two links are not behavioural responses to treatment but conditions for the experiment to identify anything at all. The first is *recommendation generation*: the soil test translates plot- and crop-specific information into a recommended nutrient package and corresponding fertilizer products, and a site-specific recommendation can fail if it is agronomically inappropriate, too complex, or based on products that are not locally feasible. The second is *recommendation delivery*: farmers must actually receive the soil report, the associated advice, and, in the voucher arm, the voucher itself, since a null intent-to-treat effect has a different meaning if the recommendation never arrived than if it arrived but was not acted upon. We treat both as first-stage and compliance diagnostics, documenting Agronaut visits, recommendation receipt, and voucher receipt or redemption before interpreting any downstream outcome.

Behavioural mechanisms. Two further links sit between assignment and behaviour and motivate the two-arm design. The first is *farmer comprehension*: even when advice is delivered, farmers may not understand the recommendation or may not find it actionable in their decision environment, so an information intervention can fail either because the information does not arrive or because it is not usable. The second is *affordability and actionability*: site-specific recommendations typically imply changes in product choice and input quantities, and farmers may understand the recommendation but still be unable to purchase the recommended inputs under high fertilizer prices, liquidity constraints, and imperfect product availability. The voucher arm is designed to test this margin: it holds the recommendation component fixed and adds a fixed-value voucher redeemable at the nearest Farmers World shop (indicated on the voucher) for any fertilizer. Because the voucher is fungible across fertilizer products, it does not mechanically force the recommended bundle; rather, it relaxes the household's fertilizer-purchasing constraint and lets the household decide whether to spend the resulting purchasing power on the recommended bundle or on other fertilizers. The voucher is also strictly more fungible than FISP itself, which at equal transfer value restricts product (urea plus NPK basal blend), package size (50 kg bags), retail channel, and composition (fixed 50/50 split), and requires a cash co-payment. The voucher relaxes all of these alongside the affordability constraint, so predictions for T2 versus an FISP-equivalent counterfactual diverge by which constraint is doing the work: if the binding constraint is purchasing power, the voucher-induced bundle should look like the FISP bundle; if the FISP composition restrictions also bind, the voucher bundle should diverge from it: toward the recommendation if the recommendation is what farmers would buy, or toward whatever products farmers and the local agro-dealer network prefer when the restrictions are lifted. The incremental effect of adding the voucher to the recommendation therefore captures the combined effect of (i) lower own-cost on fertilizer in general and (ii) any reallocation of that purchasing power toward or away from the recommended products. Because the study population also has partial access to FISP, the voucher overlaps with an existing fertilizer transfer rather than relaxing an unsubsidised price, so an increase in fertilizer use on the test plot need not imply an equivalent increase in total household input use; this motivates the displacement analysis below.

Outcome chain. Three links translate behavioural response into agronomically and economically meaningful effects. The first is *product-level implementation*: a farmer may increase fertilizer use without applying the products named in the recommendation, so we distinguish product compliance from a generic fertilizer-use response. The second is *nutrient-level implementation*: product compliance and nutrient compliance need not coincide, since a farmer can reduce a nitrogen shortfall by applying urea while still failing to implement the full recommended bundle or to close phosphorus and potassium gaps, so we separately examine actual nutrient application, application error, and nutrient shortfall.[2] The distinction between product and nutrient compliance is central to the interpretation: the policy question is not only whether farmers use more fertilizer, but whether they can implement the specific product and nutrient bundle implied by the site-specific advice. The third is *production and returns*: changes in product choice and nutrient application are intermediate outcomes, and the central economic question is whether they raise maize yield, whether the implied value of production rises once output is priced at market rates, and whether the net effect on farmer profit is positive once the household's own spending on inputs is taken into account.

A separate analytical concern, which cuts across the outcome chain, is displacement. The recommendation-plus-voucher arm operates inside a partial input-subsidy regime (FISP) and is delivered on a designated test plot, so an observed increase in fertilizer use can in principle reflect reallocation across programmes or across plots rather than additional input use. We therefore examine FISP substitution and across-plot displacement as explicit components of the outcome chain rather than as residual robustness checks.

The results section is organised along this chain. Sample realization and soil-test linkage and uptake and treatment adherence document the implementation preconditions. Recommendation compliance and the voucher-induced product mix examine product-level implementation and, together with the incremental contrast between the recommendation-plus-voucher and recommendation-only arms, inform the comprehension and affordability mechanisms. Fertilizer and nutrient use address the use margin and map the quantity response into macronutrient application, while FISP substitution and the test-plot-specificity check examine displacement across subsidy sources and plots. Maize yield and value of production, costs, and profits close the chain at production and returns.

---

[2] For each nutrient $n \in \{N, P_2O_5, K_2O\}$ we compute (i) actual application in kg/ha, (ii) absolute application error, defined as the absolute difference between actual and recommended kg/ha, $|q_n^a - q_n^r|$, which penalises both under- and over-application, and (iii) shortfall, defined as the positive part of the recommended-minus-actual gap, $\max\{q_n^r - q_n^a, 0\}$, which captures under-application only. Recommended kg/ha are derived from the product-specific recommendation; actual P and K are expressed in fertilizer-grade P_2O_5 and K_2O units to match recommended quantities.

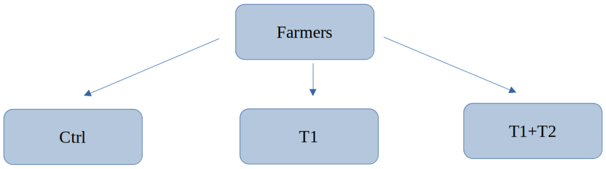


Figure 1: Experimental layout

3 Experimental Design and Treatments

The study employs a randomized controlled trial (RCT) design involving 113 villages across four districts in Malawi (Kasungu, Dowa, Ntchisi, and Mchinji). The villages are randomly assigned to one of the following three groups and the layout of the experiment is illustrated in Figure 1.

- **Treatment 1 (T1)** households receive a soil report based on a soil test conducted on a farmer-selected plot for the crop the farmer plans to grow on that plot.
- **Treatment 2 (T2)** households receive the same soil test, recommendation, and advice as T1 households, plus a voucher redeemable at the nearest Farmers World shop (indicated on the voucher) for any fertilizer product(s), with face value equal to the implicit subsidy a household would have received under the FISP.
- **Control (C)** households do not receive a customized soil-test-based fertilizer recommendation or voucher.

For the treatment arms, we cooperate with the Farm Services Unit (FSU) of a large private-sector fertilizer company operating across Malawi, Mozambique, Zambia, and Zimbabwe. The FSU was established to extend soil science and agronomic advisory services to farmers of all scales, from smallholders to commercial operations, and fields roughly 120 trained extension agents (known as Agronauts) in Malawi alone. Its core offering combines soil sampling, spectral soil testing with rapid turnaround, site- and crop-specific nutrient recommendations, customized fertilizer blending, and farmer training delivered through on-farm visits and village-level clinics.

Farmers in all three arms (including control) are visited by an enumerator accompanied by a trained Agronaut, offered a soil test on a plot of their own choosing, and asked which crop they intend to grow on that plot. All farmers are told up front that the sample will be analysed and that they will receive a soil report and recommendation based on the test. The arms differ only in when the report is delivered: T1 and T2 farmers receive it roughly one month later, before the planting season, while control farmers receive theirs only in the following season, after the study's harvest.[3] To separate household-level substitution against FISP from within-household reallocation across plots, the endline collects plot-level fertilizer-use data on two plots in every household: the soil-test plot and one non-test plot drawn at random by the survey instrument from the household's remaining plots. The random plot is not restricted to maize and reflects the household's overall plot portfolio; we use it both to test for displacement of FISP fertilizer off the recommended plot and to construct a household-level FISP total that sums fertilizer captured on both surveyed plots.

The soil samples are analyzed and a soil report is generated for every household.[4] The report is delivered as a two-page document. The first page opens with a graphical comparison of the farmer's expected current yield against the target yield, followed by a short plain-language fertility summary that names the macro- and micro-nutrients most likely to be limiting yields. Below this, a fertilizer recommendation table—specific to the farmer's intended crop—lists the recommended product, application rate (in kg/ha, kg/acre, and

---

[3] Because plot and crop choices occur under the same information set across arms, the test plot is balanced in expectation between treatment and control, and plot-level comparisons admit a standard intention-to-treat (ITT) interpretation. Farmer agency over plot and crop choice is therefore preserved without sacrificing identification.

[4] We used Crop Nutrition Laboratory Services Ltd (Cropnuts), a leading agricultural testing laboratory & agronomy advisory services company in East Africa. The model underlying the spectral test was calibrated on wet-chemistry analyses of 22,000 samples from across Malawi.

50-kg bags/acre), and timing for the planting and top-dress stages, with comments delivered in Chichewa where appropriate. The second page reports the underlying soil-test results: measured values and units for pH, available P, exchangeable K, calcium, magnesium, iron, manganese, boron, copper, zinc, cation exchange capacity, total nitrogen, organic matter, and the carbon-to-nitrogen ratio. Each parameter is also assigned a status on a four-point scale—Very Low, Low, Adequate, High—rendered as a color-coded bar (red, orange, green, blue) with accompanying emoticons.

Farmers in T2 villages additionally receive a voucher redeemable at the nearest Farmers World shop, indicated on the voucher itself, and fungible across the fertilizer products stocked there. "Nearest" was operationalised using OpenRouteService travel-time routing under a motorbike-or-bicycle profile, applied to each sample village and the universe of Farmers World shop locations. The face value of 170,000 MWK, approximately USD 100 at the official exchange rate at the time of the intervention, was calibrated to the implicit subsidy delivered by the FISP in that season. Eligible FISP households could buy one 50-kg bag of urea and one 50-kg bag of NPK at a co-payment of 15,000 MWK per bag against a market price of roughly 100,000 MWK per bag, implying an implicit subsidy of about 85,000 MWK per bag, or 170,000 MWK for the two-bag package. The voucher thus transfers an FISP-equivalent amount of fertilizer purchasing power, but without restricting how the funds are spent or tying them to the site-specific recommendation. Beyond value-matching, the voucher is strictly more fungible than FISP along several dimensions: any fertilizer product (not only urea and the NPK basal blend), any package size sold by the dealer (not the fixed 50 kg bag), and no co-payment requirement; A household that prefers a different composition or quantity at the same transfer value can express that preference here but not under FISP. Each voucher is personalized with the recipient's name and a unique redemption identifier and carries a QR code that the agro-dealer scans at the point of sale; an example is shown in Figure 2.


FARMERS WORLD
Malawi's Growing Power
AGORA
"Helping to grow your future"
fsu Farm Services Unit

K170, 000

INPUTS VOUCHER

Redeem Fertilizer from
Mponela Farmers World shop.

FWM-ML-51789
Daise Andisen

Figure 2: Example fertilizer-input voucher issued to T2 households. Face value is 170,000 MWK, redeemable at the nearest Farmers World shop (indicated on the voucher) for any fertilizer product stocked there. The voucher is personalised with recipient name and a unique redemption identifier (QR-coded for point-of-sale scanning).

Farmers in the C group of villages do not receive the soil report, the recommendation, or the voucher during the study period. As described above, however, they undergo the same soil-sampling visit and plot- and crop-selection protocol as T1 and T2 households; their reports are simply deferred until after the study's harvest, so that the arms differ only in the timing of feedback and in the voucher. The control-group soil samples therefore serve two purposes: they preserve a comparable plot-level counterfactual by holding the sampling experience constant across arms, and they provide diagnostic data on baseline soil conditions in the sample. The control group therefore serves as the omitted comparison group for the ITT estimates of the two treatment packages.

4 Empirical Specifications

The randomized assignment identifies the intent-to-treat effects of the recommendation and the recommendation-plus-voucher packages relative to control; the intermediate outcomes in Section 7 trace the process through which the packages did or did not become agronomically and economically meaningful.

The main empirical analysis estimates intent-to-treat effects of village-level assignment. For outcomes observed for the control group and both treatment arms, we estimate ordinary least squares (OLS) regressions of the following form:

$$Y_{iv} = \alpha + \beta_1 T1_v + \beta_2 T2_v + \varepsilon_{iv} \tag{1}$$

where Y_{iv} is the outcome for household i in village v. $T1_v$ and $T2_v$ are mutually exclusive village-level assignment indicators for the recommendation-only and recommendation-plus-voucher arms, with the control group omitted. The coefficient β_1 is the ITT effect of assignment to T1 relative to control, and β_2 is the ITT effect of assignment to T2 relative to control. The incremental voucher margin is tested directly as $H_0 : \beta_2 = \beta_1$ and reported in the tables as *p-value: T2 = T1*. This specification is used for the main fertilizer-use, AIP fertilizer use, test-plot displacement, nutrient-use, yield, economic-return, and appendix secondary-outcome tables. Standard errors are clustered at the village level, the level of random assignment.

