INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

IFPRI Discussion Paper 02463

November 2026

Soil Tests, Vouchers, and the Limits of Site-Specific Fertilizer Recommendations

Experimental Evidence from Malawi

Thomas Assefa
Jonathan Atkinson
Hailemariam Ayalew Tiruneh
Joachim De Weerdt
Edwin Siyame
David J. Spielman
Bjorn Van Campenhout

Agrifood Innovation and Resilience Unit
Development Strategies and Governance Unit
Markets, Trade, and Institutions Unit

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), a CGIAR Research Center established in 1975, provides research-based policy solutions to sustainably reduce poverty and end hunger and malnutrition. IFPRI's strategic research aims to foster a climate-resilient and sustainable food supply; promote healthy diets and nutrition for all; build inclusive and efficient markets, trade systems, and food industries; transform agricultural and rural economies; and strengthen institutions and governance. Gender is integrated in all the Institute's work. Partnerships, communications, capacity strengthening, and data and knowledge management are essential components to translate IFPRI's research from action to impact. The Institute's regional and country programs play a critical role in responding to demand for food policy research and in delivering holistic support for country-led development. IFPRI collaborates with partners around the world.

AUTHORS

Thomas Woldu Assefa (t.assefa@cgiar.org) is a Research Fellow in the Agrifood Innovation and Resilience (AIR) Unit of the International Food Policy Research Institute (IFPRI), Nairobi, Kenya.

Jonathan Atkinson (jonathan.atkinson@solidaridadnetwork.org) is a Managing Director at Solidaridad Southern Africa, Johannesburg, South Africa.

Hailemariam Ayalew Tiruneh (h.tiruneh@cgiar.org) is a Research Fellow in IFPRI's Development Strategies and Governance (DSG) Unit, Nairobi, Kenya.

Joachim De Weerdt (j.deweerdt@cgiar.org) is a Senior Research Fellow in IFPRI's DSG Unit, Lilongwe, Malawi.

Edwin Siyame (e.siyame@cgiar.org) is a Research Analyst in IFPRI's DSG Unit, Lilongwe, Malawi.

David J. Spielman (dspielman@wustl.edu) is the Lauren and Lee Fixel Distinguished Professor at Washington University in St. Louis, St. Louis, MO.

Bjorn Van Campenhout (b.vancampenhout@cgiar.org) is a Senior Research Fellow in IFPRI's Markets, Trade, and Institutions (MTI) Unit, Leuven, Belgium.

Notices

[1]IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by IFPRI.

[2] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

[3]Copyright remains with the authors. The authors are free to proceed, without further IFPRI permission, to publish this paper, or any revised version of it, in outlets such as journals, books, and other publications.

Abstract

Rising and volatile fertilizer prices have sharpened concerns about the efficiency of fertilizer subsidy programmes across Africa, fueling a shift toward more precise, site-specific fertilizer recommendations based on high-resolution soil intelligence such as on-farm soil testing. The case for such reform rests on the premise that better intelligence can redirect farmers' input choices, either on its own or, where financial constraints bind, when combined with the means to act on it. We test this premise in Malawi, where the national subsidy programme provides farmers with a standard fertilizer package that makes limited use of soil intelligence. In a cluster-randomized trial, farmers are assigned to an individualized fertilizer recommendation based on soil-test results, the same recommendation combined with a subsidy-equivalent voucher, or 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 production rising by roughly a fifth. However, these gains are not driven by adoption of site-specific recommendations: voucher recipients overwhelmingly replicate and intensify the standard subsidy package, regardless of soil-test results, pushing most households past their recommended nitrogen use rates, while use of recommended alternative and supplementary fertilizers remains negligible. These findings suggest that further investment in soil-testing-based recommendation services—whether as part of input subsidy programmes or as independent services—should be preceded by a better understanding of the constraints preventing farmers from acting on individualized recommendations.

Keywords: Fertilizer subsidies, soil testing, soil intelligence, technology adoption, Malawi

Acknowledgments

This research was initially funded by the U.S. Agency for International Development in 2024 as an analytical and evaluative contribution to the "Space-to-Place" initiative led by the International Fertilizer Development Center. Additional support was provided in 2025/26 by the CGIAR Policy Innovations Science Program. We thank all funders who support this research through their contributions to the CGIAR Trust Fund. We thank Latha Nagarajan, Nathan Ives, Lesley Perlman, and Zachary Stewart for their keen interest in this work. We also thank the farmers, agronauts, and field staff who participated and supported this study.

1 Introduction

Public input subsidy programmes have long been the largest single budget line in agriculture sector spending across much of Sub-Saharan Africa. Jayne et al. (2018) identified ten countries operating such programmes at scale at the time of their review and estimated that the programmes account for roughly 30 percent of public agricultural expenditure in each country. Two decades of evidence from farm-level surveys document a recurring set of performance problems: low or negative benefit-cost ratios at a programme level, distortion of commercial input markets through parallel public distribution channels, substantial crowding-out of commercial fertilizer purchases by subsidized receipts, regressive and politically captured targeting, fiscal entrenchment that makes programmes difficult to wind down once established, and concerns about environmental externalities arising from excessive fertilizer use (Basurto, Dupas, and Robinson, 2020; Holden, 2019; Mather and Jayne, 2018; Galloway et al., 2008). The 2021–22 global fertilizer price spike amplified these concerns by increasing the fiscal cost of delivering each unit of subsidized fertilizer and raising the opportunity cost of public expenditure devoted to fertilizer subsidy programmes (Vos et al., 2025). That said, few governments have backed away from their subsidy programmes, resulting in a shift in the policy narrative from whether fertilizer subsidies should exist toward how they can be delivered more efficiently.

A central concern in this debate is that conventional subsidy programmes typically distribute standardized fertilizer packages across highly heterogeneous production environments. Because crop response and fertilizer profitability vary substantially across soils, rainfall conditions, and agro-ecological settings, uniform packages can misallocate subsidized inputs and reduce the productivity of public spending (Theriault, Smale, and Haider, 2018; Wortmann, Sones et al., 2017). This concern motivates growing interest in site-specific nutrient management (SSNM), which combines decades of soil-health research with soil testing, remote sensing, agronomic field trials, and other soil intelligence efforts meant to tailor fertilizer products and use-rates to smallholder production conditions (Vanlauwe et al., 2026; Chivenge et al., 2022).

The agronomic case for SSNM is clear. Blanket recommendations are poorly matched to local soil conditions, often omit limiting micronutrients, and ignore substantial within-field heterogeneity in soil chemistry and crop response (Johnston and Bruulsema, 2014). The open question is whether this technical case translates into changes in farmer behaviour. Direct experimental tests of soil-test recommendations without accompanying input or financial support have produced mixed results. Ayalew, Chamberlin, and Newman (2022) find that soil-test recommendations alone raised 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 affected input use or yields in Bangladesh. Most closely related to our setting, Nyondo et al. (2025) report results from a cluster RCT in central Malawi in which soil-test-based extension advice raised adoption of complementary soil-management practices but did not move commercial inorganic fertilizer use. 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. Evidence from digital advisory experiments is similarly mixed (Maertens et al., 2023; Oyinbo et al., 2022; Arouna et al., 2021).[1]

More encouraging results emerge when the recommendation is paired with cash, a voucher, or a discount. Harou et al. (2022) find in Tanzania that neither recommendations nor vouchers alone substantially alter fertilizer choices, whereas combining them sharply increases adoption of the recommended product. Ammonium sulfate uptake rises from 4 percent with the voucher alone to 50 percent when the voucher is paired with a recommendation, despite sulfur deficiency on 92 percent of plots. The response operates primarily through input choice rather than total fertilizer use. Corral et al. (2020) find that recommendation-plus-grant raises yields in Mexico by 12-17 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 commonality in this strand is that information and finance behave as complements: profitable technologies may not be adopted when households face binding credit, liquidity,

[1] One possible explanation is that the technical limitations run deeper than delivery or uptake alone. Some have questioned the intrinsic value of soil-test recommendations, irrespective of delivery, financing, or other issues. 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 too weak to support genuinely field-specific recommendations.

or confidence constraints, a pattern familiar from the broader technology adoption literature (Jack, 2013; Beaman et al., 2013; Suri, 2011; Duflo, Kremer, and Robinson, 2011; Croppenstedt, Demeke, and Meschi, 2003).

In Malawi, the national fertilizer subsidy, which operated as the Affordable Inputs Programme (AIP) during our study, delivers a uniform, fixed-composition package (one 50 kg bag of nitrogen-phosphorus-potassium (NPK) basal blend and one 50 kg bag of urea for top-dressing) at heavily subsidized prices, regardless of local soil conditions.[2] This makes the programme a quasi in-kind transfer that fixes not only the level of support but also its composition. This institutional setting creates a natural composition-margin question: when farmers receive plot-specific recommendations and fertilizer purchasing power of comparable value, do they move away from the fixed AIP input mix toward a different product bundle? The Malawi fertilizer subsidy literature has documented partial crowding-out of commercial purchases by subsidized receipts (Ricker-Gilbert, Jayne, and Chirwa, 2011), but has not, to our knowledge, directly tested this composition margin.

Against this institutional backdrop, we conduct a cluster-randomized intervention that evaluates two policy-relevant alternatives to the standard subsidy package. The first is an individualized, plot-specific soil test accompanied by site-specific fertilizer recommendations provided by private-sector extension agents. The second is the same testing and recommendation treatment paired with a 170,000 Malawi Kwacha (MWK) voucher (approximately USD 100 at prevailing exchange rates) redeemable for fertilizer at a designated nearby private-sector agro-dealer participating in the intervention.[3] The voucher was calibrated to match the implicit subsidy in the AIP package, but allowed households to choose any stocked fertilizer product and bag size without a co-payment requirement. Conditional on the recommendation, this design allows us to observe whether added purchasing power is used to move away from the fixed AIP input mix toward the site-specific prescription. The trial enrolled 2,034 households across 113 villages in central Malawi, where maize is the dominant crop.

Four sets of findings emerge. First, assignment to the soil-test recommendation alone produces no detectable change in fertilizer use, nutrient application, yields, or returns. This null result is consistent with the modal finding in comparable studies (Beg, Islam, and Rahman, 2024; Harou et al., 2022; Corral et al., 2020). Second, adding the voucher raises fertilizer use on the soil-test plot by about 29–31 kg, or 35–40 percent, and increases maize production by about 19 percent. Third, the input mix reveals that this production gain is not driven by adoption of the site-specific prescription. The voucher-induced increment is concentrated in urea and NPK, in roughly the same 50/50 split as the AIP package it partly displaces. Products that distinguish the soil-test recommendation from the standard Malawi maize package, namely calcium ammonium nitrate (CAN), potassium products, lime, and mono-ammonium phosphate (MAP), are adopted at or near zero, and exact-bundle compliance is lower in the voucher arm than in the recommendation-only arm. The result is miscalibration in both directions: the share of households applying more nitrogen than their plot-specific recommendation doubles from 31 to 61 percent, while the phosphorus and potassium prescriptions remain essentially unfilled. 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. Voucher households see large profit gains driven by reduced 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 voucher also partially displaces the subsidy programme itself, reducing household AIP take-up by about 40 percent.

