Study on Adoption and Impact of the ÚNICA Variety in Peru

Prepared for International Potato Center (CIP)

December 2024

laterite
DATA | RESEARCH | ANALYTICS

laterite
DATA | RESEARCH | ANALYTICS

Table of Contents

Introduction

The potato variety ÚNICA, developed by the International Potato Center (CIP) and launched in Peru in 1998, stands out as a remarkable agricultural innovation due to its high productivity, resistance to diseases, and adaptability to diverse climates. Originally bred from the Lowland Tropics Virus Resistance (LTVR) breeding population, ÚNICA was designed to enhance food security and improve farmers' incomes globally. Its ability to thrive in both mountainous and lowland environments, coupled with resistance to potato viruses and late blight—a devastating disease—has contributed to its widespread success (CIP, 2018).

ÚNICA's attributes go beyond disease resistance. It produces abundant tubers that mature quickly, within three to four months after planting (Quevedo, M.[1]). First selected in the Andean region of Huancayo for its resilience to late blight, the variety underwent over a decade of rigorous evaluation across 17 coastal plains in Peru. These tests confirmed its tolerance to heat, drought, and saline soils. The variety was later renamed after the Universidad Nacional de Ica (UNICA), honoring the institution's contribution to its national release. ÚNICA's rapid adoption by farmers can be attributed to its unique combination of resilience and productivity (Fonseca, C.). Farmers were drawn to its ability to produce tubers quickly and resist common diseases, making it a reliable source of income and food security.

Beyond Peru, the ÚNICA variety has been introduced to countries such as Kenya, Tajikistan, Tanzania, and China, where it is known as Qingshu 9. In Eastern Africa, ÚNICA has achieved notable adoption rates, particularly in Kenya, where certified seeds were distributed to thousands of farmers, resulting in significant yield improvements. In Tanzania, the variety, locally known as Mkanano, has been adopted by farmers and is currently undergoing field evaluations as part of the process for official release. Moreover, it has been introduced and evaluated in countries such as Uzbekistan, Georgia, Rwanda, and Bhutan, further solidifying its role as a global solution for enhancing food security and increasing incomes (CIP, 2018). This widespread adoption highlights ÚNICA's remarkable adaptability to diverse environments, ranging from semi-arid regions to coastal plains and high-altitude areas.

Interviews with CIP professionals suggest that understanding the mechanisms behind its adoption, market dynamics, and barriers is essential to fully realize ÚNICA's potential. The primary goal of this research is to design and validate an **adoption and impact study** to measure the effects of ÚNICA on key indicators such as **poverty, livelihoods, and jobs (PLJ)** while addressing the factors that facilitate or hinder its adoption and sustainability. Specifically, the study seeks to answer the following questions:

1. **What are the effects of adopting ÚNICA on farmers' income, livelihoods, and employment?**
2. **What factors promote the adoption of the ÚNICA variety, and what elements contribute to ensuring its long-term sustainability?**

[1] This citation, as well as the others in the same format throughout the document, refer to the interviews conducted during the project.

This study aims to use Propensity Score Matching (PSM) and Differences in Differences (DID) methods to compare between farmers who adopted ÚNICA and those who did not adopt this variety, providing a methodological framework to estimate the impact of adopting this variety.

The relevance of this research transcends the Peruvian context, offering practical evidence to develop policies and agricultural strategies aimed at farmers worldwide. The potential impact of ÚNICA in improving livelihoods and reducing rural poverty aligns with the CIP2030 Strategy and CGIAR goals, providing insights to guide decisions toward more resilient and inclusive agriculture.

Context

Understanding the socio-economic context of Peruvian farmers is crucial to evaluate the adoption and impact of the ÚNICA potato variety. In Peru, structural transformation has led to a gradual decline in the share of agriculture in both GDP and employment. However, this transformation has been slow, with limited changes in the economic composition since the 1990s. During 1990–2015, agriculture contributed only 0.3 percentage points to GDP growth but played a significant role in reducing extreme poverty, given the high proportion of poor households reliant on agricultural livelihoods (World Bank, 2017).