The compliance tables use a different estimand because their outcomes are defined relative to treatment recommendation records. Product-level compliance and nutrient application error are estimated only among T1 and T2 households with valid treatment recommendation records and usable fertilizer-application data. In those tables the reported coefficient is the unadjusted difference between T2 and T1. This comparison asks whether adding the voucher changes implementation conditional on access to the recommendation package.

The analysis is organized by outcome family rather than by a single aggregate index. Recommendation delivery, product compliance, nutrient compliance, fertilizer use, yield, and profits answer different questions along the causal chain and should not be collapsed into one average effect. The main tables report raw p-values for the pre-specified or theory-motivated contrasts.

5 Context: Fertilizer in Malawi

Public fertilizer subsidy in Malawi predates independence and has run in some form for over seventy years, beginning with subsidised smallholder fertilizer supply through the colonial Department of Agriculture in 1952 and the post-independence ADMARC and SFFRFM parastatals (Benson et al., 2024). The modern food-security framing dates to the 1998/99 Starter Pack programme, a universal free-input scheme reaching 2.8 million smallholders that was scaled back under donor pressure to a Targeted Input Programme over 2000–2003 and then consolidated as the Farm Input Subsidy Programme (FISP) in 2005/06, after a maize harvest failure made fertilizer access an electoral issue (Harrigan, 2008; Benson et al., 2024; Lunduka, Ricker-Gilbert, and Fisher, 2013). FISP became politically transformative and was widely credited as a “Malawi miracle” that powered yield gains (Denning et al., 2009), an account that subsequent farm-level evaluations substantially revised (Lunduka, Ricker-Gilbert, and Fisher, 2013; Jayne et al., 2018; Nyondo et al., 2021b). The agronomic blueprint inherited from this period, was embedded in the Affordable Inputs Programme (AIP), which ran from 2020 to 2025), is a blanket recommendation of 50 kg/acre of NPK basal blend at planting and 50 kg/acre of urea for top-dress, uniform across regions and soils despite substantial heterogeneity in soil fertility, rainfall, and household resources.

The fiscal weight of this commitment is large and persistent. Variable-input subsidies absorbed roughly two-thirds of on-budget agrifood spending in recent years (World Bank, 2025), and the AIP alone has run at 30–45 percent of the agriculture budget since 2020 (Benson et al., 2024; Duchoslav and De Weerdt, 2023). Farm-level returns are modest. Critical reviews place the conditional maize response at roughly 2–3 kg per kg of subsidised fertilizer (Lunduka, Ricker-Gilbert, and Fisher, 2013; Nyondo et al., 2021b), and field work in Central Malawi finds the response collapsing toward zero once weeding and soil-carbon thresholds are not met (Burke, Frossard, and Jayne, 2017). Benefit–cost ratios computed for FISP at conditional response rates hover below or near one (Lunduka, Ricker-Gilbert, and Fisher, 2013), and crowding out of commercial fertilizer purchases erodes the net contribution further: roughly 0.18–0.22 kg of commercial fertilizer is displaced per kg of subsidy, with a further share lost to wholesale diversion (Ricker-Gilbert, Jayne, and

Chirwa, 2011; Jayne et al., 2018). Targeting is similarly weak: nationally representative analysis finds that over half of FISP recipients were not consistent with plausible resource-poverty criteria (Kilic, Whitney, and Winters, 2013), and within villages most allocation is done de facto by chiefs, who use local information on shocks and returns but cover only the consumption-poor imperfectly (Basurto, Dupas, and Robinson, 2020; Duchoslav and Kenamu, 2018).

National maize production has trended upward over the FISP/AIP era, but household-level food security has not followed. Only about 17 percent of Malawian households are maize self-sufficient, and among the poorest quintile 85 percent grow maize while only 5 percent are self-sufficient (Benson et al., 2024); rural headcount poverty was essentially unchanged across the previous FISP decade (Nyondo et al., 2021b). The system is also exposed to macroeconomic and fiscal stress. The 2022–24 forex crunch tightened import financing and constrained timely fertilizer procurement, AIP beneficiary numbers were cut from 3.75 million in the early years to 1.5 million in 2023/24, the package was scaled back, and the government moved to sole-importer status (Office of the Ombudsman of Malawi, 2024; Duchoslav and De Weerdt, 2023; World Bank, 2025). The documented governance failures in 2022/23–2023/24 procurement reinforced the case for reform (Office of the Ombudsman of Malawi, 2024).

A converging policy consensus, sharpened during the 2022–24 reform debate, argues that the binding constraints on AIP are no longer purchasing power per se but allocation across heterogeneous farmers and products: targeting on productivity, allowing farmer co-payments to self-select demand, and integrating soil-health and site-specific agronomy into what the subsidy buys (Duchoslav and De Weerdt, 2022; Banda et al., 2022; Nyondo et al., 2021a).

6 Study Sample, Randomization, and Analysis Samples

Sampling frame and design. The sampling frame is inherited from a previous cluster-randomized study on smallholder market participation (De Weerdt et al., 2024a), which delimited 113 villages across four districts (Kasungu, Dowa, Ntchisi, and Mchinji). The number of farmers within each village and the share of villages allocated to each arm were chosen using a series of power simulations. Drawing yield observations from the De Weerdt et al. (2024a) farmer panel (mean 2,574 kg/ha, SD 1,145 kg/ha), we ran 1,000 simulations on a grid varying the control share (10–50 percent of villages) and the number of households per village (10–20), assuming a 13 percent yield increase in each treatment arm relative to control. The selected configuration (roughly 30 percent of villages to control, the remaining 70 percent split approximately equally between T1 and T2, and 18 households per village) yields 85–90 percent power to detect a 13 percent yield effect (about 335 kg/ha) at the 5 percent level on T1-vs-control, with comparable power on T2-vs-control. The T1-vs-T2 contrast is substantially less well powered: the same simulations indicate power closer to 50 percent for a 13 percent gap between T1 and T2, so where we discuss T1-vs-T2 differences below we interpret them with that caveat. The final design enrolled 18 households in each of the 113 villages, for a total of 2,034 households in the baseline sampling frame.

Randomization and timeline. Random allocation was performed at the village level by simple random assignment, without stratification, and assigned 34 villages to control, 40 to T1, and 39 to T2, yielding 612 control, 720 T1, and 702 T2 households at baseline. Baseline data were collected in May–June 2022 as part of the source study; treatments were delivered in the 2024/25 season; endline followed in June 2025.

Baseline balance. We test balance on five pre-registered household characteristics: household-head age, whether the household head is male, household size, land area for crop production, and whether the household had difficulty meeting food needs in the year preceding the survey. Results are in Table 1. The table reports the control-group mean and the difference between each treatment arm and the control group, with standard errors clustered at the village level. No pairwise difference is statistically significant at conventional levels, and joint tests of orthogonality of treatment assignment to the five covariates (run separately for T1-vs-control and T2-vs-control) do not reject the null of orthogonality.[5]

---

[5] The pre-analysis plan also pre-registered roughly ten additional balance variables drawn from the previous agricultural season: prior-season fertilizer use by type, prior-season yield and profit, organic-fertilizer use, number of crops grown, AIP receipt, total cultivated area, and whether the prior-season crop was a legume, to be elicited a few weeks after treatment

Table 1: Balance table

	Control mean	T1 - Control	T2 - Control	N
Household head age (years)	42.434	1.158	1.136	2007
	(14.760)	(0.867)	(0.819)	
Household head is male (1=yes)	0.781	0.012	0.022	2034
	(0.414)	(0.025)	(0.024)	
Household size (number)	4.902	0.129	0.288	2032
	(1.927)	(0.125)	(0.138)	
Land area (ha)	1.002	0.009	0.016	2029
	(0.626)	(0.052)	(0.057)	
Had difficulties feeding family in last year (1=yes)	0.276	0.041	0.040	2034
	(0.447)	(0.037)	(0.034)	
Multinomial test (p-value)	13.649	(0.190)		
F-test C/T1 (p-value)	1.349	(0.241)		
F-test T1/T2 (p-value)	0.466	(0.802)		

Note: First column reports control group means (and standard deviations below); second and third columns show differences between each treatment arm and the control group (and standard errors clustered at the village level below). Land area is measured in hectares and is constructed from reported maize, groundnut, and soybean acreage. Joint tests assess whether baseline covariates predict treatment assignment in the indicated comparison.

Sample realization and soil-test linkage We also document survey realization and the soil-test linkage used for recommendation-based outcomes. Figure 3 and Appendix Table 12 track the sample through four stages. *Baseline sampled* refers to households selected into the study sample from the 2022 source-study frame and is the common denominator for all reported percentages. *Endline target* refers to households appearing in the realized endline survey master file after treatment-administration and tracking efforts identified households that had moved, could not be found, or otherwise could not be retained as endline targets. *Interview completed* refers to households with a usable endline interview after the consent screen. *Matched to soil test data* refers to endline observations successfully linked to the administrative soil-test records used in constructing soil-test-based outcomes. This final stage is relevant for outcomes that require soil-test-linked information, in particular application-error outcomes; it is not the denominator for outcomes measured from the endline survey alone.

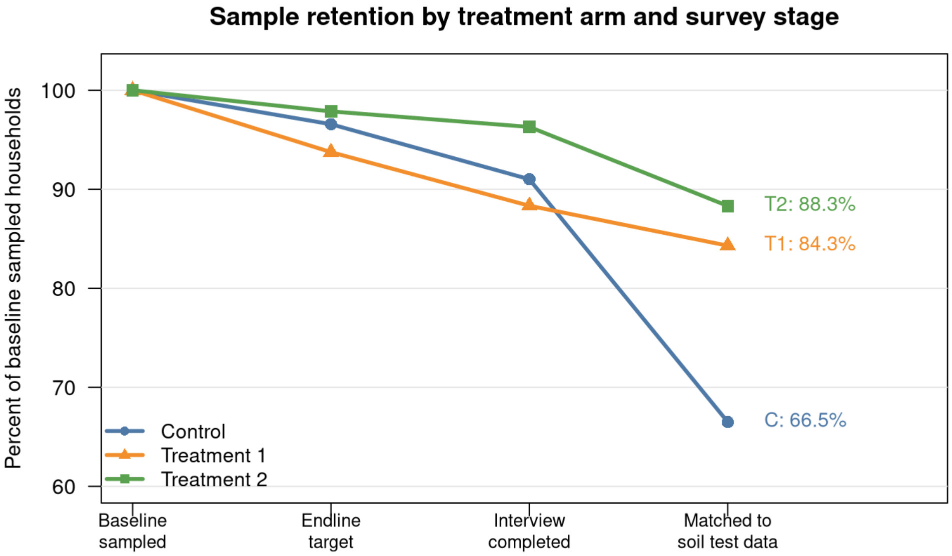
Relative to the original sampled baseline frame, the endline target frame contains 591 control households, 675 T1 households, and 687 T2 households (Appendix Table 12). Completed interviews number 557 in the control group, 636 in T1, and 676 in T2. The gap between endline targets and completed interviews is small in each arm, so differences in realized survey completion primarily arise before the interview stage, through differential retention in the endline target frame, rather than through refusal after contact, for example.

Regression-based attrition tests reported in Appendix Table 13 show no detectable difference between T1 and the control group in realized survey completion, but households assigned to T2 are about 4.2 percentage points more likely than control households to complete the endline interview ($p < 0.001$) and about 4.2 percentage points more likely than T1 households to do so ($p = 0.0016$). One plausible mechanism is that voucher households were easier to recontact or more willing to cooperate because treatment delivery created additional links through the voucher and agro-dealer process. However, retained-sample baseline balance checks in Appendix Table 14 show no statistically significant imbalance, so the differential does not mechanically imply selection on baseline observables.

---

administration. In practice the endline questionnaire scoped retrospective recall narrowly: only prior-season AIP receipt, prior-season cultivated area, and prior-season crop on the soil-test plot were collected. The remaining pre-registered retrospective variables (fertilizer use, yield, profit, organic-fertilizer use) were not elicited for the prior season and cannot be tested. We do not report a separate retrospective balance panel because the available subset is too thin and post-randomization recall on the remaining variables would in any case be vulnerable to treatment-correlated reporting differences.