This interpretation differs from the common reading of recommendation-paired transfer experiments, beginning with Harou et al. (2022), in which productivity gains are taken as evidence that finance makes agronomic information actionable. Our results show why that inference can fail: a recommendation-paired transfer can raise output even when the recommendation itself is behaviourally inert, if farmers use the additional purchasing power on familiar rather than prescribed inputs. We cannot identify whether this reflects limited agro-dealer product depth, communication or trust, uncertainty about unfamiliar products,

[2]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.

[3]The participating agro-dealers belonged to the Farmers World network, one of Malawi's largest agricultural input retail chains. Farmers World operates more than 130 outlets nationwide, most of them located in the four study districts. It takes the average household in our sample about 2 hours to reach the nearest agro-input shop.

or other constraints. What we can show is that replacing AIP with a recommendation-paired voucher, without addressing these constraints, does not deliver precision nutrient management. The case for routing fiscal support through commercial agro-dealers may still rest on targeting (Basurto, Dupas, and Robinson, 2020) or market-shaping grounds, but not on the agronomic-precision rationale underpinning site-specific nutrient management and smart-subsidy narratives.

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework and derives the hypotheses tested in the paper. Section 3 situates the trial in Malawi's fertilizer policy landscape and describes the study setting. Section 4 describes the intervention and assignment, the sample, randomization, and baseline balance, and the empirical strategy. Section 5 presents the results on implementation, fertilizer and nutrient use, recommendation compliance, substitution with the national subsidy programme, yield, and economic returns. Section 6 discusses policy implications and concludes.

2 Conceptual framework

Matching fertilizer to soil is an allocation problem under two frictions. Soils are heterogeneous: plots differ in the nutrients they lack, so the value of any given fertilizer bundle varies from plot to plot. Fertilizer, however, is traded as a small menu of standardized products that bundle nutrients in fixed proportions, and in Malawi the dominant source of fertilizer support, the subsidy programme, distributes a single fixed package to all participants. Uniform products applied to heterogeneous soils misallocate nutrients: some plots receive nutrients they already hold in sufficient supply while their binding deficiencies go untreated.[4]

Site-specific recommendations aim to close this mismatch by converting a soil test into plot-level information. But a prescription can fail at several points before it reaches the plot:

$$\text{what the soil } \textit{needs} \longrightarrow \text{what is } \textit{recommended} \longrightarrow \text{what is } \textit{stocked} \longrightarrow \text{what the farmer } \textit{buys}.$$

Three gaps separate these four objects, each a distinct constraint with a distinct policy lever. The first lies between what the soil needs and what products can express: because the menu of blends is finite, even an ideal recommendation is an approximation. This is a manufacturing constraint. The second lies between what is recommended and what is stocked: the local agro-dealer may not carry the recommended products. This is a distribution constraint. The third lies between what is stocked and what the farmer buys: even with a suitable product on the shelf, the farmer may choose the familiar package. This is a demand constraint. The premise under test is that the first two constraints leave sufficient scope for adjustment, so that the remaining demand-side gap can be closed by information and the means to act on it. The rest of this section formalizes the chain and derives the predictions we take to the data.

2.1 Constraints to precision fertilization

Consider a household cultivating a plot of soil type θ, with a nutrient and amendment requirement $n(\theta) \in \mathbb{R}^K_+$ over K nutrients (nitrogen, phosphorus, potassium, and amendments such as lime and micronutrients). Fertilizer is sold as products $j = 1, \ldots, J$ (urea, NPK blends, CAN, and so on), each delivering nutrients in fixed proportions that the farmer cannot alter, summarized by a characteristics vector $a_j \in \mathbb{R}^K_+$ giving nutrient content per kilogram (Rosen, 1974; Lancaster, 1966). From these products the household assembles a purchase bundle b, a vector of product quantities x_{bj}, which delivers total nutrients $z_b = \sum_j a_j x_{bj}$ at money cost c_b and a non-price cost k_b that collects search effort, indivisibility of bag sizes, unfamiliarity, application knowledge, and trust in the product.

Three nested objects define the three gaps. Let z_r be the nutrients delivered by the soil-test recommendation $r(\theta)$, the best product bundle for the plot that can be assembled from the blends industry manufactures; let $z_{b^\dagger}$ be the nutrients of $b^\dagger$, the closest bundle available in the local stocked menu $\mathcal{B}$; and let $z_{b^\star}$ be the nutrients the farmer actually applies, chosen in the next subsection. The shortfall between what the soil

[4] Our soil-test data show that plots vary widely in their nutrient requirements and that the standard subsidy blend is poorly matched to them: the median plot needs roughly as much potassium as nitrogen, whereas the subsidy blend supplies about one-fifth as much potassium as nitrogen, and deficiencies in sulphur, zinc, and boron, which the standard package does not carry, are widespread.

needs and what the farmer applies then decomposes into:

$$n(\theta) - z_{b^\star} = \underbrace{\big[\, n(\theta) - z_r \,\big]}_{\text{Gap 1}} + \underbrace{\big[\, z_r - z_{b^\dagger} \,\big]}_{\text{Gap 2}} + \underbrace{\big[\, z_{b^\dagger} - z_{b^\star} \,\big]}_{\text{Gap 3}}. \tag{1}$$

Gap 1 is the residual left because the recommendation is bounded by the blends that exist; Gap 2 opens when the recommendation is not stocked, so $r(\theta) \notin \mathcal{B}$ and the best buyable bundle $b^\dagger$ falls short of it; and Gap 3 is the distance between that best buyable bundle and the farmer's actual choice. Precision fertilization requires all three brackets to be small. A reform that targets only one of them leaves the rest of the shortfall open; information and finance address at most the final step unless recommended products are also manufactured, stocked, and behaviourally acceptable. Empirically, we observe compliance on two margins. The product-set margin asks whether the applied products match the recommended products; the nutrient margin asks whether the applied bundle delivers the recommended nutrient totals, whichever products carry them. Because products are fixed-proportion nutrient carriers, the two margins can move independently: a farmer who substitutes urea for recommended CAN fails on products yet can still meet, or overshoot, the nitrogen total, while a farmer who buys the recommended blend in the wrong quantity complies on products but not on nutrients. The product margin therefore tests whether the site-specific composition of the recommendation is adopted; the nutrient margin tests whether the prescribed rates are attained. The decomposition also warns that a null on precision adoption need not be a pure demand result: it can come from any of the three brackets. The experiment observes the two ends of the chain, the recommendation and the purchase, but not the stocked menu in between, so without agro-dealer stocking data it cannot fully separate a recommended product that was unavailable (Gap 2) from one that was available but not chosen (Gap 3).

2.2 The farmer's bundle choice

Given the stocked menu, the household chooses the bundle that maximises expected plot profit net of money and non-price costs:

$$b^\star = \arg\max_{b \in \mathcal{B}} \; \{\, q\, \tilde{f}(z_b; \theta, s) - c_b - k_b \,\} \qquad \text{s.t.} \qquad c_b \le m + T, \tag{2}$$

where q is the output price, $\tilde{f}$ is subjective expected yield given beliefs s, m is own resources available for fertilizer, and T is the value of any transfer. The soil-test recommendation enters through beliefs s, raising the perceived return to the recommended bundle; the voucher enters through T, relaxing the budget; the retail environment and the farmer's experience enter through the menu $\mathcal{B}$ and the non-price costs k_b. In this vocabulary, the AIP package and the soil-test prescription are particular bundles: the familiar bundle b_0, the combination of NPK basal blend and urea that two decades of subsidy have made the default, and the recommendation $r(\theta)$.

We treat m as predetermined at the point of the bundle choice: the household has already decided how much of its own resources to devote to fertilizer, so the recommendation acts on beliefs and the voucher on the budget constraint, and neither acts on that prior allocation. This is a two-step decision, and a farmer could in principle revise m on receiving a recommendation. The advice-only arm is where such a revision would appear, and it does not: T1 raises own input spending on the sampled plot by about 15,000 MWK against a control mean of 222,000 MWK, and total fertilizer applied by 2.2 kg against a control mean of 78 kg, neither distinguishable from zero (Tables 3 and 10). The standard errors are wide enough to leave room for a modest reallocation, so we treat the assumption as maintained rather than tested.

Comparing the buyable recommendation $b^\dagger$ with the familiar bundle b_0, the farmer applies the recommendation rather than the familiar bundle only if

$$q\big[\tilde{f}(z_{b^\dagger}; \theta, s) - \tilde{f}(z_{b_0}; \theta, s)\big] \;\ge\; (c_{b^\dagger} - c_{b_0}) + (k_{b^\dagger} - k_{b_0}). \tag{3}$$

The left-hand side is the perceived extra return to complying; the right-hand side is the extra money cost plus the extra non-price cost relative to the familiar bundle. Because b_0 is stocked, divisible, and well understood, its non-price cost k_{b_0} is low, so switching requires the recommendation's perceived advantage to clear the extra cost. This is Gap 3: even a recommendation that is agronomically correct and physically available is not adopted if the prescribed products are dearer, sold only in inconvenient quantities, unfamiliar, or not trusted.

2.3 The subsidy, the voucher, and the composition margin

The policy environment gives the condition in Equation (3) a concrete, testable form. The AIP subsidy is a *restricted in-kind transfer*: at a given fiscal cost it fixes both the amount of support and its composition, delivering the familiar bundle b_0 and nothing else. A long literature on in-kind versus fungible transfers holds that such a composition restriction changes behaviour only when it *binds*, that is, only when the household would otherwise have chosen a different bundle; an inframarginal recipient treats the restricted transfer as equivalent to fungible cash of the same value and simply displaces own spending (Ricker-Gilbert, Jayne, and Chirwa, 2011; Hoynes and Schanzenbach, 2009; Moffitt, 1989; Southworth, 1945).

The voucher relaxes this restriction experimentally: it carries a value T calibrated to the AIP subsidy but can be spent on any bundle in the stocked menu $\mathcal{B}$. Conditional on receiving the recommendation and AIP-equivalent purchasing power, we can therefore observe whether households use flexible support to move away from the familiar AIP bundle toward the prescribed product mix.[5] Consider a household for whom the recommended products are stocked ($b^\dagger = r(\theta)$). If it still buys the standard bundle b_0, then the recommendation and the flexible transfer did not induce movement away from the familiar product mix, and condition (3) fails despite relaxed purchasing power. If instead it switches to the recommendation, information and finance are complements on the product-composition margin. The pairing of the voucher with the recommendation is central to this interpretation: a positive yield effect among farmers given the voucher is evidence that purchasing power increased productive input use, but it is evidence that the recommendation became actionable only if the added bundle moves toward the prescribed products or nutrient totals.

2.4 Empirical predictions: which constraint binds

The premise under test says information and liquidity bind and supply does not. Exogenous variation in the information farmers hold and in their purchasing power, together with the outcome sequence, lets the data adjudicate, and the three predictions below are best read as competing accounts of where the chain breaks.