According to the 2012 National Agricultural Census (CENAGRO), there were approximately 2.2 million agricultural units nationwide, 99.3% of which were managed by individuals. Approximately 633 thousand (28%) of these units grew some variety of potato. Demographically, 69% of those managing these units were male, with an average age of about 50 years. The main mother tongue of these producers was Quechua (48%), followed by Spanish (41%). In addition, 82% of these producers reported being able to write, but about 50% had only an incomplete primary education or less.

Despite the importance of potato production, the resources available to Peruvian farmers remain limited. The median total area farmed by these producers was 1.05 hectares, of which 0.25 hectares was devoted to potatoes. Only 6% of these producers reported having received training, 2% had received technical assistance, and less than 1% had received extension services. In addition, only 5% used improved seed. Significantly, only 30% of these producers reported that they set aside part of their production for sale. In addition, only 24% reported that their farm income was sufficient to cover their expenses.

Peru has diverse sources of agricultural data; however, the Price and Supply System (SISAP) of the Ministry of Agriculture and Irrigation (MIDAGRI) is one of the few systems that disaggregates information by crop variety. This system records data on the volumes of agricultural products, including potato varieties, that arrive at Lima's Gran Mercado Mayorista. For the ÚNICA variety, data is predominantly available from 2010 onwards[2], while records for other varieties date back to 1997.

An analysis of the regional distribution of ÚNICA production volumes from 2010 to 2024 reveals notable trends, as illustrated in the graphs below. It is important to note that **the SISAP database provides information on the production volumes estimated in Lima based on regional supply, rather than the total production in those regions**. This distinction is

[2] Junín and Lima are the only regions with data available for 2008 and 2009.

crucial, as agricultural varieties are also sold in local markets, supermarkets, and other venues, meaning the data should not be interpreted as the total regional production. Despite this limitation, SISAP remains a valuable data source for assessing trends.

The graphs specifically reflect the volumes of ÚNICA reported in Lima's wholesale market, representing quantities supplied from different regions. Graph 1 shows the percentage distribution of production by region over time, while Graph 2 displays absolute production volumes, also over time.

Figure 1 shows that, over the years, Lima's share steadily declined, although it remained a dominant contributor, often surpassing 30% in later periods. Ica emerged as a significant contributor starting in 2014, with its share increasing rapidly, particularly between 2015 and 2024, where it reached peaks surpassing 30%. Similarly, Arequipa exhibited growing importance, showing a more gradual but consistent rise in production share, particularly after 2016, where it frequently maintained shares between 20% and 30%. These two regions, alongside Lima, have been the primary contributors to the production volumes of the ÚNICA variety in recent years.

Junín has maintained a more stable yet moderate share throughout the period, oscillating between 10% and 20%, contributing consistently to the overall production. In contrast, regions such as Ayacucho, Huánuco, Pasco, and La Libertad consistently showed minimal contributions, often falling below 5% of the total volume. Despite occasional increases, their shares remained relatively small compared to the larger producing regions.

Figure 1: Regional export volume (%) 2010 – 2024. Source: Own elaboration, based on data extracted from the SISAP-MIDAGRI platform for 2008-2024. Note: Data for the regions of Apurímac, Huancavelica and Lambayeque have significant gaps during this period and are therefore excluded from the graph.