Figure 3: Sample retention by treatment arm and survey stage



Note: Each stage is expressed as a percentage of the original sampled baseline frame in that treatment arm. This keeps the denominator fixed across stages and makes differences in realized sample retention directly comparable across treatment arms. Baseline sampled denotes households selected into the study sample from the 2022 source-study frame; endline target denotes households in the realized endline master file; interview completed denotes households with a usable endline interview; matched to soil test data denotes endline observations successfully linked to the administrative soil-test records used to construct soil-test-based outcomes.

7 Results

7.1 Uptake and treatment adherence

Before turning to fertilizer use, we first document treatment fidelity. This is necessary for interpreting both the information-only arm and the incremental voucher contrast. A null effect of T1 could mean that advice was not delivered, that advice was delivered but not understood, or that advice was understood but could not be acted upon because of affordability or input-access constraints. These cases have different substantive implications and should not be conflated.

Table 2 follows the implementation sequence. Agronaut visits are measured as a binary indicator for whether the respondent reported that an Agronaut, defined in the questionnaire as an extension worker from a Farmers World agro-input shop, visited the household to provide further extension services on crops grown in the 2024/25 season. Agronaut visits occurred for 17.4 percent of control households, 40.3 percent of T1 households, and 39.0 percent of T2 households, with no detectable difference between T1 and T2. Recommendation receipt is also self-reported. After being shown an example recommendation sheet on the tablet computer, respondents were asked whether they received a similar document with fertilizer recommendations based on the soil test for their plot. Reported receipt was near universal: 96.7 percent of T1 households and 98.5 percent of T2 households report receipt. The difference in recommendation receipt across the two treatment arms is small in absolute terms, about 1.8 percentage points, and only marginally significant at conventional levels.

Larger differences between T1 and T2 appear after reported receipt. Conditional on receiving a recommendation, 80.7 percent of T1 households and 89.9 percent of T2 households report that it was easy or very easy to understand. Both arms received the same advice from the same agronaut, so the gap cannot reflect the recommendation itself becoming more comprehensible. It is consistent with having an executable action attached to the advice, which can raise self-reported ease through engagement at the point of purchase, retrospective justification of the action taken, or actionability bleeding into the comprehension question; the downstream compliance results reported below rule out a genuine comprehension channel. Conditional on recommendation receipt, 60.8 percent of T1 households and 92.6 percent of T2 households report following the recommendation, a pattern more in line with expectations and consistent with liquidity constraints limiting follow-through in the recommendation-only arm. These high self-reported adherence rates should be interpreted with caution. As shown in the compliance analysis below, observed fertilizer choices indicate substantially lower compliance with the specific products recommended by the soil tests.

Implementation of the voucher treatment itself was also nearly complete in T2. Among interviewed T2 households, 97.5 percent report voucher receipt. Conditional on voucher receipt, 98.6 percent report redemption. At the same time, conditional on redemption, only 14.2 percent report that the voucher value was enough to buy the full recommended bundle.

These first-stage results rule out a simple non-receipt explanation for any weak T1 effects in the fertilizer-use tables below. The shared information component was reportedly received at very high rates in both treatment arms. What differs sharply is the extent to which farmers report acting on the recommendation, which is consistent with affordability remaining a binding margin even when advice reaches farmers. We therefore interpret the T1 versus T2 contrast primarily as a difference in the ability to implement the recommendation rather than a difference in basic treatment receipt.

Table 2: Treatment fidelity and first-stage means

Outcome	Control	T1	T2	p-value: T1 = T2	N
Panel A: Delivery among interviewed households					
Agronaut visited household	0.174	0.403	0.390	0.780	1868
Received recommendation	0.002	0.967	0.985	0.096	1868
Received voucher	0.000	0.000	0.975	0.000	1868
Panel B: Recommendation comprehension and adherence, conditional on receipt					
Found recommendation easy/very easy		0.807	0.899	0.000	1280
Followed recommendation		0.608	0.926	0.000	1280
Panel C: Voucher implementation, T2 only					
Redeemed voucher, conditional on receipt			0.986		658
Voucher enough for full recommended purchase, conditional on redemption			0.142		649

Notes: The sample is households with a completed endline interview. Control, T1, and T2 columns report arm-specific means. The p-value column tests equality of the T1 and T2 means using a linear probability model with standard errors clustered at the village level. The Agronaut-visit row is an indicator for at least one reported visit by a Farmers World extension worker providing further extension services on crops grown in the 2024/25 season. The recommendation-receipt row is based on the respondent's report of receiving the fertilizer recommendations based on the soil test for their plot after being shown an example recommendation sheet on the tablet. Panel B is restricted to households that received a recommendation. Panel C is restricted to T2 households, with row-specific denominators: voucher redemption is conditional on voucher receipt, and voucher sufficiency is conditional on voucher redemption.

7.2 Fertilizer and nutrient use

We begin with total fertilizer use on the soil-test plot. The endline survey asked each household which types of fertilizer they applied to the soil-test plot during the 2024/25 season and, for each type selected, the quantity in kilograms. Households who indicated they never applied any fertilizer were coded as zero. Total fertilizer applied simply sums quantities across fertilizer types recorded in the instrument; kilograms per acre divides by self-reported plot size in acres. Table 3 reports ITT estimates for three outcomes. Panel A is a binary indicator for any fertilizer use. Panel B covers total kilograms applied; Panel C normalises by plot area. Within each panel, the first row uses all soil-test plots and the second restricts to households whose soil-test plot's main crop was reported to be maize.

Table 3: Effect of the Program on Total Fertilizer Applied

Sample	Control mean	T1 − Control	T2 − Control	p-value: T2 = T1	N
Panel A: Any fertilizer used (proportion)					
All soil-test plots	0.70	0.03	0.25***	0.000	1950
		(0.03)	(0.03)		
Maize soil-test plots	0.82	0.02	0.17***	0.000	1625
		(0.03)	(0.02)		
Panel B: Total kg applied					
All soil-test plots	78.46	2.24	30.52***	0.000	1950
		(7.29)	(6.40)		
Maize soil-test plots	83.36	3.37	28.62***	0.000	1625
		(6.87)	(6.01)		
Panel C: Kg per acre					
All soil-test plots	52.61	2.20	31.63***	0.000	1866
		(4.83)	(3.19)		
Maize soil-test plots	54.52	2.32	31.30***	0.000	1625
		(4.98)	(3.60)		

Notes: This table reports unadjusted ITT effects on fertilizer use on the soil-test plot. Panel A is a binary indicator for any fertilizer use (linear probability model). Panel B reports total kilograms applied; Panel C reports kilograms per acre. No panel sums fertilizer across the soil-test and randomly selected non-test plots. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher redeemable for any fertilizer product at the nearest Farmers World shop. The column labelled *p-value: T2 = T1* reports the test of equality between the two treatment effects. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

T1 does not generate a detectable change in any fertilizer outcome. Panel A shows that T2 raises the probability of using any fertilizer by 26 percentage points across all soil-test plots and by 17 percentage points on maize plots, both highly significant. In other words, essentially every voucher household applied fertilizer to the maize plot designated by the soil test. This is consistent with the 98.6 percent voucher-redemption rate reported above and confirms that the voucher arm reaches a near-ceiling on the use margin. T2 increases total fertilizer applied by 30.5 kg across all soil-test plots and by 28.6 kg among maize soil-test plots (Panel B), both significant at the 1% level. Panel C shows per-acre effects of 31.6 kg/acre and 31.3 kg/acre respectively, slightly larger than the total-kg estimates, consistent with T2 plots being somewhat smaller on average. Equality tests reject T1 = T2 in both samples across all panels. Site-specific recommendations alone do not measurably change fertilizer quantities. The absence of a response to recommendations alone is particularly notable given the high reported rates of recommendation receipt and understanding. This suggests that information constraints were not the primary barrier to adoption in this setting.

The null effect of T1 on fertilizer use is not simply an absence of response — it is an informative finding that sharpens the boundary conditions of the information-only literature. Existing evidence from Ethiopia (Ayalew, Chamberlin, and Newman, 2022) demonstrates positive fertilizer responses to plot-specific recommendations in a setting where the recommended inputs were already familiar to farmers and commercially available at the point of recommendation delivery. The contrast with the present null result points to a critical moderating condition: the agronomic signal contained in a soil-test recommendation generates behavioral change only when it aligns with the existing product ecology — the set of fertilizers farmers recognize, trust, and can obtain. A plausible reading is that in Malawi, where AIP has for two decades anchored farmer expectations around urea and NPK 23:10:5, recommendations that prescribe CAN, potassium products, or lime sit outside the institutional infrastructure through which fertilizer knowledge and access are jointly organized, and that this misalignment, rather than the agronomic content of the advice, drives the null. We cannot identify this channel directly from the experiment, but the pattern is harder to reconcile with a pure information-content explanation.

Table 4 decomposes this increment by product, reporting cell means for each arm. The additional fertilizer in the voucher arm consists almost entirely of urea and NPK 23:10:5: T2 households apply 17.8 kg more urea and 18.7 kg more NPK 23:10:5 on the soil-test plot than control households. Together these two increments account for nearly all of the 29–31 kg total-kg increase in Panel B of Table 3, in a roughly 1:1 split that mirrors the 50/50 composition of the AIP bundle (one bag urea topdressing, one bag NPK basal blend). Despite the voucher's full fungibility across products and package sizes, T2 households appear to simply reproduce the AIP product mix. Despite complete flexibility in voucher use, farmers overwhelmingly selected the same fertilizer products promoted under the national subsidy program, indicating strong persistence in fertilizer-use patterns and preferences.

Table 4: Descriptive product composition of fertilizer applied on the soil-test plot

Product	Control	T1	T2	T2 − Control
Panel A: AIP-style products				
Urea	34.6	34.0	52.4	+17.8
NPK 23:10:5	30.5	33.8	49.2	+18.7
Panel B: Other common products				
NPK 15:23:16	0.8	1.1	2.8	+2.1
CAN	2.5	2.9	1.3	−1.2
DAP	1.9	0.2	0.07	−1.9
Panel C: Site-specific or low-adoption products				
NPK 8:18:15	2.8	4.5	2.7	−0.1
Potassium product	1.5	1.2	0.0	−1.5
NPK 14:14:20	0.0	0.4	0.3	+0.3
Lime	0.0	0.0	0.0	0.0
MAP	0.0	0.0	0.0	0.0
N households	591	675	687	—

Notes: Mean kilograms applied per household on the soil-test plot by treatment arm. This table is descriptive: it reports unadjusted arm means and the raw T2–control difference by product. Adjusted ITT estimates for total fertilizer use and nutrient application are reported in Tables 3 and 5. "Potassium product" combines MOP, SOP, and potassium sulphate. "Lime" combines calcitic and dolomitic lime. DAP denotes diammonium phosphate.

Table 5 disaggregates the maize-plot fertilizer response by nutrient type. This is important because kilograms of fertilizer do not map one to one into agronomic intensity when farmers use different blends. T2 increases nitrogen application by about 25.7 kg/ha. It also raises phosphorus and potassium application, generating an increase of roughly 31.8 kg/ha in total nutrients. T1 again shows no detectable effect. The equality tests confirm that T2 differs from T1 for each nutrient outcome. The fertilizer response in the voucher arm is therefore not only an increase in product weight; it also implies higher application of the main macronutrients relevant for maize production.

The asymmetric pattern across nutrients, a large nitrogen response paired with modest phosphorus and near-zero potassium gains under T2, needs explicit interpretation. The nitrogen increase mechanically reflects the composition of the voucher-induced increment: as Table 3 documents, the additional kilograms are overwhelmingly urea (46% N) and NPK 23:10:5 (23% N), both high-nitrogen products. The phosphorus gain is a by-product of NPK 23:10:5 application rather than a deliberate response to soil phosphorus deficiency. And the near-zero potassium response, in a sample where soil-test results indicate widespread K deficiency, reflects the complete absence of potassium product purchases, not a considered agronomic judgment. This decomposition matters for the yield results that follow. The T2 yield gain is generated by nitrogen intensification of an already nitrogen-centric bundle, not by correction of the nutrient imbalances the recommendation was designed to address. To the extent that potassium and phosphorus are co-limiting with nitrogen, as the soil-test evidence suggests, the observed yield gain represents only a fraction of the agronomic potential that full-compliance fertilization would unlock. Consequently, the estimated treatment effects should be interpreted as the returns to additional conventional fertilizer use rather than the returns to precision nutrient management.