H1 Information binds: If farmers simply lack credible plot-specific information, then delivering the recommendation shifts beliefs s enough to satisfy (3) on its own. Farmers who receive the recommendation, but no relaxation of their budget, then move products, nutrient application, and yields toward it.

H2 Liquidity binds: If farmers accept the recommendation but are liquidity-constrained ($c_r > m$), relaxing the budget in (2) makes the prescription actionable. Farmers who are informed *and* no longer liquidity-constrained then comply with the recommended products more than those who are informed alone, with matching gains in nutrient alignment and output.

H3 Product-choice frictions bind: If the recommended products are not manufactured in a matching form, are unstocked ($r(\theta) \notin \mathcal{B}$), or carry high non-price costs, then relaxing the budget raises total fertilizer use without raising compliance: the added purchasing power is spent on the familiar bundle b_0. Output can still rise, but the gain is a return to additional conventional fertilizer, not to precision nutrient management. This is the alternative to the premise, in which the binding constraint lies after recommendation delivery, either in local product availability or in farmer willingness to switch products.

A further question: additionality. Beyond which constraint binds, the voucher's effect on the quantity of fertilizer raises a distinct question: how much of the transfer finances fertilizer the household would not otherwise have bought, and how much displaces purchases it would have made anyway, commercially or through the subsidy. If the transfer is additional, total fertilizer use rises by the full voucher-financed quantity and own input spending is unchanged; if it is inframarginal, use rises by less and own spending falls, so a household can gain, and its input costs fall, while little new output is created. This is not a further

[5] AIP is rationed to a minority of households each season, so for most recipients the voucher is a fungible transfer of AIP-equivalent value rather than the flexible counterpart of a package they would actually have received. The experiment therefore does not compare a fixed AIP entitlement with a flexible AIP entitlement for the same household; it identifies the composition margin generated by flexible fertilizer purchasing power when paired with the soil-test recommendation.

account of where the chain breaks, which is why we do not number it among the hypotheses; it bears on the programme's cost-effectiveness, and we take it up in the analysis of economic returns.

3 Context

3.1 Fertilizer policy in Malawi

Public fertilizer subsidy in Malawi predates independence and has run in some form for over seventy years, beginning with subsidized 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 (Benson et al., 2024; Lunduka, Ricker-Gilbert, and Fisher, 2013; Harrigan, 2008). 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 (Nyondo et al., 2021; Khonje et al., 2022; Jayne et al., 2018; Lunduka, Ricker-Gilbert, and Fisher, 2013). The agronomic blueprint inherited from this period was embedded in the Affordable Inputs Programme (AIP), which ran from 2020 to 2025.

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 subsidized fertilizer (Nyondo et al., 2021; Lunduka, Ricker-Gilbert, and Fisher, 2013), and evidence from Zambia finds the response collapsing toward zero once weeding and soil-carbon thresholds are not met (Burke, Jayne, and Black, 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 (Jayne et al., 2018; Ricker-Gilbert, Jayne, and Chirwa, 2011). 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, 2015), and within villages most allocation is done de-facto by chiefs, who use local information on shocks and returns but cover 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., 2021). The system is also exposed to macroeconomic and fiscal stress. The 2022–24 foreign exchange shortage 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 (World Bank, 2025; Office of the Ombudsman of Malawi, 2024; Duchoslav and De Weerdt, 2023). 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., 2021).

3.2 Study site

The study was conducted in 2024/25 with the participation of 2,034 maize-cultivating farm-households located in 113 villages across four districts in central Malawi. These households are small-scale, resource-

poor farming families who grow maize as their staple food crop on small, rain-fed holdings, cultivating about 2.5 acres in total and roughly one acre of maize at the median. They rely mainly on family labour and hand tools, sell only a modest share of what they harvest, and look to the national fertilizer subsidy as their principal source of subsidized inputs, though the programme is rationed and reaches fewer than one household in five in a given season. Most are male-headed (about three-quarters), with a household head in their mid-forties. The districts in which the study was conducted (Kasungu, Dowa, Ntchisi, and Mchinji) lie in Malawi's central-region maize belt and are broadly characterized by a single, unimodal rainy season running from roughly November to April, a maize-dominated farming system commonly intercropped or rotated with legumes such as groundnuts and pigeonpea, with soya, grown largely for sale, having overtaken tobacco as the principal cash crop, and soils that, as our own soil-test data confirm, vary widely in fertility and are frequently deficient in potassium, sulphur, and micronutrients and, in places, acidic.

The intervention was implemented in cooperation with Meridian Group, a private company that operates in Malawi's agricultural inputs and commodity sectors.[6] Soil tests were conducted on participating farmers' plots by Meridian's Farm Services Unit (FSU) at no cost to the household, recommendations were provided to farmers by trained extension agents ("agronauts") attached to the FSU, and fertilizer vouchers were redeemable solely at Meridian's Farmers World shops in the study area.[7]

4 Study design and empirical strategy

The trial is registered in the American Economic Association RCT Registry as AEARCTR-0014293.[8] The pre-analysis plan was filed in October 2024, before any endline data were collected, and is available at the registry.

4.1 Experimental design and treatments

The study is designed as a cluster randomized controlled trial in which the village is the unit of randomization. The 113 study villages, drawn from the sampling frame of an earlier cluster-randomized study in the same four districts (De Weerdt et al., 2024), were assigned to the three arms by computer-generated random assignment; the sampling frame, baseline balance, and sample flow are reported in Section 4.2 below. Villages were assigned to one of the following three groups:

- **Treatment 1 (T1)** households receive a soil-test report and plot-specific fertilizer recommendation for a farmer-selected plot and intended crop.
- **Treatment 2 (T2)** households receive the same soil-test report and plot-specific fertilizer recommendation as T1 households, plus a voucher redeemable for any fertilizer product stocked at the nearest designated Farmers World shop, with face value equal to the implicit subsidy under the AIP package.
- **Control (C)** households do not receive the soil-test report, plot-specific fertilizer recommendation, or voucher during the study period.

Farmers in all three arms (including the control arm) were enrolled in the study during a visit by an enumerator trained by the study team and accompanied by an agronaut, who obtained each household's informed consent before any data were collected. Once enrolled, the farmer was offered a soil test on a plot of their own choosing, and asked which crop they intend to grow on that plot. All farmers were told at the time of enrollment that their sample would be analyzed and that they would receive a soil-test report

[6] The Meridian Group is headquartered in Mauritius and specializes in the manufacturing and distribution of fertilizers and other commodities in Malawi, Mozambique, Zambia and Zimbabwe. Meridian was acquired in 2019 by Maaden, a global mining company headquartered in Saudi Arabia. The analysis, results, and interpretation of results, and policy recommendations presented in this paper were in no way influenced by any of the companies listed above.

[7] The Meridian Group owns and operates both the Farmers World retail shops and the Farm Services Unit used in this study, and employs the Agronauts who participated in this study. The FSU's 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 World has more than 130 retail stores across central and northern Malawi that sell fertilizers and other inputs as well as other farm and household items.

[8] https://www.socialscienceregistry.org/trials/14293

and recommendations based on it. They were also told that the testing laboratory operated under capacity constraints: the study team would return results as quickly as possible, but limited processing throughput meant that not every farmer could be served before planting, and some would receive their report only after the season, when it would still be useful for fertilizer decisions in subsequent seasons. Because this early-season processing capacity was genuinely scarce, we rationed it by random assignment, and it is this rationing that generates the information component of the experiment: T1 and T2 farmers receive the report roughly one month after enrollment, prior to the season's planting period, while control farmers receive theirs in the following season, after harvest. No farmer is denied the recommendation; randomizing the order of delivery is both a fair way to allocate a scarce service and the source of the exogenous variation in early access to the recommendation.[9]

The soil samples were analyzed and soil reports generated for every household.[10] The report was delivered to farmers as a two-page document. The first page provides 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 50-kg bags/acre), and timing for the planting and top-dressing stages, together with short action comments. The report is written in Chichewa throughout; English is used only for nutrient and product names, units, and dates. An English translation of the example report is given in the caption to Figure 1. The second page reports the underlying soil-test results: measured values and units for pH, available P, exchangeable K, calcium (Ca), magnesium (Mg), iron (Fe), manganese (Mn), boron (B), copper (Cu), zinc (Zn), cation exchange capacity, total nitrogen (N), organic matter, and the carbon-to-nitrogen (C:N) ratio. Each parameter is also assigned a status on a four-point scale—Very Low, Low, Adequate, High—rendered as a colour-coded bar (red, orange, green, blue) with accompanying emoticons. An example of a report is shown in Figure 1.

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 in the shop. "Nearest" was operationalized using OpenRouteService travel-time routing under a motorbike-or-bicycle profile, applied to each sample village and the universe of Farmers World shop locations.[11] 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 (Figure 2). Vouchers were distributed together with the soil-test report in September–October 2024, ahead of the season's planting window, and could be redeemed over the course of the 2024/25 season; redemption was therefore feasible in time for both basal and top-dress applications.

4.2 Sample, randomization, and balance

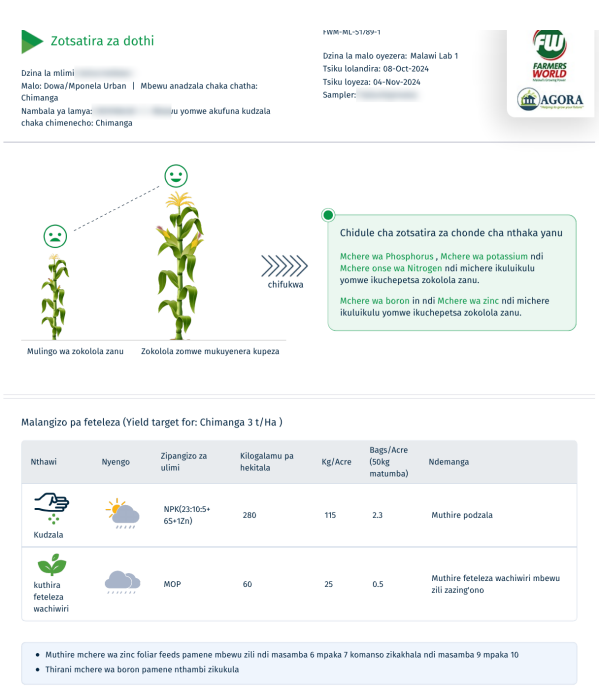
Sampling frame and design. The sampling frame is inherited from a previous cluster-randomized study on smallholder market participation (De Weerdt et al., 2024), which delimited 113 villages across four districts in Central Malawi. The number of households per village and the allocation of villages across arms were chosen by power simulation to detect a 13 percent yield effect (about 135 kg/acre).[12] The final design assigns roughly 30 percent of villages to control and splits the remainder approximately equally between T1 and T2, enrolling 18 households in each of the 113 villages, for a total of 2,034 households in the baseline sampling frame. The T1-vs-T2 contrast is substantially less well powered than the arm-vs-control comparisons: 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.

[9] Unfortunately, during the course of the study, a major donor pulled out, and as a result, the implementing partner did not deliver on this promise; control farmers never received their recommendations.