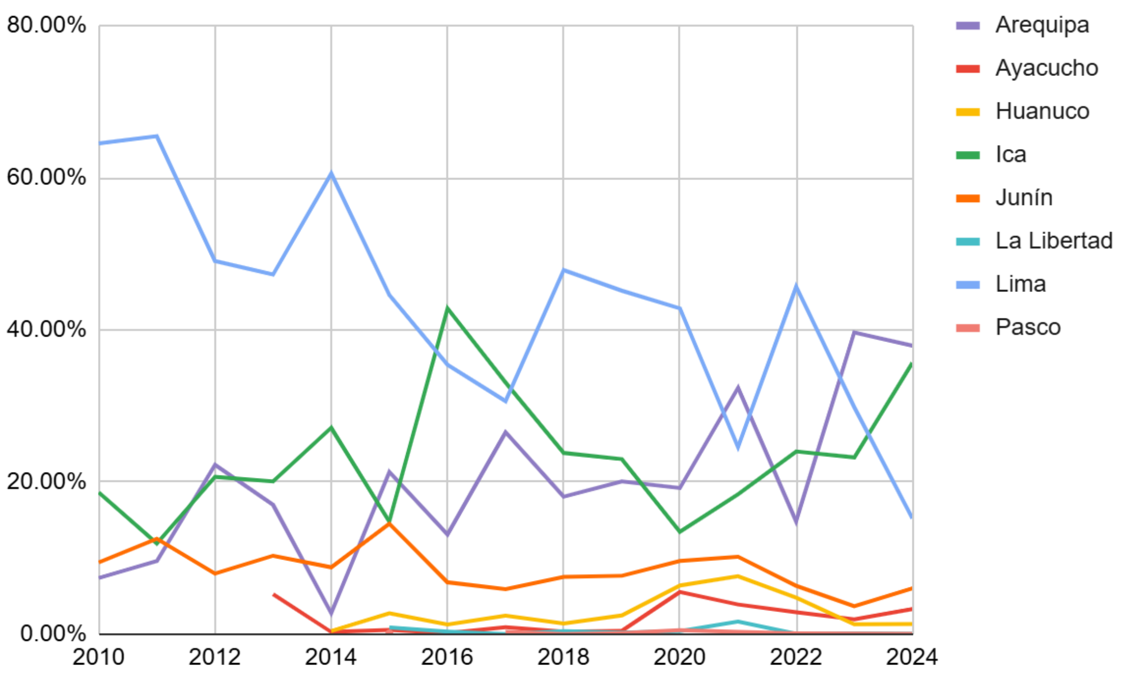
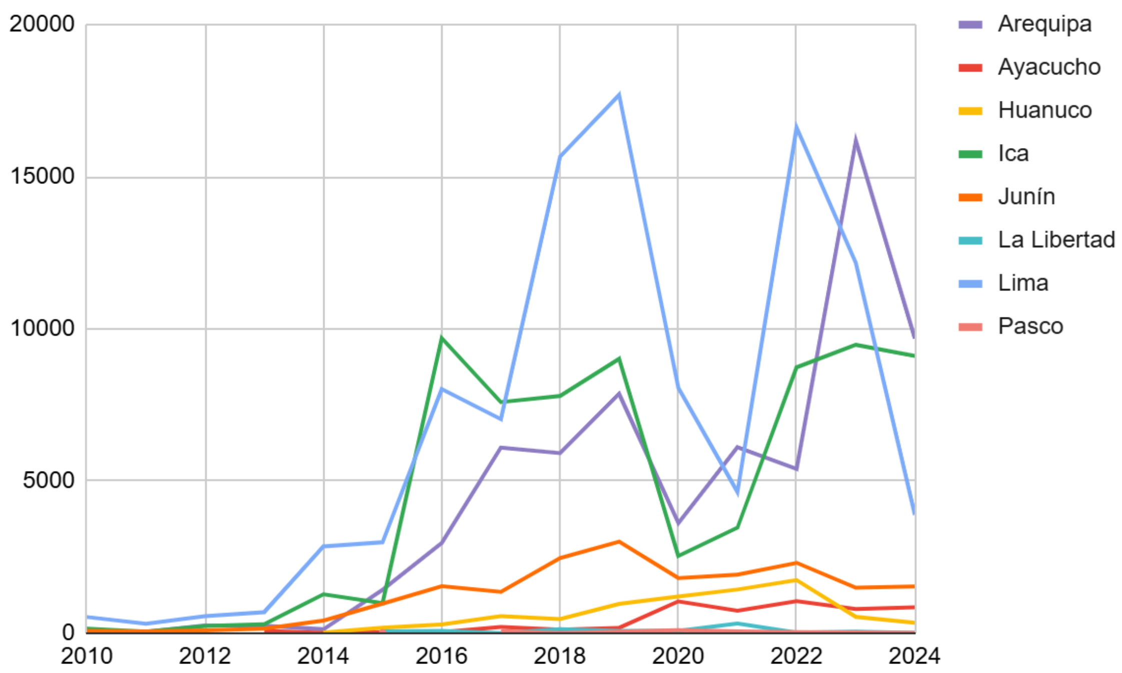