Table 5: Effect of the Program on Nutrient Application on Maize Plots

Outcome	Control mean	T1 − Control	T2 − Control	p-value: T2 = T1	N
Nitrogen	45.50	1.58	25.72***	0.000	1625
		(4.88)	(2.93)		
Phosphorus	6.18	0.33	4.03***	0.000	1625
		(0.49)	(0.44)		
Potassium	3.11	0.19	2.06***	0.000	1625
		(0.24)	(0.22)		
Total nutrients	54.78	2.10	31.81***	0.000	1625
		(5.24)	(3.54)		

Notes: This table reports unadjusted ITT effects on nutrient application in kg/ha among households whose main crop is maize. Nutrient quantities are derived by multiplying each product's applied weight by its label-grade fractions: urea (46% N); NPK 23:10:5 (23% N, 10% P_2O_5, 5% K_2O); CAN (26% N); NPK 8:18:15 (8% N, 18% P_2O_5, 15% K_2O). Phosphorus and potassium are expressed as P_2O_5 and K_2O. Nutrient totals are measured on the soil-test plot and divided by plot area after converting acres to hectares. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher redeemable for any fertilizer product at the nearest Farmers World shop. The column labelled *p-value: T2 = T1* reports the test of equality between the two treatment effects. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

A natural concern is that the 29–31 kg test-plot increase in Panel B of Table 3 reflects reallocation from other household plots rather than genuinely additional fertilizer use. The endline survey addresses this directly. The survey recorded identical information for a randomly-selected non-test plot: which fertilizer types were applied and the quantity of each in kilograms, together with the plot's area in acres. Households who reported never applying any fertilizer to the random plot (883 respondents) were coded as zero rather than excluded. The analysis therefore covers all 1,302 households with at least one non-test plot.[6] Of these, only 144 have a randomly-selected non-test plot that is itself maize, confirming that the majority of non-test plots are secondary crops that receive little or no fertilizer.

Table 6 reports the ITT estimates. Neither arm shows a detectable effect on any outcome. Neither T1 nor T2 changes the probability of using any fertilizer on the random plot (control mean: 0.33; T2: −0.02, $p = 0.65$), ruling out even an extensive-margin spillover. The T2 estimate for total non-test household fertilizer is −18.9 kg (SE = 16.7, $p = 0.26$) and the T1 estimate is −21.2 kg (SE = 15.6, $p = 0.18$). The kg-per-acre outcome is also null; the equality p-value of 0.99 provides no evidence of differential reallocation between T1 and T2. Total household fertilizer use therefore rises by approximately the test-plot increment, with no offsetting reduction or spillover gain on other plots in the household's portfolio.

---

[6] A potential concern with the single randomly-selected plot analysis is that it may mask displacement concentrated among households with many non-test plots: if a household reallocates fertilizer away from, say, four other plots, the effect per plot is diluted and may fall below detectable thresholds on any one of them. To account for this, we apply a Horvitz–Thompson correction, multiplying each household's random-plot fertilizer use by its total number of non-test plots. Under random plot selection within each household, this product is an unbiased estimator of total non-test fertilizer use. In the analysis sample, 53.5% of households have exactly one non-test plot (scaling factor of one), 34.8% have two, and the remaining 11.7% have three to five.

Table 6: Test for fertilizer reallocation to non-test plots

Outcome	Control mean	T1 − Control	T2 − Control	p-value: T2 = T1	N
Any fertilizer used (proportion)	0.33	0.00	-0.02	0.653	1301
		(0.05)	(0.05)		
Total kg applied on non-test plots (HT-scaled)	78.1	-21.2	-18.8	0.864	1301
		(15.6)	(16.7)		
Kg per acre on non-test plot (HT-scaled)	52.8	-3.6	-3.4	0.987	1300
		(10.2)	(9.9)		

Notes: Unadjusted ITT estimates for fertilizer use on the randomly-selected non-test plot. The first row is a binary indicator for any fertilizer use (linear probability model). The second and third rows report Horvitz–Thompson (HT)-scaled outcomes: each household's random-plot fertilizer use is multiplied by its total number of non-test plots, giving an unbiased estimator of total non-test household fertilizer under random plot selection. Households that reported never using fertilizer on the random plot are included with a value of zero. Sample is restricted to households with at least one non-test plot ($N = 1{,}302$). Standard errors are clustered at the village and reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

7.3 Recommendation compliance

Three facts from the preceding tables set up the compliance question. First, T2 farmers self-report higher adherence to the recommendation (Table 2). Second, they apply roughly 30 kg more fertilizer on the soil-test plot (Table 3). Third, Table 4 shows that this increase is concentrated almost entirely in urea and NPK 23:10:5, the two products that also constitute the AIP bundle, while products prescribed by the recommendation but outside the AIP bundle (calcium ammonium nitrate (CAN) as topdress, potassium products, lime, mono-ammonium phosphate (MAP)) are not purchased in greater quantities. The question is therefore whether the voucher arm's fertilizer increase reflects genuine recommendation compliance, or whether farmers redeemed the voucher for the familiar AIP bundle regardless of what the recommendation prescribed.

Unlike Table 2, which reports arm-specific first-stage means, Tables 7 and 8 report the T2–T1 contrast among treated households with valid recommendation records. This is the relevant comparison for recommendation compliance because the control group did not receive the soil-test recommendation during the study period. Table 7 restricts attention to T1 and T2 households that report receiving a recommendation, have a valid treatment recommendation record, and have at least one positive parsed product recommendation.

The recommendation data identify up to eight fertilizer products or product groups: NPK 23:10:5, urea, CAN, NPK 14:14:20, NPK 8:18:15, potassium products, lime, and MAP. Panel A summarizes product-set overlap between these recommended products and the products farmers report applying on the plot where the soil was tested. The outcome "any recommended product applied" equals one if the farmer applied at least one of the products in the recommendation. The "share of recommended products applied" is the number of recommended products applied divided by the number of products recommended. The exact-bundle outcome is the strictest product-set measure: it equals one only when the farmer applies every recommended product and no additional non-recommended product from the parsed product set, though it still does not require applied quantities to match recommended quantities.

We find that, in T1, 57.1 percent of households apply at least one recommended product; the voucher raises this to 73.7 percent in T2, an increase of 16.6 percentage points. The average share of recommended products applied rises from 34.8 percent in T1 to 43.4 percent in T2. The gain is concentrated in the familiar products that farmers could plausibly obtain through the voucher-supported package: application of recommended NPK 23:10:5 rises from 64.0 percent in T1 to 90.1 percent in T2. Recommended urea is also applied at high rates in T2, rising from 66.7 percent to 94.0 percent, but this row is defined for only 95 households because urea is rarely the product prescribed by the soil-test recommendation. By contrast, potassium products are recommended for 900 households in this sample, often in place of the more familiar urea topdress, yet potassium compliance is essentially zero and is lower in T2 than in T1.

The table also shows a hard limit of the intervention. T2 does not increase the probability of applying all recommended products, and exact product-bundle compliance is not only near zero in both arms but is significantly lower in T2 than in T1. This pattern is consistent with several mechanisms that the design cannot separate. A habit-formation account, in which two decades of AIP participation have anchored the modal fertilizer bundle as the reference point for input decisions, would generate exactly this observation: increasing purchasing power without changing the reference bundle raises the probability that farmers execute the familiar plan at larger scale, mechanically increasing the applied share of familiar products while reducing exact-bundle compliance on the prescription's non-standard components. But the same observation is generated by supply absence (potassium, lime, MAP, or CAN may simply not be stocked at the redeeming Farmers World shop), by a budget constraint that pushes farmers to maximise usable kilograms within a fixed voucher value (only 14 percent of redeeming households reported the voucher value as sufficient for the full recommended bundle), by divisibility and timing mismatches between voucher redemption and the application windows for lime and potassium, and by a mechanical penalty in the exact-bundle measure, which falls whenever T2 households add a non-recommended product alongside the recommended ones.

Conditioning on each T2 household's product-level recommendation does not change this picture. Among T2 households that applied urea or NPK 23:10:5 on the test plot and have a valid product-level recommendation, the applied urea share, urea/(urea+NPK 23:10:5) in kg, has median 0.50 both for the 427 households whose recommendation does not list urea and for the 47 households whose recommendation does list urea; the difference in means is 1.1 percentage points (cluster-robust standard error 2.8, $p = 0.68$). Within the subset of households for whom both products are recommended, regressing the applied urea share on the recommended urea share returns a slope of -0.19 (cluster-robust standard error 0.27, $p = 0.49$). The voucher arm applies the FISP product ratio irrespective of whether the site-specific recommendation prescribes urea, NPK 23:10:5 only, or both in a non-50/50 ratio.

Even when farmers did not fully match the recommended product bundle, farmers may still move closer to the recommended nutrient totals. This is explored in Table 8. We again restrict this analysis to T1 and T2 households with valid treatment recommendation records. The table reports recommendation amounts, actual application, absolute application error, and nutrient shortfall for N, P_2O_5, and K_2O. As expected, recommended nutrient quantities are similar across treatment arms. The voucher arm applies substantially more nutrients: actual N application is about 24.3 kg/ha higher than in the recommendation-only arm, while actual P_2O_5 and K_2O application are also higher, though by smaller amounts. This narrows the N gap: absolute N application error falls by about 8.6 kg/ha and N shortfall falls by about 15.7 kg/ha in T2 relative to T1. The shortfall falls by more than the absolute error because the two measures treat over-application differently: shortfall is zero once a farmer exceeds the recommendation, whereas absolute error continues to penalise excess application. The gap between the two reductions implies that a portion of T2 farmers cross from under-applying to over-applying nitrogen, eliminating their shortfall but generating a new absolute error on the upside. The same pattern does not appear for P_2O_5 or K_2O. For these nutrients, recommendation levels remain far above actual application in both treatment arms, so the voucher-induced change in fertilizer use mainly improves compliance with the nitrogen component of the recommendation. It is important to register the level of residual non-compliance explicitly: even in T2, mean nitrogen application remains roughly 11 kg/ha below the recommended rate, with phosphorus and potassium shortfalls of similar or larger magnitude. The voucher therefore narrows but does not close the application-error gap, and the headline positive T2 result coexists with a substantial residual shortfall relative to the agronomic recommendation.

The compliance evidence permits a partial decomposition of non-adoption across three product classes defined by their supply and demand characteristics. First, products that are both recommended and familiar — urea and NPK 23:10:5 — show high compliance rates in both arms, with the voucher amplifying uptake along the intensive margin. This is the baseline: when recommendation and product familiarity coincide, compliance is achievable. Second, CAN as a topdress substitute for urea is functionally equivalent to urea as a nitrogen source and falls within the voucher's value, yet compliance with CAN is lower in T2 than in T1. To the extent that CAN was stocked at the redeeming Farmers World shop, a possibility we cannot verify with the data at hand, this cell is the most diagnostic in the compliance table: an affordable, agronomically equivalent product would be rejected in favor of the familiar one when purchasing power is relaxed, pointing to a demand failure rooted in product familiarity, branding, or risk aversion over untested substitutes rather than to a supply constraint. If CAN was not consistently stocked, the same observation is generated by supply absence, which we cannot rule out. Third, potassium products, lime, and MAP show zero compliance in every arm, consistent with either supply absence or demand failure — without agro-dealer stocking data these cases cannot be resolved, but they represent distinct policy problems that demand-side instruments alone cannot address. A reform that treats all three classes as a single compliance failure will misdiagnose the binding constraint.

Taken together, Tables 7 and 8 distinguish partial implementation from full recommendation compliance. The voucher raises the probability that farmers apply at least one recommended product and reduces nitrogen shortfalls, but it does not produce exact product-bundle compliance or close the phosphorus and potassium gaps; on the strictest compliance measure, the voucher arm falls significantly below the soil-test-only arm.