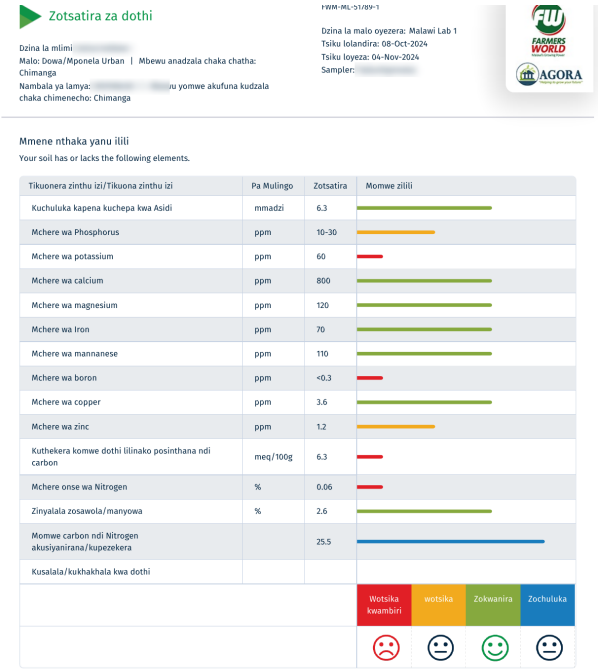
[10] We used the Crop Nutrition Laboratory Services Ltd. (Cropnuts), an agricultural testing laboratory & agronomy advisory services company located in East Africa. The model underlying the spectral test was calibrated on wet-chemistry analyses of 22,000 samples from across Malawi.

[11] https://openrouteservice.org/

[12] Drawing yield observations from the De Weerdt et al. (2024) farmer panel (mean 1,042 kg/acre, SD 463 kg/acre), 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 yields 85–90 percent power at the 5 percent level for both the T1-vs-control and T2-vs-control contrasts. A fuller account of the simulation, including the grid of configurations considered and the basis for the assumed effect size, is given in the registered pre-analysis plan (AEARCTR-0014293).



(a) Front page: yield gap and crop-specific recommendation



(b) Back page: measured soil-test results and status

Figure 1: Example two-page soil-test report delivered to farmers. The report is written in Chichewa throughout; English appears only for nutrient and product names, units, and dates. The front page (a) shows the household's estimated current yield against a target yield, a plain-language fertility summary, and a crop-specific fertilizer recommendation table. In this example the summary reads, in translation, "phosphorus, potassium and total nitrogen are the main nutrients limiting your yield" and "boron and zinc are also main nutrients limiting your yield". The recommendation table is headed "fertilizer recommendations (yield target: maize, 3 t/ha)" and carries columns for stage, weather, input, kg/ha, kg/acre, 50-kg bags/acre, and comment. It prescribes NPK 23:10:5+6S+1Zn at 280 kg/ha (115 kg/acre; 2.3 bags/acre) at planting, with the comment "apply at planting", and muriate of potash (MOP) at 60 kg/ha (25 kg/acre; 0.5 bags/acre) at top-dressing, with the comment "apply the second fertilizer while the plants are still small". Two further instructions follow the table: "apply zinc foliar feed when the plants have six to seven leaves, and again at nine to ten leaves", and "apply boron as the branches develop". The back page (b), table headed "the condition of your soil", reports the measured soil parameters with their units and values alongside a colour-coded four-point status scale whose labels read, in translation, very low, low, adequate, and high.



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 the recipient's name (redacted here) and a unique redemption identifier, QR-coded for point-of-sale scanning.

Randomization and timeline. Randomization was at the village level, with the number of villages per arm fixed in advance: 34 villages were allocated to control, 40 to T1, and 39 to T2 by randomly permuting that allocation across the 113 study villages. Assignment was not stratified. This yields 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 agricultural 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 C and T2 vs C) do not reject the null of orthogonality. While these covariates were measured in 2022, two years before treatment delivery, they are slow-moving household characteristics unaffected by later assignment, so balance on them speaks to the validity of the randomization despite the gap between frame construction and enrollment. These five are the household characteristics carried by the source study's pre-treatment round. Beyond them, and not pre-specified, we test balance on village-level access to fertilizer retail, measured as the straight-line distance from each village to the agro-dealer it was matched to: control villages average 9.71 km (standard deviation 4.72), and the differences are -0.27 km for T1 and 0.32 km for T2, with standard errors of 1.21 in both cases and a joint test that does not reject orthogonality ($p = 0.88$). We report this separately from Table 1 because it is measured at the village rather than the household level. Fertilizer use and participation in the subsidy programme, which would be natural complements, are recorded at endline and are themselves affected by treatment, so they cannot serve as baseline covariates.

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 (acres)	2.476	0.023	0.040	2029
	(1.546)	(0.129)	(0.141)	
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: The first column reports control group means (with standard deviations in parentheses); the second and third columns report differences between each treatment arm and the control group (with standard errors clustered at the village level in parentheses). Land area is measured in acres 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 A.1 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. The lower linkage rate in the control group is immaterial: the matched data are used only for the recommendation-compliance and application-error outcomes, which are defined for treated households and estimated as T2–T1 comparisons, so control households never enter these estimates. All other outcomes are 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 A.1). 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 enrollment, for example.

Regression-based attrition tests reported in Appendix Table A.2 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 A.3 show little evidence of selection on baseline observables: nine of the ten treatment-control differences are statistically insignificant, the exception being that retained T2 households are slightly larger (0.31 members, $p = 0.03$), about what chance would produce across ten tests.

4.3 Empirical strategy

The randomized assignment allows for causal identification of the intent-to-treat (ITT) effects of both treatments relative to the control. The design supports this interpretation at the plot level as well: because farmers chose the soil-test plot and its intended crop under a common information set, before any arm

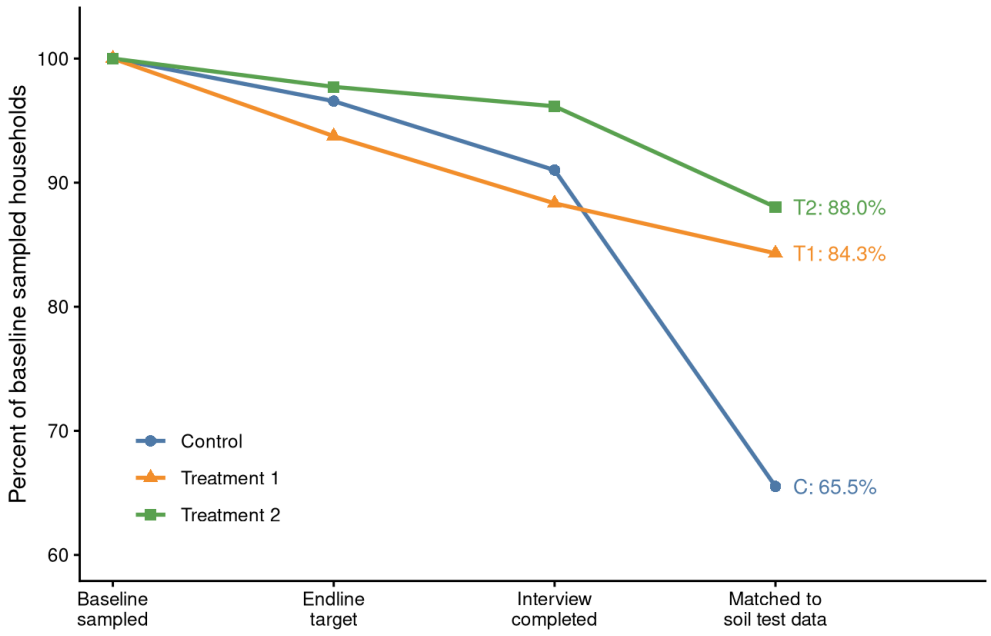


Figure 3: Sample retention by treatment arm and survey stage. Each stage is expressed as a percentage of the original sampled baseline frame in that treatment arm, keeping the denominator fixed across stages so that differences in realized retention are directly comparable across arms. Baseline sampled 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. The linkage stage is not used as the analysis denominator for endline-survey outcomes; compliance and application-error outcomes use treated households with valid recommendation records and compare T2 with T1, so the lower control linkage rate is not part of those estimands.

learned its assignment, that plot is balanced in expectation between treatment and control, so plot-level comparisons retain a clean ITT interpretation while preserving farmer agency over plot and crop choice.

Fertilizer-use outcomes are observed at the plot level. Data was collected for two plots per household: the soil-test plot and one non-test plot randomly selected from the list of remaining plots provided by each study participant. The random plot captures reallocation effects, whether treatment shifts or adds fertilizer, including AIP fertilizer, on plots beyond the test plot, and permits household-level totals that sum fertilizer use on both surveyed plots.

We estimate these ITT effects for a set of outcomes using ordinary least squares (OLS) in the following form:

$$Y_{iv} = \alpha + \beta_1 T1_v + \beta_2 T2_v + \varepsilon_{iv} \tag{4}$$

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 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, nutrient-use, non-test-plot reallocation, yield, and economic-return tables, and for the AIP-substitution and secondary-outcome tables in the appendix. Standard errors are clustered at the village level, the level of random assignment.

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. For household i in village v in this restricted sample, we estimate

$$C_{iv} = \gamma_0 + \gamma_1 T2_v + u_{iv}, \tag{5}$$

where C_{iv} is the compliance or application-error outcome, $T2_v$ indicates assignment to the voucher arm with T1 the omitted category, and γ_1 is the incremental effect of the voucher conditional on access to the recommendation package. Standard errors are clustered at the village level.

The analysis is organized by *outcome family* rather than by a single aggregate index. An outcome family is a group of outcomes that bear on the same link in the causal chain, and it is the unit within which we correct for multiple hypothesis testing: the adjustment is applied across the outcomes within a family rather than across every outcome in the paper. We use seven families—delivery and adherence, fertilizer and nutrient use, product compliance, nutrient compliance, AIP substitution, production and economic outcomes, and secondary soil-management practices—which 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. The pre-analysis plan specified summary indices to address multiple hypothesis testing within these families; because our contribution is the disaggregated causal chain rather than a family-average effect, we instead retain the individual outcomes and report sharpened two-stage false-discovery-rate (q) values (Anderson, 2008) within each family, alongside the raw p-values, in Appendix Table A.7. The AIP-substitution family was not pre-specified and is included for completeness. Every outcome underlying the paper's conclusions survives at $q < 0.05$, most at $q < 0.001$.

5 Results

5.1 Uptake and treatment adherence

Before reporting results related to fertilizer use, we first document treatment fidelity. This is necessary for interpreting results for both the recommendation-only arm (T1) and for contrasting it with the voucher arm (T2). A null effect in T1 could suggest 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 fertilizer availability, affordability, or access constraints. These cases have different yet 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 shops, visited the household to provide further extension advice on crops grown in

the 2024/25 season. Agronaut visits occurred at 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. This indicator captures in-season advisory visits and is distinct from the enrollment and soil-sampling visit, which agronaut-accompanied teams made in all three arms, and from delivery of the report itself; recommendation receipt can therefore exceed the visit rate. The nonzero control rate reflects the agronauts' routine commercial extension presence in the study area.