Figure 2 illustrates the reported volumes of the ÚNICA variety arriving at Lima's Gran Mercado Mayorista. Lima consistently registers the highest volumes during this period, with significant peaks between 2016 and 2019, followed by a sharp decline and subsequent fluctuations. Arequipa shows notable growth after 2014, reaching its highest levels around 2022 before a slight decrease in recent years. Ica also demonstrates a strong upward trend beginning in 2015, peaking in 2018, and stabilizing thereafter. Junín exhibits moderate but steady volumes with peaks around 2018 and 2020, while regions like Pasco, Ayacucho, and La Libertad record lower overall volumes with occasional increases, particularly in 2020 and 2021.

The data reveals a general growth trend in reported supply between 2015 and 2018, with many regions peaking between 2018 and 2022. However, several regions experience declines post-2020.

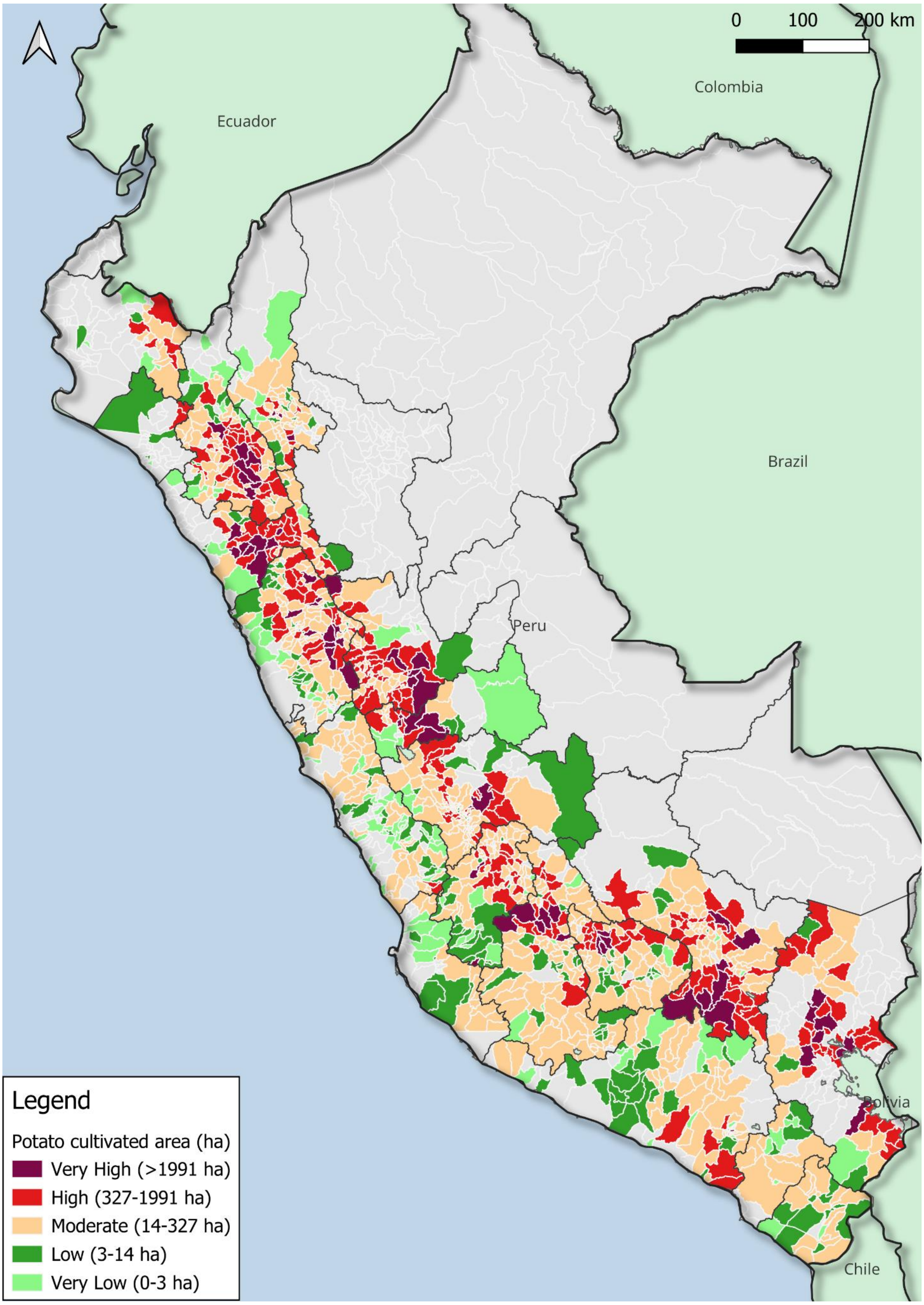
Figure 2: Regional export volume (tons) 2010 – 2024. Source: Own elaboration, based on data extracted from the SISAP-MIDAGRI platform for 2010-2024. Note: Data for the regions of Apurímac, Huancavelica and Lambayeque have significant gaps during this period and are therefore excluded from the graph.