Table 7: Voucher effects on product-level recommendation compliance

Outcome	T1 mean	T2 − T1	N
Panel A: Product-set compliance summary			
Applied at least one recommended product	0.572	0.164^{***}	1161
		(0.037)	
Share of recommended products applied	0.350	0.086^{***}	1161
		(0.031)	
Applied all recommended products	0.135	0.013	1161
		(0.041)	
Applied exact recommended product set	0.038	-0.021^{**}	1161
		(0.010)	
Panel B: Product-specific compliance, conditional on product recommendation			
Recommended NPK 23:10:5 applied	0.642	0.258^{***}	937
		(0.032)	
Recommended urea applied	0.667	0.272^{***}	94
		(0.088)	
Recommended CAN applied	0.250	-0.145^{*}	97
		(0.081)	
Recommended NPK 14:14:20 applied	0.047	-0.047	108
		(0.032)	
Recommended NPK 8:18:15 applied	0.018	0.000	115
		(0.023)	
Recommended potassium product applied	0.021	-0.021^{**}	910
		(0.010)	
Recommended lime applied	0.000	0.000	143
		(0.000)	
Recommended MAP applied	0.000	0.000	51
		(0.000)	

Notes: The sample is T1 and T2 households with valid treatment recommendation records, at least one positive parsed product recommendation, self-reported recommendation receipt, and usable fertilizer-application data. The table reports T1 means and unadjusted estimates of T2 minus T1. Product recommendations are parsed from product-specific recommendation rates for eight products or product groups: NPK 23:10:5, urea, CAN, NPK 14:14:20, NPK 8:18:15, potassium products, lime, and MAP. Actual product use is measured from the endline product-list questions for the test plot. All Panel A outcomes compare the set of recommended products with the set of products applied, ignoring quantities. The share outcome is the fraction of recommended products that were applied at any positive amount. The all-products outcome equals one if all recommended products were applied, regardless of quantities or additional products. The exact-bundle outcome equals one only if the applied product set matches the recommended product set within the eight parsed products. Potassium products combine muriate of potash (MOP), sulphate of potash (SOP), and potassium sulphate; lime combines calcitic and dolomitic lime. Product-specific rows are defined among households for whom that product or product group was recommended. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

Table 8: Recommendation compliance and application error

Outcome	T1 mean	T2 − T1	N
Panel A: Nitrogen (kg/ha)			
Recommended amount	59.08	1.64	1180
		(2.50)	
Actual application	40.49	27.67***	1180
		(2.51)	
Absolute application error	37.88	-8.55***	1180
		(1.65)	
Shortfall	28.23	-17.29***	1180
		(1.98)	
Panel B: Phosphorus, P_2O_5 (kg/ha)			
Recommended amount	36.42	2.99	1180
		(4.57)	
Actual application	6.57	3.68***	1180
		(0.48)	
Absolute application error	30.30	-0.98	1180
		(4.44)	
Shortfall	30.07	-0.84	1180
		(4.44)	
Panel C: Potassium, K_2O (kg/ha)			
Recommended amount	51.95	7.05	1180
		(5.92)	
Actual application	3.61	1.83***	1180
		(0.34)	
Absolute application error	48.56	5.04	1180
		(5.80)	
Shortfall	48.45	5.13	1180
		(5.80)	

Notes: The sample is T1 and T2 households with valid treatment recommendation records, non-sentinel plot area, and usable fertilizer-application data. The table reports unadjusted estimates of T2 minus T1. Recommended nutrients are computed from the product-specific recommendation rates. Actual P and K application are measured in fertilizer-grade P_2O_5 and K_2O units, matching the cleaned fertilizer-use data. Actual application rates divide plot-level nutrient totals by plot area after converting acres to hectares. Absolute application error is the absolute difference between actual and recommended kg/ha; shortfall is the positive part of recommended minus actual kg/ha. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

7.4 AIP substitution

The fertilizer-use results in Panel B of Table 3 show that T2 households apply about 29–31 kg more fertilizer than control households. A natural concern is whether this gross increase translates into net additional fertilizer at the household level, or whether it is partly offset by reduced AIP receipt: if voucher-arm households are less likely to receive AIP, the true additionality is smaller than the 29–31 kg figure suggests.

The AIP distributes a single in-kind bundle of approximately 100 kg (typically one 50 kg bag of NPK basal blend and one 50 kg bag of urea topdressing) to each beneficiary household at heavily subsidised prices. The programme is large but rationed: in our control group, 19 percent of households report receiving AIP fertilizer in the study season.

A treatment-induced reduction in AIP fertilizer observed on the soil-test plot could mean two very different things. The first is *genuine household-level substitution*: voucher-arm households either no longer receive AIP at all, or receive less of it. The second is *within-household reallocation*: voucher-arm households still receive their full 100 kg AIP bundle but redirect it away from the soil-test plot toward their other plots, because voucher-supplied fertilizer now covers the soil-test plot. Total household AIP would then be unchanged, and any apparent reduction on the test plot would be a measurement artefact of plot selection rather than a substantive change in subsidy receipt.

The challenge is sharpened by how the soil-test plot was selected. At baseline each household nominated, with the field team's prompting, their primary or most important plot; households could nominate any plot, but most selected their main plot, which in the majority of cases was a maize field. This is the plot the soil sample was taken from, the plot the recommendation was written for, and the plot the endline survey collected detailed fertilizer-use data for. Households also typically apply AIP fertilizer preferentially to this same primary plot, because maize is the crop AIP is designed for and the primary plot is where the household seeks the largest return from fertilizer use. The soil-test plot is therefore a high-AIP plot relative to the household's other plots in the control group: control households apply on average 9.6 kg of AIP fertilizer on the soil-test plot but only 1.9 kg on the randomly-selected non-test plot (Table 9).

Table 9 uses the data collected on the randomly selected plot to separate substitution from reallocation through four outcomes. AIP receipt is a binary indicator for whether the household received any AIP fertilizer at all; this is the household-level extensive margin, unaffected by which plot the AIP eventually reached. AIP on the test plot is the kilograms applied on the soil-test plot, the result typically reported in the plot-level subsidy displacement literature. AIP on the random plot is the kilograms applied on the randomly selected non-test plot, restricted to households that cultivated at least one additional plot. The two-plot total sums the test-plot and random-plot AIP; households with no other plot contribute zero.

The pattern is consistent across outcomes. T2 reduces the probability that a household receives AIP at all by 7.7 percentage points, a relative reduction of about 40 percent against the 19 percent control receipt rate. T2 also reduces AIP fertilizer applied on the soil-test plot by 7.0 kg, representing about 73 percent of the 9.6 kg control mean. The random-plot outcome addresses the reallocation hypothesis: if voucher-arm households were shifting AIP fertilizer away from the soil-test plot toward other plots, the random-plot AIP regression should show a positive T2 effect. It does not. The T2 estimate is +0.21 kg (SE = 1.09), statistically indistinguishable from zero. The data do not support the within-household reallocation interpretation. The two-plot total corroborates this: it falls by 6.7 kg under T2, very close to the 7.0 kg test-plot reduction, and the gap of under one kilogram is consistent with the null on the random-plot outcome. T1 is statistically indistinguishable from zero on every outcome, indicating that the AIP-receipt response is specific to the voucher arm and is not produced by the recommendation alone.

The extensive-margin reduction in AIP receipt, combined with the absence of detectable offsetting increases on random plots, points toward partial substitution rather than simple within-household reallocation across observed plots. Voucher-arm households are less likely to receive AIP at all, and the reduction in AIP fertilizer on the soil-test plot is not offset by higher AIP use on the randomly selected non-test plot. Two mechanisms are consistent with this pattern. The first is *demand-side substitution*: voucher-arm households, having already secured fertilizer through the voucher, may be less likely to pursue or pay the AIP co-payment for an additional subsidized package. The second is *supply-side re-targeting*: AIP committees or village leaders may deprioritize households known to have received voucher support (Basurto, Dupas, and Robinson, 2020). The data do not fully separate these channels, although the absence of comparable effects in T1 suggests that the response is tied to voucher receipt rather than to recommendation delivery alone.

Table 9: AIP substitution: effects on receipt and on AIP fertilizer applied at the plot and household level

Outcome	Control mean	T1 − Control	T2 − Control	p-value: T2 = T1	N
AIP receipt (0/1)	0.187	0.013 (0.028)	-0.079*** (0.025)	0.001	1868
AIP on test plot (kg)	9.59	-1.10 (1.49)	-7.13*** (1.25)	0.000	1952
AIP on random plot (kg)	1.90	0.89 (0.97)	0.21 (1.09)	0.521	1304
Two-plot AIP total (kg)	10.82	-0.69 (1.60)	-6.79*** (1.43)	0.000	1952

Notes: Unadjusted ITT effects on four AIP outcomes, all-crops sample. AIP receipt is a linear probability model for whether the household received any AIP fertilizer. AIP on the test plot and AIP on the random plot are kilograms applied on the respective plots; the random-plot sample is restricted to households that cultivated at least one additional plot. The two-plot total sums the two plot-level outcomes without scaling by plot count; households with no other plot contribute zero. AIP allocates a single subsidised bundle of approximately 100 kg per beneficiary household (one bag of NPK and one bag of urea); plot-level reports above 100 kg are recall errors and are capped. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher. The column labelled *p-value: T2 = T1* reports the test of equality. Standard errors clustered at the village are in parentheses. Asterisks: 1% (***), 5% (**), 10% (*).

Appendix Table 15 reports effects on broader soil nutrient management practices. T1 is broadly null. T2 significantly reduces use of fresh vegetative material, dairy or poultry manure, and Mbeya fertilizer (a hybrid compost that blends locally available organic material with a small quantity of chemical fertilizer) by 7–10 percentage points, while increasing adoption of ridges or check dams (+9.0 pp). The reductions in organic inputs are consistent with rational substitution: collecting and applying organic material is labour-intensive, and when mineral fertilizer becomes accessible at low voucher cost, the marginal return to that effort falls. The ridges-and-check-dams increase points in the opposite direction and is more naturally read as a complementary investment to protect the mineral fertilizer applied.

7.5 Maize production and yield

The endline survey recorded farmer-reported total kilograms of maize harvested from the soil-test plot in the 2024/25 season. Yield per acre divides this self-reported harvest total by the maize plot area in acres, adjusted for intercropped area shares. Table 10 presents ITT estimates for both outcomes among households whose soil-test plot's main crop is maize. Panel A uses winsorized total kilograms harvested (1st–99th percentile) as the outcome; Panel B uses log yield in kilograms per acre.

Panel A shows that T2 raises total maize harvested by 151 kg (SE = 72, $p = 0.038$), an increase of 19 percent on a control mean of 779 kg. T1 has no detectable effect (−12 kg, $p = 0.87$). Panel B converts to log yield per acre. T1 again shows no detectable effect. T2 increases log maize yield by 0.36 log points, equivalent to approximately 44 percent.[7] The equality test rejects T1 = T2 in both panels. The evidence supports the causal chain: the voucher arm raised fertilizer application on maize plots, and this input response translated into substantially higher total maize production. To situate the magnitude of this response, consider the implied maize-to-fertilizer ratio. T2 applies approximately 29–31 additional kg of fertilizer and produces approximately 151 additional kg of maize, implying a gross response of approximately 4.9–5.2 kg maize per kg fertilizer. This exceeds the 2–3 kg/kg range that critical reviews of FISP report

---

[7] Sixteen households with zero reported maize yield (3 in control, 8 in T1, 5 in T2) are dropped from the log specification. Robustness checks in Appendix Table 17 confirm the positive T2 effect across alternative sample restrictions. The per-acre percentage (44 percent) substantially exceeds the levels percentage (19 percent) because T2 plots are smaller on average, so the same total production is divided by a smaller area; the 19 percent figure from Panel A is the more conservative and plot-size-neutral estimate of the production gain. Furthermore, if T2 plots are systematically smaller — consistent with the per-acre estimate substantially exceeding the levels estimate — the per-acre yield gain may partly reflect intensive cultivation of high-productivity micro-plots rather than a general improvement in fertilizer productivity across the maize portfolio. A robustness check restricted to plots above median size would verify that the T2 yield result is not driven by the smallest plots in the distribution.

Table 10: Effect of the Program on Maize Production and Yield

Ctrl. mean	T1 − C	T2 − C	T2 % effect	p: T2 = T1	N
Panel A: Total kg harvested					
779	-11.60	150.83**	19.4	0.035	1609
	(68)	(72)			
Panel B: Log yield (kg/acre)					
623.07	-0.13	0.36***	43.8	0.000	1609
	(0.11)	(0.10)			

Notes: Unadjusted ITT effects among households whose soil-test plot's main crop is maize. Panel A outcome is total kilograms of maize harvested, winsorized at the 1st and 99th percentiles; the control-group mean is in kilograms. Panel B outcome is log maize yield in kg/acre; the control-group mean is in kg/acre. The T2 % effect in Panel A is the T2 coefficient as a percentage of the control mean; in Panel B it is $\exp(\hat{\beta}_{T2}) - 1$. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher. The column labelled *p: T2 = T1* reports the test of equality. Standard errors clustered at the cluster level are in parentheses. Asterisks: 1% (***), 5% (**), 10% (*).

as the conditional maize response (Lunduka, Ricker-Gilbert, and Fisher, 2013; Nyondo et al., 2021b). The most likely explanation is that the test plot is farmer-selected and likely among the household's highest-productivity plots, so the plot-level response will tend to exceed a sample-average response measured on randomly assigned plots. This does not imply that the intervention outperforms AIP on a like-for-like basis; it indicates that the favorable comparison reflects a selected plot rather than a generalizable improvement in fertilizer use efficiency.