Receipt of recommendations is also measured based on self-reports from respondents. After being shown an example recommendation sheet on the enumerator's 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 does not survive adjustment for multiple hypothesis testing (raw $p = 0.10$, sharpened $q = 0.24$; Appendix Table A.7).

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. A more likely explanation is that acting on the advice, rather than the advice itself, raised the reported ease: T2 households redeemed the voucher and bought fertilizer, and having acted they may rate the recommendation as clearer, whether because the purchase built familiarity, because they rationalize a choice already made, or simply because acting on advice makes people call it easy to understand.

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 (Section 5.3), observed fertilizer choices indicate substantially lower compliance with the specific products recommended by the soil tests. The endline instrument recorded whether farmers found the recommendation easy to understand and whether they believed they had followed it; it did not ask them to state what had been recommended, so we have no objective measure of comprehension. The gap between reported adherence and observed compliance therefore admits more than one reading. Farmers may have understood the recommendation and departed from it for the reasons examined in Section 5.3, or they may have believed they had understood it when they had not, in which case they would report both that the advice was easy to follow and that they had followed it while applying something other than what was prescribed. The data cannot separate these, and doing so would require a comprehension measure the survey does not contain.

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 package.

These first-stage results rule out a simple non-receipt explanation for weak T1 effects. 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 points to affordability as a binding margin even when advice reaches farmers.

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.

5.2 Fertilizer and nutrient use

Next, we examine treatment effects on fertilizer use on the soil-test plot. For each household, the endline survey recorded which fertilizer types were applied to that plot during the 2024/25 season and the quantity (kg) of each, with non-users coded as zero. Table 3 reports ITT estimates for three outcomes: a binary indicator for any fertilizer use (Panel A); total kilograms applied, summed across fertilizer types (Panel B); and kilograms per acre, the total divided by self-reported plot size (Panel C). Each panel is estimated first on all soil-test plots and then on the subset whose main crop was 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 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.

Our results indicate that T1 does not generate a detectable change in any fertilizer outcome. However, T2 raises the probability of using any fertilizer by 25 percentage points across all soil-test plots and by 17 percentage points on maize plots, both highly significant (Table 3, Panel A). Because control households already used fertilizer at 70 percent across all soil-test plots and 82 percent on maize plots, these increases lift T2 use to about 95 and 99 percent respectively, at the voucher-redemption ceiling reported above (98.6 percent conditional on receipt): essentially every household that redeemed applied fertilizer to the maize plot the soil test targeted. T2 specifically increases total fertilizer applied by 30.5 kg (about 39 percent of the control mean) across all soil-test plots and by 28.6 kg (about 34 percent) 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 (about 60 and 57 percent of the control mean). Equality tests reject T1 = T2 in both samples across all panels. In short, 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. In the terms of Section 2, this is evidence against H1: information was not the binding constraint in this setting.

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 (about 51 percent of the control mean) and 18.7 kg more NPK 23:10:5 (about 61 percent) 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

package (one bag NPK basal blend, one bag urea top-dressing). Despite the voucher's full fungibility across products and bag sizes, T2 households simply reproduce the AIP product mix, 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 kg applied per household on the soil-test plot by treatment arm. This table is descriptive: it reports arm means and the raw T2–control difference by product. ITT estimates for total fertilizer use and nutrient application are reported in Tables 3 and 5. "Potassium product" combines muriate of potash (MOP) and sulphate of potash (SOP, also recorded as potassium sulphate). "Lime" combines calcitic and dolomitic lime. CAN denotes calcium ammonium nitrate, MAP mono-ammonium phosphate, and DAP di-ammonium phosphate.

Table 5 expresses the maize-plot response in nutrient terms, and its striking feature is the asymmetry across nutrients. The T2 increment is almost entirely nitrogen, about 10.4 kg/acre more N, against small phosphorus gains and a near-zero potassium response, for a total increase of roughly 12.9 kg/acre in macronutrients; T1 again shows no detectable effect, and the equality tests confirm that T2 differs from T1 on each nutrient. This pattern is not an agronomic choice but a mechanical consequence of what the voucher bought: as Table 4 shows, the added fertilizer is overwhelmingly urea (46% N) and NPK 23:10:5 (23% N), both high-nitrogen products. The phosphorus gain is a by-product of the NPK blend rather than a response to soil P deficiencies, and the near-zero potassium response, in a sample where the soil tests flag widespread K deficiency, reflects the complete absence of potassium purchases. This decomposition matters for the yield results that follow: the T2 gain comes from adding more nitrogen to a bundle already dominated by it. To the extent that phosphorus and potassium are co-limiting with nitrogen, as the soil-test evidence suggests, the observed yield gain is only a fraction of what applying the full recommended bundle could deliver, and the treatment effects should be read as the returns to additional conventional fertilizer rather than to precision nutrient management.

A natural concern is that the 29–31 kg test-plot increase in Panel B of Table 3 may reflect reallocation from other household plots rather than a genuine increase in fertilizer use. The randomly selected non-test plot introduced in Section 4.3, for which the endline survey recorded the same fertilizer and area information as for the test plot, addresses this directly. Households that 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.[13] Of these, only 144 have a randomly-selected non-test plot

[13] 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

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 (kg/acre)	18.41	0.64	10.41***	0.000	1625
		(1.97)	(1.19)		
Phosphorus (kg/acre)	2.50	0.13	1.63***	0.000	1625
		(0.20)	(0.18)		
Potassium (kg/acre)	1.26	0.08	0.83***	0.000	1625
		(0.10)	(0.09)		
Total nutrients (kg/acre)	22.17	0.85	12.87***	0.000	1625
		(2.12)	(1.43)		

Notes: This table reports ITT effects on nutrient application in kg/acre 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 in acres. 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.

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 change 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 −18.8 kg (SE = 16.7, $p = 0.26$) and the T1 estimate is −21.2 kg (SE = 15.6, $p = 0.18$). The kg/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 increase on other plots in the household's portfolio.

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: 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.

Having established that the voucher-induced fertilizer increase is not drawn from the household's other plots (Table 6), we turn to a different offset margin: the subsidy programme itself. If households in the voucher arm respond to the transfer by forgoing AIP fertilizer they would otherwise have obtained, part of their increased fertilizer use replaces subsidized fertilizer rather than adding to it, and the net gain in household fertilizer access is smaller than the gross figure. Our results (detailed in Appendix 3) show that T2 reduces the probability of any AIP take-up by 7.7 percentage points, about 40 percent of the control-group rate of 19 percent, and reduces AIP fertilizer on the soil-test plot by 7.1 kg, about 74 percent of the control mean. This indicates partial substitution, consistent with two mechanisms: *demand-side substitution*, in which households that have already secured fertilizer through the voucher are less likely to pursue or pay the co-payment for an additional subsidized package, and *supply-side re-targeting*, in which AIP committees or village leaders deprioritise households known to have received voucher support (Basurto, Dupas, and Robinson, 2020). The data do not fully separate these channels, but do indicate that substitution effects are at play in our setting; we return to the design implications in the concluding section.

The adjustment to the voucher extends beyond inorganic fertilizer. Appendix Table A.4 reports effects on broader soil nutrient management practices. T1 is broadly null. T2 significantly reduces use of fresh vegetative material and dairy or poultry manure by 7–10 percentage points and increases adoption of ridges or check dams (+9.0 pp); these effects survive the multiple-testing adjustment in Appendix Table A.7. Use of *mbeya* fertilizer, a hybrid compost that blends locally available organic material with a small quantity of chemical fertilizer, also declines, but that reduction is significant only on the unadjusted p-value. The reductions in organic inputs are consistent with rational substitution: collecting and applying organic material is labour-intensive, and when inorganic 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 inorganic fertilizer applied. This substitution carries a potential long-run soil-health cost, since sustained reductions in organic-matter inputs deplete soil carbon and moisture retention, and neither mineral nor organic inputs alone can sustain productivity and soil health in smallholder systems (Vanlauwe et al., 2026, 2015); we return to this in the conclusion.

5.3 Recommendation compliance

Three facts from the preceding tables set up the compliance question: T2 farmers self-report higher adherence (Table 2); they apply about 31.6 kg per acre more fertilizer, some 60 percent of the per-acre control mean (Table 3); and that increase is concentrated almost entirely in urea and NPK 23:10:5, the two products that constitute the AIP package, while prescribed products outside it are not bought in greater quantities (Table 4). 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 package regardless of what the recommendation prescribed.[14,15]

Unlike the arm-specific means in Table 2, the compliance tables (Tables 7 and 8) report the T2–T1 contrast, since the control group never received the recommendation. The sample is thus T1 and T2 households that report receiving a recommendation, hold a valid recommendation record naming at least one product, and have usable fertilizer-application data.

Up to eight products were recommended to T1 and T2 farmers: NPK 23:10:5, urea, CAN, NPK 14:14:20, NPK 8:18:15, potassium products, lime, and MAP. "Potassium products" here means straight potassium fertilizers—muriate of potash (potassium chloride) and sulphate of potash (potassium sulphate)—rather than potassium-bearing compound blends or foliar formulations, and they are typically prescribed as a top-dressing in place of urea. 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-package outcome is the strictest product-set measure: it equals

[14] We use the term *compliance* descriptively, as the degree to which the products and nutrients a farmer applies match those the soil-test recommendation prescribes, and not as a judgment about the farmer or the recommendation.

[15] A compliance shortfall is consistent with the recommended products being unstocked at the agro-dealer as much as with farmers declining to buy them; as our conceptual framework makes clear, the design cannot separate these, so low compliance should not be read as farmer error.

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, about 78 percent of this sample, often in place of the more familiar urea top-dressing, yet potassium compliance is essentially zero and is *lower* in T2 than in T1.

The voucher does not, however, move farmers toward the full prescribed package: T2 does not increase the probability of applying all recommended products, and exact product-package 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 package as the reference point for input decisions, would generate exactly this observation: increasing purchasing power without changing the reference package raises the probability that farmers execute the familiar plan at larger scale, mechanically increasing the applied share of familiar products while reducing exact-package compliance on the prescription's non-standard components. But the same observation can also result from supply absence (e.g., if K, lime, MAP, or CAN were not in stock at the time the participant redeemed their voucher at a 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 K and lime, and by a mechanical penalty in the exact-package 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 a median of 0.50 both for the 427 households whose recommendation does not list urea and 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, $p = 0.49$). In sum, the voucher arm applies the AIP 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 package, 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. By design, recommended nutrient quantities are balanced across treatment arms. The voucher arm applies substantially more nutrients: actual N application is about 11.2 kg/acre 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 3.5 kg/acre and N shortfall falls by about 7.0 kg/acre in T2 relative to T1. The shortfall falls by more than the absolute error, suggesting 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 final row of Panel A shows how many farmers cross this line: 31 percent of T1 households apply more nitrogen than their plot-specific recommendation, and the voucher raises this share by 30 percentage points, so that a majority of T2 households over-apply. Mean actual application in T2 (27.6 kg/acre) in fact exceeds the mean recommended rate (24.6 kg/acre). The voucher thus acts as a quantity instrument rather than a calibration instrument: it shifts the nitrogen distribution upward past the target instead of concentrating application at the recommended rate, consistent with farmers spending a fixed voucher value on the familiar nitrogen-heavy products documented above. The same pattern does

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 estimates of T2 minus T1. Product recommendations are parsed from product-specific recommendation rates for the eight products or product groups listed in the text above. 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-package outcome equals one only if the applied product set matches the recommended product set within the eight parsed products. 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.

not appear for P_2O_5 or K_2O: fewer than 5 percent of households over-apply either nutrient in either arm, and the voucher leaves these shares unchanged. For these nutrients, recommendation levels remain far above actual application in both treatment arms, so the voucher-induced change in fertilizer use operates almost entirely on the nitrogen component of the recommendation, and there it overshoots for most households. Substantial under-application nonetheless persists alongside the overshoot: because the shortfall measure is zero for over-appliers, the mean T2 shortfall of 4.4 kg/acre is carried entirely by the 39 percent of households still below their recommendation, who fall short by about 11 kg/acre on average; phosphorus and potassium shortfalls are of similar or larger magnitude. The voucher therefore does not close the application-error gap so much as change its nature, converting a one-sided problem of under-application into a two-sided problem of dispersion around the 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 class establishes that compliance is achievable when the recommendation and product familiarity coincide. Second, CAN as a top-dressing 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 favour 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, which is a point we will return to later.