7.6 Returns and profits to farmers

The yield results do not by themselves establish whether the intervention improved farmers' economic returns. Higher fertilizer use mechanically raises costs, and the relevant question is whether the additional production value is large enough to offset those costs. The endline data do not measure household-level revenue across all crops. We therefore construct economic outcomes for the sampled plot's main crop. In the pooled panel, we include all sampled main crops and value production using crop-specific median sale prices observed among sellers of the same crop. This avoids conditioning the outcome on whether a household sold the crop, but it also means that the value measure is an imputed production value rather than realized household crop revenue. In the maize panel, we restrict to households whose sampled plot's main crop is maize and value output using the median observed maize unit value among maize sellers. We prefer the median to the mean because seller-reported unit values are thin and strongly right-skewed for several crops.

Table 11 separates value of production from costs and then reports profits in levels. We report each outcome both in total and per acre. The cost measure is intentionally transparent: it adds seed expenditure, pesticide expenditure, separately reported other sampled-plot expenses, and fertilizer expenditures. The questionnaire elicits those components in separate modules, so this construction does not mechanically double count fertilizer or seed costs inside the residual expense term. The fertilizer-cost component counts only fertilizer the household purchased with its own resources. The endline question elicits the cost of all fertilizer applied to the plot, and respondents report voucher-funded fertilizer at its full commercial value even though the voucher is a transfer they did not pay for. We therefore net the voucher-funded kilograms out of each fertilizer record, valued at that record's own reported price per kilogram, so that costs reflect farmer-private outlays rather than the value of the transfer. Fertilizer obtained through the national subsidy programme is retained, because respondents report it at the subsidised price they actually paid.[8]

---

[8] Two cost components are deliberately excluded. First, family labour is not priced in the cost vector; only cash outlays for hired labour are captured, and these enter through the residual sampled-plot expense term. To the extent that the voucher-induced fertilizer increase crowds in additional own-labour for basal and top-dressing application, omitting an opportunity-cost wage on family labour overstates T2 net returns. Beaman et al. (2013) document a similar labour-cost-driven attenuation of profit gains in Mali and are the cleanest design analogue for the direction of bias here. Second, the soil-test cost itself is excluded; this is correct for farmer-private ITT returns but should be borne in mind for any social or programme-level cost-effectiveness

Table 11: Effects on Value of Production, Costs, and Profits

Outcome	Control mean	T1	T2	p-value: T2 = T1	N
Maize main crop					
Value of production	605,184	-13,432	96,133	0.084	1,625
		(59,976)	(62,014)		
Total costs	221,781	15,229	-93,683***	0.000	1,625
		(29,270)	(27,098)		
Profits	383,404	-28,660	189,816***	0.000	1,625
		(46,760)	(52,386)		
Value per acre	464,244	4,972	320,814**	0.052	1,625
		(49,497)	(159,924)		
Costs per acre	184,474	12,339	-48,417	0.018	1,625
		(37,165)	(35,880)		
Profits per acre	279,770	-7,367	369,231**	0.014	1,625
		(38,467)	(147,640)		
All sampled main crops					
Value of production	832,945	988,043	-63,844	0.324	1,868
		(1,067,100)	(120,805)		
Total costs	235,542	7,535	-100,549***	0.000	1,868
		(29,362)	(27,130)		
Profits	597,403	980,508	36,705	0.375	1,868
		(1,063,651)	(113,458)		
Value per acre	508,306	771,249	45,202	0.373	1,867
		(813,110)	(63,606)		
Costs per acre	145,973	-101	-55,708***	0.000	1,867
		(13,235)	(11,993)		
Profits per acre	362,333	771,349	100,911*	0.410	1,867
		(811,859)	(60,154)		

Notes: Unadjusted ITT effects in Malawi kwacha. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher redeemable for any fertilizer product at the nearest Farmers World shop. Value of production equals harvested quantity times the crop-specific median observed sale price among sellers of that crop. Costs sum seed expenditure, pesticide expenditure, separately reported other sampled-plot expenses, and household-purchased fertilizer costs; fertilizer funded by the voucher is netted out so that costs measure farmer-private outlays. Profits equal value of production minus costs. Per-acre outcomes use maize area in the maize panel and sampled-plot area in the all-sampled-main-crop panel. The column labelled *p-value: T2 = T1* reports the test of equality between the two treatment effects. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

The economic-return estimates sharpen the interpretation of the production results. In the maize-main-crop sample, the voucher raises the value of production by about 96,000 MWK, though this level effect is not precisely estimated. It also *lowers* the household's own input costs, by about 94,000 MWK: because the voucher fertilizer substitutes for fertilizer the household would otherwise have bought commercially, T2 households spend less of their own cash on inputs than the control group. The two effects combine into a farmer-private profit gain of about 190,000 MWK, statistically significant and roughly half the control-group profit level; the per-acre estimates tell the same story. Appendix Table 16 shows that the profit gain is not an artifact of a single coding choice: it remains in the 190,000–205,000 MWK range and significant when harvested quantity is winsorized and when both harvested quantity and total costs are winsorized, and the positive-profit log specification is directionally consistent.

Importantly, not all of the voucher's value to the household comes through higher output. The voucher could purchase roughly 75 kg of fertilizer, yet total fertilizer use on the test plot rose by only about 30 kg, so much of the voucher fertilizer displaced fertilizer the household would otherwise have bought commercially. This displacement is what drives the negative cost effect in Table 11: T2 households spend about 94,000 MWK less of their own money on inputs than the control group. For the farmer this saving is a genuine benefit, and it is already embedded in the profit estimate reported above. From a programme standpoint, however, the saving is a transfer rather than newly created value, since the displaced kilograms would have been purchased in any case.

7.7 Cost-benefit analysis

A central question in the debate on fertilizer subsidies is whether the induced yield gains exceed the subsidy's cost. The voucher costs 170,000 MWK per beneficiary, while the estimated yield gain in the maize sample is 151 kg. Whether this gain justifies the subsidy depends critically on the price used to value maize. At the sample median sales price of 750 MWK/kg, the implied production value is 113,250 MWK, which is substantially below the voucher cost.

This valuation, however, may understate the private benefits to farmers. About 44 percent of maize consumption in Malawi is met through market purchases (De Weerdt, Duquennois, and Oliveres-Mallol, 2026). The additional output of roughly three 50 kg bags would sustain a typical household for about three months. To the extent that the additional production displaces purchases, it should be valued at prevailing retail prices rather than at farmgate prices. While we cannot know the timing these displaced purchases, retail prices rose to 1,360 MWK/kg within the consumption window following the 2025 harvest, and reached as high as 1,742 MWK/kg in the preceding year (IFPRI, 2026). These large price fluctuations — and, crucially, the choice between farmgate and retail prices —substantially affect the profitability assessment.

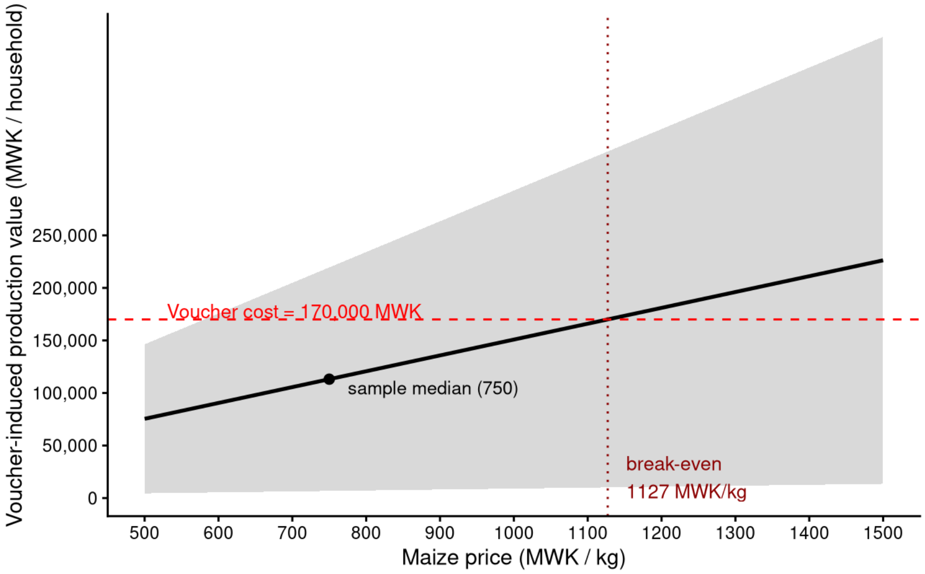
Figure 4 illustrates this sensitivity by plotting the value of voucher-induced production against maize prices alongside the voucher cost, but excluding the cost of the soil test. Given the estimated yield gain of 151 kg, the implied break-even price is approximately 1,126 MWK/kg.

8 Conclusion

Fertilizer subsidy programmes dominate agricultural spending across much of sub-Saharan Africa, and in recent years Malawi's Affordable Inputs Programme (AIP) - now renamed FISP - has consistently been among the largest of these in budget share. With fertilizer prices high and fiscal space tight, the case for moving from a blanket subsidy bundle to a site-specific, soil-test-anchored alternative has been pressed in both the policy and academic literatures. The argument rests on two empirical premises: that recommendations derived from soil tests redirect farmer behaviour toward an agronomically appropriate input mix, and that pairing such recommendations with the financial means to act on them delivers both higher yields and a more efficient use of the fiscal envelope. This paper tests both premises in a cluster-randomized trial across 113 villages in Malawi. Treatment 1 households received a plot-specific soil-test recommendation; Treatment 2 received the same recommendation together with a fertilizer voucher of AIP-equivalent value, redeemable at the nearest Farmers World shop for any fertilizer product; the control group received neither. Parsed

---

calculation. Both omissions push the T2 profit point estimate upward relative to a fully-costed accounting.



Yield effect: 150.83 kg per household (Table 5, Panel A); shaded band: 95% CI.

Figure 4: Voucher-induced value of maize production as a function of the maize price. Solid line uses the T2 yield effect of 151 kg per household; shaded band is the 95 percent confidence interval. Dashed horizontal line marks the voucher cost (170,000 MWK). The break-even maize price is approximately 1,126 MWK/kg.

product-level recommendation data allow us to compare the voucher arm's realised input mix against the prescription on a product-by-product basis.

The recommendation arm changed neither input use, nutrient application, nor maize yield. The voucher arm raised fertilizer use on the soil-test plot by about 30 kg, raised maize total production by about 19 percent, and partially crowded out AIP receipt at the household level. Yet the composition of that increment is not what the site-specific recommendation framing predicts. The voucher arm delivers intensification of the existing Malawi maize fertilizer bundle rather than precision fertilization. The voucher-induced increment is composed almost entirely of 18 kg more urea and 19 kg more NPK 23:10:5, a 50/50 split that mirrors the AIP bundle the voucher is partially displacing. Where the recommendation differs from the modal AIP prescription, the realised mix does not shift toward it: compliance with CAN is lower in the voucher arm than in the recommendation-only arm, and compliance with potassium products, lime, and MAP is zero or near-zero in every arm.

The additional maize gain is generated almost entirely by nitrogen intensification of an already nitrogen-centric bundle: urea and NPK 23:10:5 quantities rise, but the phosphorus, potassium, and lime gaps the recommendation was designed to close remain open. To the extent that these nutrients are co-limiting with nitrogen, as the soil-test evidence in our sample suggests, the measured response is a lower bound on the agronomic potential that full-compliance fertilization would unlock.

The break-even calculation reported earlier is a programme-level one, weighing the value of the extra output against the full cost of the voucher. From the farmer's own standpoint the picture is more favourable. Because the voucher fertilizer largely displaces fertilizer the household would otherwise have bought commercially, treated households spend about 94,000 MWK less of their own money on inputs, and their farm profit rises by roughly 190,000 MWK, an effect that is statistically significant and robust to alternative valuations. The voucher is thus privately profitable for recipients even where, at prevailing maize prices, it does not recover its cost to the public purse. The displaced spending is a transfer rather than newly created value, so it improves the household's position without strengthening the fiscal case.