One direct, if coarse, piece of evidence speaks to the availability channel. Among voucher recipients who had received a recommendation and redeemed, 95.7 percent reported that the recommended fertilizer was available at the redeeming shop, and the small minority who reported it unavailable almost all bought a different product instead. Taken at face value, a blanket stock-out is therefore not what drives the aggregate compliance failure: most voucher households could obtain the recommendation and nonetheless purchased the familiar urea and NPK 23:10:5 package, which points to a demand-side rather than a supply-side account of the composition result. Two qualifications stop this from settling the question. First, the question refers to *the recommendation* as a whole rather than to individual products, so a household that found urea and NPK on the shelf could report availability even if the prescribed potassium, lime, MAP, or CAN were absent; it cannot confirm shelf presence of the very non-standard products whose compliance is zero. Second, it is self-reported and may be coloured by the same product familiarity that shapes the purchase. The farmer-reported evidence thus reinforces a demand-side reading of the aggregate result while leaving the product-specific supply question, for potassium, lime, MAP, and CAN, unresolved without agro-dealer stocking data.

Taken together, Tables 7 and 8 distinguish partial from full recommendation compliance. The voucher raises the probability that farmers apply at least one recommended product and reduces N shortfalls, though it pushes a majority of its recipients past the recommended nitrogen rate; it does not produce exact product-package compliance or close the P and K gaps, and on the strictest compliance measure, the voucher arm falls significantly below the recommendation-only arm. In the terms of Section 2, this pattern rejects the liquidity hypothesis (H2) as a complete account and matches the product-choice-frictions hypothesis (H3): relaxing the budget raised total fertilizer use without raising compliance, with the added purchasing power spent on the familiar bundle.

Table 8: Recommendation compliance and application error

Outcome	T1 mean	T2 − T1	N
Panel A: Nitrogen			
Recommended amount (kg/acre)	23.91	0.66	1180
		(1.01)	
Actual application (kg/acre)	16.39	11.20***	1180
		(1.01)	
Absolute application error (kg/acre)	15.33	-3.46***	1180
		(0.67)	
Shortfall (kg/acre)	11.43	-7.00***	1180
		(0.80)	
Applied more than recommended (share)	0.31	0.30***	1180
		(0.03)	
Panel B: Phosphorus, P_2O_5			
Recommended amount (kg/acre)	14.74	1.21	1180
		(1.85)	
Actual application (kg/acre)	2.66	1.49***	1180
		(0.19)	
Absolute application error (kg/acre)	12.26	-0.40	1180
		(1.80)	
Shortfall (kg/acre)	12.17	-0.34	1180
		(1.80)	
Applied more than recommended (share)	0.04	0.01	1180
		(0.01)	
Panel C: Potassium, K_2O			
Recommended amount (kg/acre)	21.02	2.85	1180
		(2.40)	
Actual application (kg/acre)	1.46	0.74***	1180
		(0.14)	
Absolute application error (kg/acre)	19.65	2.04	1180
		(2.35)	
Shortfall (kg/acre)	19.61	2.08	1180
		(2.35)	
Applied more than recommended (share)	0.03	-0.01	1180
		(0.01)	

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 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 in acres. Absolute application error is the absolute difference between actual and recommended kg/acre; shortfall is the positive part of recommended minus actual kg/acre. Applied more than recommended is an indicator equal to one when actual application of the panel's nutrient exceeds the plot-specific recommended rate. Standard errors clustered at the cluster level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

5.4 Maize production and yield

The endline survey recorded farmer-reported total maize harvested (in kg) 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 9 presents ITT estimates for both outcomes among households for which the soil-test plot's main crop is maize. Panel A uses winsorized total kg harvested (1st–99th percentile) as the outcome; Panel B uses log yield in kg/acre.

Table 9: 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: 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% (*).

Panel A shows that T2 raises total maize harvested by 151 kg (SE = 72, $p = 0.038$), an increase of 19 percent over a control mean of 779 kg. T1 shows no detectable effect (−12 kg, $p = 0.87$). Panel B, reporting log yield in kg/acre again shows no detectable effect for T1. However, T2 increases log maize yield by 0.36 log points, equivalent to approximately 44 percent.[16] The equality test rejects T1 = T2 in both panels.

These results support the causal chain described earlier. 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 rate of approximately 4.9–5.2 kg maize per kg of fertilizer. This exceeds the 2–3 kg/kg range that critical reviews of AIP report as the conditional maize response (Nyondo et al., 2021; Lunduka, Ricker-Gilbert, and Fisher, 2013). 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 favourable comparison reflects a selected plot rather than a generalizable improvement in fertilizer use efficiency.

A natural further question is what the recommendation delivers among farmers who actually followed it. This design cannot identify that quantity. Adherence is not randomly assigned, so the estimate would have to rest on exogenous variation in whether farmers followed the advice, and the experiment supplies very little of it. In the recommendation-only arm about one household in seven already applied every recommended product, and the voucher raises that share by only 1.3 percentage points (standard error 4.1), while exact-package compliance falls (Table 7). The margin the voucher does shift, applying at least one recommended product, rises by 16.4 percentage points, but it is satisfied largely by urea and NPK 23:10:5, products that

[16] 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 A.6 confirm the positive T2 effect across alternative sample restrictions. The per-acre percentage (44 percent) exceeds the levels percentage (19 percent) because the production gain is divided by reported plot area. Because plot area is farmer-reported at endline, reported area is itself potentially affected by treatment, so we treat the levels estimate in Panel A as the primary measure of the production gain.

are at once recommended and familiar; an estimate on that margin would therefore recover the return to applying more of the conventional package rather than the return to precision nutrient management. Comparing farmers who followed the recommendation with those who did not would confound adherence with plot quality and management. The 19 percent estimate should accordingly be read as the effect of the voucher as implemented and, given the near-zero adoption of the products that distinguish the prescription from the familiar package, as a lower bound on what full adherence would deliver.

5.5 Cost-benefit of the yield gain

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 calculation, 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 of 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, 2017–present). These large seasonal fluctuations imply that conclusions regarding the subsidy's profitability are highly sensitive both to the timing of consumption and to whether maize is valued at farmgate or retail prices.

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 production gain of 151 kg per household, the implied break-even price is approximately 1,127 MWK/kg.

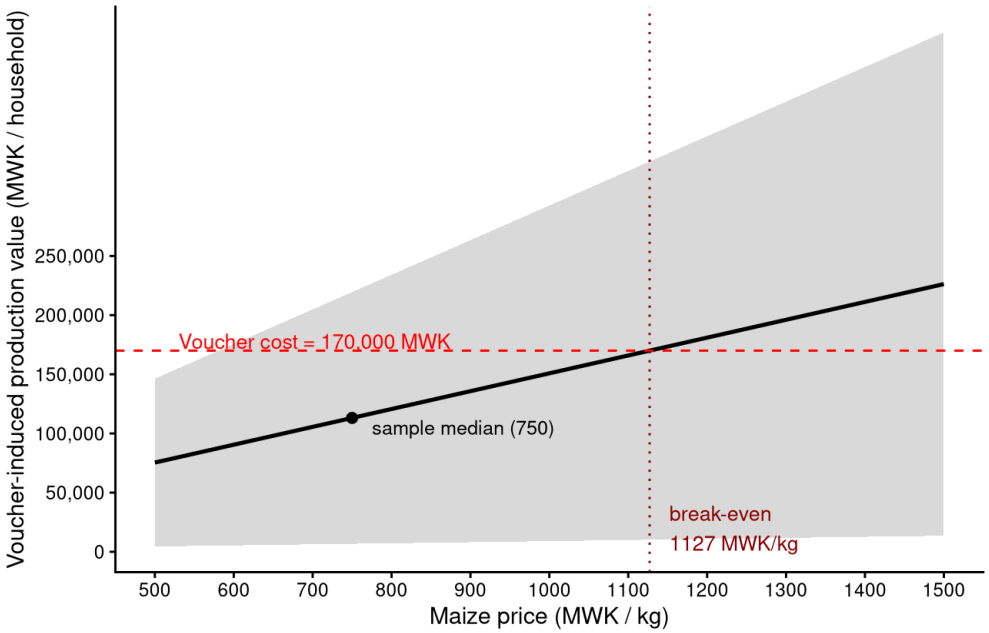


Figure 4: Voucher-induced value of maize production as a function of the maize price. The solid line plots the value of the voucher-induced production gain of 151 kg per household (Table 9, Panel A); the shaded band is the 95 percent confidence interval. The dashed horizontal line marks the voucher cost (170,000 MWK). The implied break-even maize price is approximately 1,127 MWK/kg.

Two features of this break-even calculation warrant emphasis. First, it counts the voucher but not the

soil test. The tests were provided at no cost to the household, so excluding them is the right convention for the farmer-private returns reported in the next subsection, where the household paid for neither the test nor the voucher-funded fertilizer. For a fully costed programme calculation the test is a genuine expense: the true programme cost per beneficiary is therefore higher than the 170,000 MWK voucher face value, and the break-even price of 1,127 MWK/kg is correspondingly a lower bound on what such a programme would need to clear. Second, as argued above on nutrient use, the observed yield gain reflects only intensification of the existing product mix, not the full agronomic correction the soil test was designed to deliver. A programme that successfully implemented the complete soil-test prescription would likely generate larger yield gains and a correspondingly lower break-even price. The current break-even estimate should therefore be read as an upper bound on what is needed to recover programme costs under the observed behaviour, not as the break-even under ideal agronomic compliance.

5.6 Farmer costs and returns

The previous subsection values the voucher-induced maize production gain under alternative price scenarios and shows that its profitability depends critically on how output is priced. That calculation, however, abstracts from changes in input expenditures and therefore does not capture net economic returns.

To capture this broader margin, we construct plot-level economic outcomes for the sampled plot's main crop. The endline data do not measure household-level revenue across all crops, so the analysis focuses on the sampled plot. 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—750 MWK/kg, the same price that anchors the break-even calculation in the previous subsection.

Table 10 reports value of production, costs, and profits in levels, each in total and per acre. Costs aggregate expenditures on seed, pesticides, fertilizer, and other sampled-plot expenses, elicited in separate modules so that the construction avoids double counting. Fertilizer costs reflect farmer-private outlays: because respondents report voucher-funded fertilizer at its full commercial value even though they did not pay for it, we net the voucher-funded kilograms out of each fertilizer record at that record's own reported price. Fertilizer obtained through the AIP is retained, as it is reported at the subsidized price actually paid.[17]

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.

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 10: 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.

The profit decomposition thus reveals a fundamental tension in the programme's design: the private gains that make the voucher attractive to recipient households—primarily cost savings through displacement of

[17] Two exclusions push the T2 profit estimate upward relative to a fully-costed accounting. Family labour is not priced—only cash outlays for hired labour enter, through the residual expense term—so if the voucher crowds in own labour for fertilizer application, T2 net returns are overstated; Beaman et al. (2013) document this bias direction in Mali. The soil-test cost is likewise excluded, which is correct for farmer-private ITT returns but not for programme-level cost-effectiveness.

Table 10: Effects on Value of Production, Costs, and Profits

Outcome	Control mean	T1	T2	p-value: T2 = T1	N
Panel A: 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)		
Panel B: 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: 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.

commercial purchases—are the same mechanism that reduces the programme's social value. A subsidy that transfers purchasing power to households that would have bought the same fertilizer anyway generates substantial inframarginal transfers with limited productive additionality, a finding consistent with the broader crowding-out literature on input subsidies in Sub-Saharan Africa (Mason, Jayne, and Mofya-Mukuka, 2013; Ricker-Gilbert, Jayne, and Chirwa, 2011). The parallel with the AIP is direct: displacement of commercial purchases is the signature critique of Malawi's own subsidy programme, and the recommendation-paired voucher does not escape it, since the same inframarginal logic applies whenever transfers reach households that would have purchased fertilizer anyway. Two features nonetheless distinguish the voucher's version of the problem. First, whereas AIP crowd-out routes sales away from private dealers into parallel public distribution, the voucher keeps even inframarginal sales within the commercial retail channel, preserving the market-development rationale for reform (Basurto, Dupas, and Robinson, 2020). Second, the two transfers displace each other: as shown in Appendix 3, the voucher reduces AIP take-up by about 40 percent. Whether this second crowd-out yields fiscal savings or equity gains depends on what happens to the forgone packages: if the AIP envelope contracts, the government recovers its subsidy outlay, whereas under a fixed, rationed envelope the more plausible outcome is reallocation to other households, extending programme coverage. Either way the offset is modest, since the displaced AIP fertilizer is worth less than one-tenth of the voucher's commercial-price equivalent. Targeting links the two critiques: transfers are inframarginal precisely for households wealthy enough to have purchased fertilizer anyway (Ricker-Gilbert, Jayne, and Chirwa, 2011), so directing support toward liquidity-constrained households would raise additionality and improve equity simultaneously—though our randomized allocation cannot speak to how a targeted deployment would perform.

6 Policy Implications and Conclusion

Fertilizer subsidy programmes dominate agricultural spending across much of sub-Saharan Africa, and Malawi's programme has consistently been among the largest in budget-share terms. With fertilizer prices high and volatile and fiscal space tight, the case for replacing the blanket subsidy package with a site-specific alternative anchored in soil intelligence has been pressed in both the policy and academic literatures. The argument rests on the premise that more precise nutrient management recommendations—in our study context, based on farm-level soil testing—can redirect farmer behaviour toward an agronomically appropriate input mix, whether on their own or, where financial constraints bind, when paired with the means to act on them. This paper tests this premise in a cluster-randomized trial across 113 villages in central Malawi.

The premise does not hold in either form. The recommendation alone changed neither input use, nutrient application, nor maize yield, so information is not the binding constraint; paired with purchasing power, it changed quantity but not composition. Adding a voucher raised fertilizer use on the soil-test plot by about 30 kg and total maize production by about 19 percent, but the increment is composed almost entirely of about 18 kg more urea and 19 kg more NPK 23:10:5, a 50/50 split mirroring the AIP bundle it partly displaces, while CAN, potassium products, lime, and MAP—the products that distinguish the prescription from the familiar package—remain at or near zero. The evidence thus favours product-choice frictions, the pull of familiar products over prescribed ones, rather than lack of liquidity as the constraint that binds once the recommendation is delivered. The result is nitrogen intensification rather than precision: the voucher pushes a majority of recipients past their plot-specific nitrogen recommendation, financing nitrogen beyond the site-specific optimum while the phosphorus, potassium, and lime gaps the soil test was designed to close remain open. To the extent that these nutrients are co-limiting, the measured yield response is a lower bound on what full compliance would unlock.

The returns to the voucher look very different to the farmer and to the public purse. At the median sale price the value of the additional maize falls short of the voucher's face value; break-even requires 1,127 MWK/kg, above the median price though below peak retail prices, arguably the relevant benchmark when own production displaces lean-season purchases. For the farmer, by contrast, the voucher is unambiguously attractive: because voucher fertilizer displaces commercial purchases, households spend about 94,000 MWK less of their own money on inputs and farm profits rise by roughly 190,000 MWK. But displaced spending is a transfer rather than newly created value, and the voucher also reduces the probability that a household takes up AIP by 7.7 percentage points, about 40 percent of the control rate. Private

profitability therefore coexists with a weak fiscal case.

The null recommendation effect replicates the modal result in comparable studies (Beg, Islam, and Rahman, 2024; Harou et al., 2022; Corral et al., 2020) and diverges from Ayalew, Chamberlin, and Newman (2022), who find positive effects of recommendations alone in Ethiopia. Our data cannot isolate the source of the divergence: product availability (Sida et al., 2023), baseline nutrient deficits, and extension delivery all differ. One underemphasized candidate, however, is the structure of the recommendation itself. In Ethiopia the prescription adjusted the quantity of products farmers already knew. In our study context, the modal recommendation asks for a product switch: calcium ammonium nitrate in place of urea, potassium where none was used, lime on plots that never received it. A product switch faces a qualitatively different behavioural barriers than a quantity adjustment, and information delivery alone may not bridge it in a single season. The conditions that make this barrier bind are not unusual: a long-running subsidy programme that has anchored two products as the default, thin agro-dealer product depth beyond those products, and farmers wary of spending on inputs they have never used. Where these conditions hold, and they hold across much of sub-Saharan Africa, we would expect the same pattern: recommendations alone leaving fertilizer choices unchanged, and voucher spending gravitating toward the familiar package.

Two policy implications follow. First, information-only instruments (soil-health cards, blanket extension messaging, stand-alone decision-support tools) are unlikely to redirect fertilizer behaviour at scale in settings like ours, so the case for replacing the AIP package with a recommendation-paired voucher cannot be made on agronomic-precision grounds. If that case survives, it is on commercial-channel grounds: routing fiscal support through agro-dealers may improve targeting and build private input markets (Basurto, Dupas, and Robinson, 2020), although these are benefits we do not measure. Any such reform would also need to be integrated with the existing subsidy architecture rather than run alongside it, because the voucher partially crowds out AIP take-up, whether because households forgo the subsidy once the voucher covers their needs or because village-level allocation shifts away from voucher recipients, and the policy-relevant counterfactual is therefore voucher versus AIP, not voucher versus nothing. Second, recommendation design should distinguish a core package, made up of products the farmer already uses and for which the recommendation adjusts quantities and timing, from an aspirational package whose products require supply-chain development and demonstrated farmer familiarity before they are actionable. Ranking products by marginal agronomic contribution would give farmers a basis for prioritizing under budget and supply constraints, echoing the Chandrasekhar et al. (2022) insight that actionable advice must map onto what farmers can credibly vouch for.

Several open questions warrant consideration in the design of future programmes. First, whether the voucher response persists after withdrawal, and whether recommendations generate lagged effects through learning or diffusion, requires a medium-run follow-up. Such a follow-up could also track whether the substitution away from organic inputs documented above accumulates into declining soil carbon. Second, whether the zero compliance on potassium and lime use reflects farmer demand or agro-dealer supply cannot be resolved without stocking data; collecting point-of-sale inventories alongside voucher redemption, or randomizing the pre-positioning of less common products, would separate the two, and would show whether compliance failure is milder where the missing products are routinely stocked. Finally, a coordinated multi-country replication varying recommendation content, agro-dealer product depth, and the scale of parallel subsidy channels would be more informative than further single-country trials.

Note

Author names are listed in alphabetical order. The trial was registered in the American Economic Association (AEA) RCT Registry on 14 October 2024 (AEARCTR-0014293). Ethics approval was obtained from the International Food Policy Research Institute (IFPRI) Institutional Review Board (IPS-24-0932) and from the National Committee on Research in the Social Sciences and Humanities (NCRSH) in Malawi (P.08/24/917). During the preparation of this work the authors used AI-assisted tools (Anthropic's Claude and OpenAI's Codex) to support drafting, editing, and reference management; the authors reviewed and verified all content and take full responsibility for the content of the publication.

References

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.

Anderson, M. L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

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, C., 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., T. Jayne, and J. R. Black. 2017. "Factors explaining the low and variable profitability of fertilizer application to maize in Zambia." *Agricultural Economics* 48 (1): 115–126.

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.

Chivenge, P., S. Zingore, K. Ezui, S. Njoroge, M. Bunquin, A. Dobermann, and K. Saito. 2022. "Progress in research on site-specific nutrient management for smallholder farmers in sub-Saharan Africa." *Field Crops Research* 281: 108503.

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.

De Weerdt, J., B. Dillon, E. Hami, B. Van Campenhout, and L. Nabwire. 2024. "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.

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.

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.

Holden, S. T. 2019. "Economics of farm input subsidies in Africa." *Annual Review of Resource Economics* 11: 501–522.

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.

International Food Policy Research Institute (IFPRI). 2017–present. "IFPRI Malawi Monthly Maize Market Reports." Washington, DC: International Food Policy Research Institute. https://hdl.handle.net/10568/181752.

Jack, B. 2013. *Constraints on the adoption of agricultural technologies in developing countries.* White paper, Agricultural Technology Adoption Initiative, J-PAL (MIT) and CEGA (UC Berkeley).

Jayne, T., N. M. Mason, W. J. Burke, and J. Ariga. 2018. "Review: Taking stock of Africa's second-generation agricultural input subsidy programs." *Food Policy* 75: 1–14.

Johnston, A. M. and T. W. Bruulsema. 2014. "4R nutrient stewardship for improved nutrient use efficiency." *Procedia Engineering* 83: 365–370.

Khonje, M. G., C. Nyondo, L. Chilora, J. H. Mangisoni, J. Ricker-Gilbert, and W. J. Burke. 2022. "Exploring adoption effects of subsidies and soil fertility management in Malawi." *Journal of Agricultural Economics* 73 (3): 874–892.