Our T1 null replicates the modal result in the comparable literature (Harou et al., 2022; Beg, Islam, and Rahman, 2024; Corral et al., 2020) and diverges from Ayalew, Chamberlin, and Newman (2022), who find positive effects of recommendations alone in Ethiopia. We cannot identify the source of that divergence with our data. Candidates include product availability (Sida et al., 2023), AIP-induced habit formation around the standard maize bundle, lower baseline nutrient deficits, and differences in extension delivery. A further candidate, underemphasized in the existing literature, is the structure of the recommendation itself. In Ayalew, Chamberlin, and Newman (2022) the prescribed inputs were already known to Ethiopian smallholders, and the recommendation represented a quantitative adjustment, how much of a known product to apply, rather than a product switch. In the present setting the modal recommendation asks farmers to substitute CAN for urea, add a potassium product where none was previously used, or apply lime on a plot that has never received it. A recommendation that requires a product switch faces a qualitatively different behavioural barrier than one requiring a quantity adjustment, and information delivery alone may not bridge it in a single season.

Several policy implications follow. Information-only programmes such as soil-health-card schemes, blanket extension messaging, and stand-alone decision support tools are unlikely to redirect smallholder fertilizer behaviour at scale in environments resembling ours. A recommendation-paired voucher of AIP-equivalent value, as deployed here, intensifies the AIP-composition bundle on the soil-test plot without redirecting it toward the prescription. The case for swapping AIP for a recommendation-paired voucher cannot be made on agronomic-precision grounds in this setting. It can still be made on commercial-channel grounds, since routing fiscal support through agro-dealers rather than parallel public distribution may improve targeting and build private input markets (Basurto, Dupas, and Robinson, 2020), but those benefits are not measured here and should not be conflated with the precision-agronomy benefits this experiment refutes. The AIP-substitution finding (partial, at less than one-tenth of the voucher's commercial-price equivalent) further argues for integrating any future recommendation-paired voucher with the existing AIP architecture rather than running the two in parallel: the policy-relevant counterfactual is not voucher versus zero but voucher versus AIP.

Two further implications concern recommendation design. The voucher induced a shift away from organic inputs (fresh vegetative material, manure, Mbeya fertilizer) toward mineral fertilizer. This may carry a long-run soil-health cost: sustained substitution depletes organic matter, reducing moisture retention and long-term soil fertility. A future programme could address this by incorporating organic input advice alongside the fertilizer prescription. The compliance evidence also reveals a systematic divergence between the agronomic and the behaviourally feasible optimum. A behaviourally informed architecture would distinguish between a core bundle, the products the farmer already uses and for which the recommendation adjusts quantity and timing, and an aspirational bundle of products that address additional deficiencies but require supply-chain development and demonstrated farmer familiarity before they are actionable. A staged approach, consolidating core-bundle compliance first and then introducing novel products through demonstration, supply pre-positioning, and extension on product equivalence, is more likely to generate durable input reallocation than a single-season full-prescription delivery.

The experiment leaves open questions that bear on the design of future programmes. Whether the T2 response persists once the voucher is withdrawn, and whether the recommendation-only arm produces lagged effects through learning or social diffusion, cannot be answered with single-season data; a medium-run follow-up is the most direct way to ask whether the behavioural gains persist. The conditions under which information alone shifts behaviour are also not yet established: candidate moderators include the content of the recommendation, agro-dealer product depth, the magnitude of subsidy support running through parallel channels, and farmer baseline use, and a coordinated multi-country replication varying these features within a common design would be more informative than additional single-country trials. A complementary supply-side experiment that randomises pre-positioning of less common products at agro-dealers would speak directly to whether the missing phosphorus, potassium, MAP, and lime compliance reflects farmer demand, agro-dealer stocking, or both. Our results also raise a question about what recommendations should optimise. Recommendations that target only the agronomic optimum give farmers no basis for prioritising when full compliance is infeasible; one that additionally provides a marginal ranking of products would be more actionable under budget and supply constraints, and randomising such priority orderings would identify whether the compliance gap is bridgeable by better recommendation design or requires deeper supply-side reform.

Three structural features of the Malawi context are likely portable: the dominance of familiar products anchored by a long-running national subsidy programme; thin agro-dealer product depth outside the subsidy bundle; and liquidity constraints that make farmers risk-averse toward unfamiliar input expenditures. The null recommendation effect and the AIP-bundle mimicry in the voucher arm are likely to replicate in settings that share these features. Two features are less portable: the 40 percent crowd-out of AIP receipt reflects Malawi's specific institutional architecture in which village chiefs mediate AIP allocation, so in more administratively managed subsidy regimes crowd-out may be smaller; and zero potassium compliance may be specific to Malawi's agro-dealer product mix, with K compliance failure likely less severe where potassium products are more commonly stocked, as in parts of Kenya and Rwanda. More broadly, our results caution against assuming that improved agronomic information and relaxed liquidity constraints will automatically translate into precision fertilizer adoption. Farmers may continue to rely on familiar fertilizer products when recommended alternatives are perceived as unfamiliar, unavailable, or risky. Future subsidy reforms seeking to promote site-specific nutrient management may therefore require complementary interventions that address product availability, farmer learning, and confidence in alternative fertilizer technologies.

The broader lesson for scaling site-specific fertilizer advisory systems is that the bottleneck is not only information quality but the translation from agronomic prescription to locally feasible input choices. Programmes that integrate recommendations with product availability, liquidity support, and delivery infrastructure are more likely to affect behaviour, but even then the likely margin of change may be partial and product-specific rather than complete adoption of the agronomic bundle.

Acknowledgements

This research was supported by the CGIAR Science Program on Policy Innovations. We acknowledge the contributions of all donors and partners who have generously supported this work through their foundational contributions to CGIAR.

Note

The trial was registered on the American Economic Association (AEA) RCT Registry on 14 October 2024 (De Weerdt et al., 2024b). Ethics approval was obtained from the International Food Policy Research Institute (IFPRI) Institutional Review Board and from the National Committee on Research in the Social Sciences and Humanities (NCRSH) in Malawi.

References

African Union. 2024. *African Fertilizer and Soil Health Action Plan.* Tech. rep., African Union.

Aggarwal, S., D. Jeong, N. Kumar, D. S. Park, J. Robinson, and A. Spearot. 2024. “Shortening the path to productive investment: Evidence from input fairs and cash transfers in Malawi.” *Journal of Development Economics* 170: 103288.

Aker, J. C. 2017. “Comparing cash and voucher transfers in a humanitarian context: Evidence from the Democratic Republic of Congo.” *The World Bank Economic Review* 31 (1): 44–70.

Arouna, A., J. D. Michler, W. G. Yergo, and K. Saito. 2021. “One size fits all? Experimental evidence on the digital delivery of personalized extension advice in Nigeria.” *American Journal of Agricultural Economics* 103 (2): 596–619.

Ayalew, H., J. Chamberlin, and C. Newman. 2022. “Site-specific agronomic information and technology adoption: A field experiment from Ethiopia.” *Journal of Development Economics* 156: 102788.

Banda, H., J. De Weerdt, J. Duchoslav, and A. Jolex. 2022. *Pricing farmer contributions under the Affordable Inputs Programme.* MaSSP Policy Note 47, International Food Policy Research Institute (IFPRI) Malawi Strategy Support Program.

Basurto, M. P., P. Dupas, and J. Robinson. 2020. “Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi.” *Journal of Public Economics* 185: 104047.

Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry. 2013. “Profitability of fertilizer: Experimental evidence from female rice farmers in Mali.” *American Economic Review: Papers & Proceedings* 103 (3): 381–386.

Beg, S., M. Islam, and K. W. Rahman. 2024. “Information and behavior: Evidence from fertilizer quantity recommendations in Bangladesh.” *Journal of Development Economics* 166: 103195.

Benson, T., J. De Weerdt, J. Duchoslav, and W. Masanjala. 2024. *Fertilizer subsidies in Malawi: From past to present.* MaSSP Working Paper 44, International Food Policy Research Institute (IFPRI) Malawi Strategy Support Program.

Berazneva, J., A. Maertens, W. Mhango, and H. Michelson. 2023. “Paying for agricultural information in Malawi: The role of soil heterogeneity.” *Journal of Development Economics* 165: 103144.

Burke, W. J., E. Frossard, and T. Jayne. 2017. “Agricultural input credit in Africa.” *Applied Economic Perspectives and Policy* 39 (4): 569–591.

Chandrasekhar, A. G., E. Duflo, M. Kremer, J. a. F. Pugliese, J. Robinson, and F. Schilbach. 2022. “Blue Spoons: Sparking Communication About Appropriate Technology Use.” NBER Working Paper 30423; BFI Working Paper 2022-123.

Corral, C., X. Giné, A. Mahajan, and E. Seira. 2020. *Appropriate Technology Use and Autonomy: Evidence from Mexico.* Working Paper 27681, National Bureau of Economic Research.

Croppenstedt, A., M. Demeke, and M. M. Meschi. 2003. “Technology adoption in the presence of constraints: The case of fertilizer demand in Ethiopia.” *Review of Development Economics* 7 (1): 58–70.

Cunha, J. M. 2014. “Testing paternalism: Cash versus in-kind transfers.” *American Economic Journal: Applied Economics* 6 (2): 195–230.

De Weerdt, J., B. Dillon, E. Hami, B. Van Campenhout, and L. Nabwire. 2024a. “Expecting too much, foreseeing too little? Behavioral explanations for the sell low-buy high puzzle in smallholder market participation.” Mimeo.

De Weerdt, J., C. Duquennois, and A. Oliveres-Mallol. 2026. “Changing rural labor markets and welfare in Malawi.” *World Development Perspectives* 42: 100784.

De Weerdt, J. et al. 2024b. “The Impact of Site Specific Soil Fertility Recommendations: experimental evidence from Malawi.” AEA RCT Registry.

Denning, G., P. Kabambe, P. Sanchez, A. Malik, R. Flor, R. Harawa, P. Nkhoma, C. Zamba, C. Banda, C. Magombo, M. Keating, J. Wangila, and J. Sachs. 2009. “Input subsidies to improve smallholder maize productivity in Malawi: Toward an African Green Revolution.” *PLoS Biology* 7 (1): e1000023.

Duchoslav, J. and J. De Weerdt. 2022. *Guiding principles of AIP reform.* MaSSP Policy Note 43, International Food Policy Research Institute (IFPRI) Malawi Strategy Support Program.

———. 2023. “Three outstanding questions about AIP 2.0.” IFPRI Malawi Strategy Support Program (MaSSP) blog post. Accessed 2026-05-15.

Duchoslav, J. and E. Kenamu. 2018. *Are social safety nets and input subsidies reaching the poor in Malawi?* MaSSP Working Paper 26, International Food Policy Research Institute (IFPRI) Malawi Strategy Support Program.

Duflo, E., M. Kremer, and J. Robinson. 2011. “Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya.” *American Economic Review* 101 (6): 2350–2390.

Galloway, J. N., A. R. Townsend, J. W. Erisman, M. Bekunda, Z. Cai, J. R. Freney, L. A. Martinelli, S. P. Seitzinger, and M. A. Sutton. 2008. “Transformation of the nitrogen cycle: recent trends, questions, and potential solutions.” *Science* 320 (5878): 889–892.

Gars, J., R. Fishman, A. Kishore, Y. Rothler, and P. S. Ward. 2025. “Confidence and information usage: Evidence from soil testing in India.” *American Journal of Agricultural Economics* 107 (5): 1406–1437.

Harou, A. P., M. Madajewicz, H. Michelson, C. A. Palm, N. Amuri, C. Magomba, J. M. Semoka, K. Tschirhart, and R. Weil. 2022. “The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania.” *Journal of Development Economics* 155: 102707.

Harrigan, J. 2008. “Food insecurity, poverty and the Malawian Starter Pack: Fresh start or false start?” *Food Policy* 33 (3): 237–249.

Hebebrand, C., J. Glauber, R. Vos, and B. Rice. 2026. *The Iran wars impacts on global fertilizer markets and food production.* Tech. rep., International Food Policy Research Institute (IFPRI).

Hidrobo, M., J. Hoddinott, A. Peterman, A. Margolies, and V. Moreira. 2014. “Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador.” *Journal of Development Economics* 107: 144–156.

Hill, R. V. and D. Resnick. 2025. *Realistic Options for Repurposing Fertilizer Subsidy Spending.* Project note, International Food Policy Research Institute (IFPRI), Washington, DC.

Holden, S. T. 2019. “Input subsidies and market development: A review of recent policy experiences in Africa.” *Food Security* 11 (2): 379–396.

Hoynes, H. W. and D. W. Schanzenbach. 2009. “Consumption responses to in-kind transfers: Evidence from the introduction of the Food Stamp Program.” *American Economic Journal: Applied Economics* 1 (4): 109–139.