Kilic, T., E. Whitney, and P. Winters. 2015. "Decentralised beneficiary targeting in large-scale development programmes: Insights from the Malawi Farm Input Subsidy Programme." *Journal of African Economies* 24 (1): 26–53.

Lancaster, K. J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2): 132–157.

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.

Mason, N. M., T. S. Jayne, and R. Mofya-Mukuka. 2013. "Zambia's input subsidy programs." *Agricultural Economics* 44 (6): 613–628.

Mather, D. and T. Jayne. 2018. "Fertilizer subsidies and the role of targeting in crowding out: Evidence from Kenya." *Food Security* 10 (2): 397–417.

Moffitt, R. 1989. "Estimating the Value of an In-Kind Transfer: The Case of Food Stamps." *Econometrica* 57 (2): 385–409.

Nyondo, C., M. Khonje, J. Mangisoni, W. J. Burke, J. Ricker-Gilbert, and L. Chilora. 2021. *Lessons learnt: Promises, achievements, shortcomings, and pitfalls of inputs subsidy programs in Malawi*. Working Paper 21/05, 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.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1): 34–55.

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.

Southworth, H. M. 1945. "The Economics of Public Measures to Subsidize Food Consumption." *Journal of Farm Economics* 27 (1): 38–66.

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

Theriault, V., M. Smale, and H. Haider. 2018. "Economic incentives to use fertilizer on maize under differing agro-ecological conditions in Burkina Faso." *Food Security* 10: 1263–1277.

Vanlauwe, B., K. Descheemaeker, K. E. Giller, J. Huising, R. Merckx, G. Nziguheba, J. Wendt, and S. Zingore. 2015. "Integrated soil fertility management in sub-Saharan Africa: unravelling local adaptation." *SOIL* 1: 491–508.

Vanlauwe, B., T. Amede, P. Bindraban, R. Cardinael, A. Couëdel, M. Corbeels, A. Dobermann, O. A. Fatunbi, G. Falconnier, K. E. Giller et al. 2026. "Fertilizer and soil health for sustained crop production in sub-Saharan Africa: Theory and conceptual framework." *npj Sustainable Agriculture* 4 (1): 56.

Vos, R., J. Glauber, C. Hebebrand, and B. Rice. 2025. "Global shocks to fertilizer markets: Impacts on prices, demand and farm profitability." *Food Policy* 133: 102790.

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

Wortmann, C. S., K. R. Sones et al. 2017. *Fertilizer use optimization in sub-Saharan Africa.* CABI Wallingford, UK.

Online Appendices

1 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 A.1: 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. This linkage stage is not the denominator for endline-survey outcomes; recommendation-compliance and application-error outcomes are estimated only among T1 and T2 households with valid recommendation records, so the lower control linkage rate is not used in those comparisons.

Table A.2: 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. The main recommendation-compliance and application-error estimates use T1 and T2 households with valid recommendation records and do not compare either treatment arm with the lower-linked control group. Asterisks: 1% (***), 5% (**), 10% (*).

2 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

Table A.3: 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 (acres)	2.499	0.007	0.016	1863
		(0.137)	(0.145)	
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. Asterisks: 1% (***), 5% (**), 10% (*).

Table A.4: 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 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 village level are reported in parentheses. Asterisks denote significance at the 1% (***), 5% (**), and 10% (*) levels.

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.

3 AIP substitution

This appendix details the AIP-substitution analysis summarized in the main text: whether the voucher-induced fertilizer increase partly replaces fertilizer from the Affordable Inputs Programme (AIP) rather than adding to it.

A treatment-induced reduction in AIP fertilizer on the soil-test plot admits two readings. Under *genuine household-level substitution*, voucher-arm households take up less AIP, or none at all; because trial participation did not affect AIP eligibility, which is allocated independently through government channels, any such reduction reflects household or committee decisions rather than administrative exclusion. Under *within-household reallocation*, households still take up the full 100 kg package but redirect it toward their other plots now that voucher fertilizer covers the soil-test plot, leaving household-level take-up unchanged and making the test-plot reduction an artifact of plot selection.

The distinction matters because of how the soil-test plot was selected: households nominated their primary plot, typically their main maize field, and AIP fertilizer is applied preferentially to precisely this plot. 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, so a test-plot-only comparison cannot separate the two readings.

Table A.5 therefore reports four outcomes: AIP take-up, a binary indicator for whether the household obtained any AIP fertilizer (the household-level extensive margin, unaffected by where the fertilizer was applied); AIP quantities on the test plot (the outcome typically reported in the plot-level displacement literature) and on the randomly selected non-test plot, the latter restricted to households cultivating at least one additional plot; and the two-plot total, to which households with no other plot contribute zero.

The estimates point to substitution. T2 reduces the probability of any AIP take-up by 7.7 percentage points, about 40 percent of the control-group rate of 19 percent, and reduces AIP fertilizer on the soil-test plot by 7.1 kg, about 74 percent of the control mean. Reallocation would instead surface as a positive effect on the random plot, where the estimate is statistically indistinguishable from zero (+0.21 kg, SE = 1.09); consistently, the two-plot total falls by 6.8 kg, nearly identical to the test-plot reduction. T1 is indistinguishable from zero on every outcome, so the response is specific to the voucher and not to the recommendation alone.

This partial substitution is consistent with two mechanisms: *demand-side substitution*, in which households that have already secured fertilizer through the voucher are less likely to pursue or pay the co-payment for an additional subsidized package, and *supply-side re-targeting*, in which AIP committees or village leaders deprioritise households known to have received voucher support (Basurto, Dupas, and Robinson, 2020). The data do not fully separate these channels.

Table A.5: AIP substitution: effects on take-up and on AIP fertilizer applied at the plot and household level

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

Notes: ITT effects on four AIP outcomes, all-crops sample. AIP take-up is a linear probability model for whether the household obtained 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 subsidized package 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% (*).

4 Sensitivity checks

This section reports robustness checks for the yield results, showing how sensitive the main conclusions are to alternative handling of extreme yield values. They do not change the main interpretation: the voucher arm produces clear effects on implementation, fertilizer use, and maize productivity.

Table A.6: 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 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.

5 Multiple hypothesis testing

For each outcome family, Table A.7 reports the raw p-value and the sharpened two-stage false-discovery-rate q-value (Anderson, 2008) for the preferred contrast. The q-values are computed within each family across the outcomes shown, with families ordered as the outcomes appear in the paper. All families were pre-specified except AIP substitution, which we add because the crowd-out result features among the paper's conclusions. Under this adjustment every outcome on which the paper's conclusions rest, the effects of the voucher on fertilizer and nutrient use, product and nutrient compliance, AIP take-up and displacement, maize production and yield, costs, and profits, remains significant at the 5% level, most at the 0.1% level. The only outcome that is significant on its raw p-value but not after adjustment is one secondary soil-management practice, the use of *mbeya* fertilizer, which is not part of the main argument.

Table A.7: Raw p-values and sharpened FDR q-values within outcome families

Outcome	Contrast	Raw p	Sharpened q
Delivery and adherence			
Received recommendation	T2 = T1	0.096	0.237
Agronaut visited household	T2 = T1	0.780	0.640
Fertilizer and nutrient use			
Maize nutrient use: Potassium (kg/acre)	T2 - control	< 0.001	< 0.001
Maize nutrient use: Phosphorus (kg/acre)	T2 - control	< 0.001	< 0.001
Maize nutrient use: Total nutrients (kg/acre)	T2 - control	< 0.001	< 0.001
Maize nutrient use: Nitrogen (kg/acre)	T2 - control	< 0.001	< 0.001
Total fertilizer use: All crops	T2 - control	< 0.001	< 0.001
Total fertilizer use: Maize only	T2 - control	< 0.001	< 0.001
Product compliance			
Recommended NPK 23:10:5 applied	T2 - T1	< 0.001	< 0.001
Applied at least one recommended product	T2 - T1	< 0.001	< 0.001
Share of recommended products applied	T2 - T1	0.004	0.007
Recommended urea applied	T2 - T1	0.010	0.010
Recommended potassium product applied	T2 - T1	0.027	0.018
Applied exact recommended product set	T2 - T1	0.063	0.033
Applied all recommended products	T2 - T1	0.563	0.088
Nutrient compliance			
Actual N applied	T2 - T1	< 0.001	< 0.001
N shortfall	T2 - T1	< 0.001	< 0.001
Actual P_2O_5 applied	T2 - T1	< 0.001	< 0.001
Absolute N application error	T2 - T1	< 0.001	< 0.001
Actual K_2O applied	T2 - T1	< 0.001	< 0.001
K_2O shortfall	T2 - T1	0.339	0.247
Absolute K_2O application error	T2 - T1	0.347	0.247
Absolute P_2O_5 application error	T2 - T1	0.901	0.696
P_2O_5 shortfall	T2 - T1	0.923	0.696
AIP substitution			
AIP fertilizer on test plot	T2 - control	< 0.001	< 0.001
Two-plot AIP fertilizer total	T2 - control	< 0.001	< 0.001
AIP take-up (0/1)	T2 - control	0.002	0.002
AIP fertilizer on random plot	T2 - control	0.846	0.268
Production and economic outcomes			
Profits	T2 - control	< 0.001	0.002
Log maize yield	T2 - control	< 0.001	0.002
Total costs	T2 - control	< 0.001	0.002
Profits per acre	T2 - control	0.014	0.018
Total maize harvested	T2 - control	0.039	0.033
Value per acre	T2 - control	0.047	0.033
Value of production	T2 - control	0.124	0.067
Costs per acre	T2 - control	0.180	0.067
Secondary SNM practices			
Dairy or poultry manure	T2 - control	0.004	0.017
Ridges/check dams	T2 - control	0.004	0.017
Fresh vegetative material	T2 - control	0.009	0.022
Mbeya fertilizer	T2 - control	0.046	0.075
Green legume incorporation	T2 - control	0.216	0.276
Farmyard manure	T2 - control	0.266	0.284
Pit planting	T2 - control	0.638	0.638
Minimum tillage	T2 - control	0.759	0.638
Compost	T2 - control	0.850	0.638

Notes: For each outcome the table reports the raw p-value and the sharpened two-stage false-discovery-rate (q) value (Anderson, 2008), computed within each outcome family across the outcomes shown. The reported contrast is the incremental voucher test (T2 = T1) for delivery and adherence, the T2-minus-T1 difference among treated households with valid recommendation records for the product- and nutrient-compliance families, and the T2-versus-control intent-to-treat effect for fertilizer and nutrient use, AIP substitution, production and economic outcomes, and secondary soil-management practices. All families were pre-specified except AIP substitution. All p-values are from the unadjusted ITT specifications reported in the paper's tables.

6 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 expenditures of fertilizer, seed , pesticide , 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 ITT specifications omit pre-treatment controls. Post-treatment variables such as Agronaut visits or extension contact during implementation are not included.

ALL IFPRI DISCUSSION PAPERS

All discussion papers are available **here**

They can be downloaded free of charge

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
www.ifpri.org

IFPRI HEADQUARTERS
1201 Eye Street, NW
Washington, DC 20005 USA
Tel.: +1-202-862-5600
Fax: +1-202-862-5606
Email: ifpri@cgiar.org