Islam, M. and S. Beg. 2021. “Rule-of-Thumb Instructions to Improve Fertilizer Management: Experimental Evidence from Bangladesh.” *Economic Development and Cultural Change* 70 (1): 237–281.

Jack, B. 2013. “Constraints on the adoption of agricultural technologies in developing countries.” *Annual Review of Resource Economics* 5: 245–269.

Jayne, T., N. M. Mason, W. J. Burke, and J. Ariga. 2018. “Smart subsidies and sustainable agricultural intensification in Africa.” *Food Policy* 76: 34–50.

Johnston, A. M. and T. W. Bruulsema. 2014. “Balanced fertilizers to improve crop nutrient use efficiency.” *Better Crops* 98 (2): 4–6.

Kilic, T., E. Whitney, and P. Winters. 2013. *Decentralised beneficiary targeting in large-scale development programmes: Insights from the Malawi Farm Input Subsidy Programme.* Policy Research Working Paper 6713, World Bank.

Lunduka, R., J. Ricker-Gilbert, and M. Fisher. 2013. “What are the farm-level impacts of Malawi’s farm input subsidy program? A critical review.” *Agricultural Economics* 44 (6): 563–579.

Maertens, M., O. Oyinbo, T. Abdoulaye, and J. Chamberlin. 2023. “Sustainable maize intensification through site-specific nutrient management advice: Experimental evidence from Nigeria.” *Food Policy* 121: 102546.

Mather, D. and T. Jayne. 2018. “Input subsidy programs and agricultural productivity in Africa: A review of recent evidence.” *Food Policy* 76: 88–100.

Nyondo, C., M. Khonje, J. Mangisoni, W. J. Burke, J. Ricker-Gilbert, and L. Chilora. 2021a. *The Affordable Inputs Programme in Malawi: Promises and pitfalls.* Policy Brief 7, MwAPATA Institute.

———. 2021b. *Lessons learnt: Promises, achievements and pitfalls of the Farm Input Subsidy Programme in Malawi.* Tech. rep., MwAPATA Institute.

Nyondo, C. J., J. Minofu, J. Goeb, W. J. Burke, D. Kambewa, R. Chikowo, and S. Snapp. 2025. “The impact of site-specific soil-test-based extension advice on farm management in Malawi.” *Food Policy* 133: 102850.

Office of the Ombudsman of Malawi. 2024. “A Report on Systemic Investigations on the Affordable Inputs Programme Implementation.” Office of the Ombudsman, Lilongwe. Accessed 2026-05-15.

Oyinbo, O., J. Chamberlin, T. Abdoulaye, and M. Maertens. 2022. “Digital extension, price risk, and farm performance: experimental evidence from Nigeria.” *American Journal of Agricultural Economics* 104 (2): 831–852.

Ricker-Gilbert, J., T. S. Jayne, and E. Chirwa. 2011. “Subsidies and crowding out: A double-hurdle model of fertilizer demand in Malawi.” *American Journal of Agricultural Economics* 93 (1): 26–42.

Schut, A. G. and K. E. Giller. 2020. “Soil-based, field-specific fertilizer recommendations are a pipe-dream.” *Geoderma* 380: 114680.

Sida, T. S., S. Gameda, J. Chamberlin, J. A. Andersson, M. Getnet, L. Woltering, and P. Craufurd. 2023. “Failure to scale in digital agronomy: An analysis of site-specific nutrient management decision-support tools in developing countries.” *Computers and Electronics in Agriculture* 212: 108060.

Suri, T. 2011. “Selection and comparative advantage in technology adoption.” *Econometrica* 79 (1): 159–209.

Van Campenhout, B., L. Nabwire, B. Kramer, C. Trachtman, and G. T. Abate. 2025. *Hybrid Gains and Tasty Grains: The Role of Agronomic and Consumption Traits in Technology Adoption.* IFPRI Discussion Paper 2331, International Food Policy Research Institute (IFPRI), Washington, DC.

World Bank. 2025. *Malawi Agriculture Public Expenditure Review.* Tech. rep., World Bank.

A Design and realization diagnostics

This section reports diagnostics that speak to the integrity of the experimental sample before turning to secondary outcomes and robustness checks. The purpose is not to add additional outcome tests, but to document the realized analysis sample: how the baseline frame maps into completed endline interviews, whether realization differs by treatment arm, and whether the retained endline sample remains balanced on baseline covariates.

Table 12: Sample realization by treatment arm: counts and percentages

Stage	Control	Treatment 1	Treatment 2	Total
Baseline sampled	612 (100.0%)	720 (100.0%)	702 (100.0%)	2034 (100.0%)
Endline target	591 (96.6%)	675 (93.8%)	686 (97.7%)	1952 (96.0%)
Interview completed	557 (91.0%)	636 (88.3%)	675 (96.2%)	1868 (91.8%)
Matched to soil test data	401 (65.5%)	607 (84.3%)	618 (88.0%)	1626 (79.9%)

Note: Entries report counts followed by the percentage of the original sampled baseline frame in parentheses. *Baseline sampled* denotes households selected into the Space2Place (S2P) study sample from the 2022 source-study frame. *Endline target* denotes households appearing in the realized endline survey master file after treatment-administration and tracking efforts identified households that had moved, could not be found, or otherwise could not be retained as endline targets. *Interview completed* denotes households with a usable endline interview after the consent screen. *Matched to soil test data* denotes endline observations successfully linked to the administrative soil-test records used to construct soil-test-based outcomes.

Table 13: Differential attrition by survey stage

Stage	Control mean	T1 − Control	T2 − Control	p-value: T1 = T2	N
Endline target	0.966	-0.028	0.012	0.022	2034
		(0.018)	(0.010)		
Interview completed	0.942	0.000	0.041	0.002	2034
		(0.015)	(0.011)		
Matched to soil test data	0.679	0.221	0.222	0.964	2034
		(0.070)	(0.070)		

Note: Each row reports a linear probability regression of the indicated stage indicator on mutually exclusive T1 and T2 assignment dummies, with the control group omitted. The sample is the full baseline study frame. Standard errors clustered at the village level are shown in parentheses. The fifth column reports the p-value for equality of the T1 and T2 coefficients. The final row, matched to soil test data, is a data-linkage margin relevant only for outcomes that require soil-test-linked information; it is not a survey-realization stage for outcomes measured from the endline survey alone.

B Secondary outcomes

The main text focuses on the causal chain from treatment delivery to recommendation compliance, fertilizer use, nutrient application, maize yield, and economic returns. The table below reports pre-specified secondary soil nutrient management practices that are useful for interpretation but are not central to the paper's main mechanism.

The results do not point to a broad increase in complementary soil nutrient management (SNM) practices. T1 is broadly null. T2 increases the use of ridges or check dams by about 9.0 percentage points, but it reduces reported use of fresh vegetative material, dairy or poultry manure, and Mbeya fertilizer. These patterns suggest that the voucher-induced fertilizer response did not operate through a general shift toward more intensive soil nutrient management.

C Sensitivity checks

This section reports robustness checks for the economic and yield results. These checks are included to show how sensitive the main conclusions are to alternative handling of prices, profits, and extreme yield values.

Table 14: Baseline balance among households retained through completed endline interview

Variable	Control mean	T1 − Control	T2 − Control	N
Household head age	42.514	1.049	1.103	1843
		(0.892)	(0.820)	
Household head male	0.779	0.018	0.025	1868
		(0.027)	(0.026)	
Household size	4.892	0.106	0.309	1866
		(0.128)	(0.138)	
Land area (ha)	1.011	0.003	0.007	1863
		(0.056)	(0.059)	
Difficulty feeding family	0.271	0.058	0.047	1868
		(0.040)	(0.035)	

Note: This table repeats the baseline balance exercise on the subsample of households with a usable completed endline interview. Coefficients are estimated relative to the control group using village-clustered standard errors shown in parentheses. If attrition were strongly selective on these baseline covariates, large treatment-control differences would emerge here.

Table 15: Effects on Other Soil Nutrient Management Practices

Outcome	Control mean	T1 − Control	T2 − Control	p-value: T2 = T1	N
Green legume incorporation	0.129	0.038	0.046*	0.737	1867
		(0.024)	(0.027)		
Fresh vegetative material	0.162	-0.037	-0.068**	0.220	1867
		(0.027)	(0.027)		
Farmyard manure	0.262	-0.010	-0.046	0.218	1866
		(0.034)	(0.035)		
Dairy or poultry manure	0.264	0.007	-0.098***	0.001	1866
		(0.032)	(0.031)		
Compost	0.237	-0.028	-0.005	0.473	1865
		(0.030)	(0.034)		
Mbeya fertilizer	0.065	0.031	-0.032**	0.000	1866
		(0.019)	(0.013)		
Minimum tillage	0.054	-0.005	-0.003	0.916	1863
		(0.014)	(0.016)		
Ridges/check dams	0.637	0.074**	0.092***	0.586	1867
		(0.032)	(0.033)		
Pit planting	0.079	-0.003	-0.018	0.407	1867
		(0.019)	(0.018)		

Notes: This table reports unadjusted ITT estimates from linear probability models. Outcomes equal one if the household reports using the indicated practice on the sampled plot and zero otherwise; blank responses are treated as missing. T1 provides site-specific fertilizer recommendations; T2 provides the same recommendations together with a voucher redeemable for any fertilizer product at the nearest Farmers World shop. The column labelled *p-value: T2 = T1* reports the test of equality between the two treatment effects. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

They do not change the main interpretation: the recommendation-plus-voucher arm produces clear effects on implementation, fertilizer use, maize productivity, and farmer-private profit, even though the level value of production is less precisely estimated.

Table 16: Maize-profit robustness checks

Robustness check	Control mean	T1	T2	p-value: T2 = T1	N
Level profits, baseline valuation	383,404	-28,660	189,816***	0.000	1625
		(46,760)	(52,386)		
Level profits, winsorized harvest quantity	358,448	-28,573	205,061***	0.000	1625
		(41,661)	(49,541)		
Level profits, winsorized harvest and costs	372,464	-37,328	194,109***	0.000	1625
		(40,098)	(48,201)		
Log profits, positive-profit sample only	12.387	-0.075	0.574***	0.000	1478
		(0.110)	(0.111)		

Note: Each row reports the unadjusted maize-main-crop specification under an alternative profit construction. Baseline profits value harvested maize using the median observed maize unit value among maize sellers. The winsorized specifications cap harvested quantity and, in the final row of Panel A, total costs at the 1st and 99th percentiles of the maize-main-crop sample. The positive-profit log specification is included only as a diagnostic because it drops households with nonpositive profits. Standard errors are clustered at the village level. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

Table 17: Maize-yield estimates under alternative extreme-value rules

Scenario	N	T1	T2	p-value: T2 = T1
Preferred adjusted specification	1564	-0.12	0.29***	0.000
Drop largest yield observation	1563	-0.12	0.28***	0.001
Drop yield above 10,000 kg/acre	1560	-0.13	0.27***	0.000
Drop yield above 5,000 kg/acre	1550	-0.13	0.27***	0.000
Drop maize area below 0.1 acre	1562	-0.12	0.28***	0.001

Note: Each row reports the unadjusted specification for log maize yield among households whose main crop is maize. The scenarios examine whether the T2 yield effect is driven by extreme yield values created by small maize-area denominators. Standard errors are clustered at the village level. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

D Variable construction notes

Economic outcomes are constructed for the sampled plot's main crop, not for total household crop revenue across all plots and crops. Value of production equals harvested quantity multiplied by the crop-specific median sale price observed among sellers of the same crop. Costs sum fertilizer expenditure, seed expenditure, pesticide expenditure, and separately reported other sampled-plot production expenses; fertilizer expenditure counts only fertilizer the household purchased with its own resources, with voucher-funded kilograms netted out at each record's own price per kilogram so that costs reflect farmer-private outlays. Profits equal value of production minus these costs. The maize-main-crop panel is the preferred economic-return sample because maize has broad seller support and avoids the thin-price problem that affects several non-maize crops in the pooled panel.

Recommendation-compliance outcomes are defined only where the relevant recommendation records exist. Product compliance compares recommended product names to products farmers report applying. Nutrient compliance compares realized nutrient application to the nutrient totals implied by the recommendation. The paper does not use shadow recommendations for the control group because the control soil tests were analyzed under a different laboratory and calibration environment, which would make cross-arm shadow shortfalls difficult to interpret.

All specifications report unadjusted ITT estimates without pre-treatment controls. Post-treatment variables such as Agronaut visits or extension contact during implementation are not included.