An additional data source is the Profile of the Agricultural Producers Registry[3] (Padrón de Productores Agrarios" (PPA-MIDAGRI)). Established in 2021, this registry gathers exhaustive declarative information from agricultural producers, including detailed data on personal, household, and agricultural activities. It is an essential instrument for policy making, aiming to optimize agricultural productivity and sustainability. Specifically, the section on territorial economic valuation offers a detailed geographic analysis at the district level, providing insights into the spatial distribution and specific characteristics of agricultural production across Peru.

[3] Link to online information.

Figure 3: Distribution of crop at district level in Perú. Source: Own elaboration, based on data extracted from the SISAP - MIDAGRI platform. 2023.



While previous information provided an overview of the volume of the ÚNICA variety arriving in Lima for sale, it is equally essential to identify the regions and districts where the variety is cultivated. From the 25 regions of Peru, 19 regions cultivate potatoes, while six—Callao, Loreto, Madre de Dios, San Martín, Tumbes, and Ucayali—do not produce this crop. Using the collected data, we focused on identifying regions and districts where the ÚNICA variety is specifically grown.

Based on this analysis, a district-level database was developed to geographically pinpoint areas of production for the ÚNICA variety. This database, which will be attached in Excel format, allows for a more granular understanding of its cultivation across the country. The findings from this database are visualized in Figures 3 and 4, which illustrate both the overall potato production at the district level and the specific production areas of the ÚNICA variety.

First, Figure 3 illustrates potato production in Peru, measured by the cultivated area in hectares at the district level. The data is represented through a color-coded classification, where dark purple indicates very high production (>1991 hectares), red corresponds to high production (327-1991 hectares), orange represents medium production (14-327 hectares), light green indicates low production (3-14 hectares), and dark green shows very low production (0-3 hectares).

The results reveal a significant concentration of potato production in the central and southern highlands of Peru. Districts with very high and high production are predominantly located in regions such as Puno, Cusco, Huánuco, Ayacucho, and Huancavelica, which are traditionally known for their extensive agricultural activity (Pradel et al, 2017). In contrast, coastal and jungle areas display lower levels of production, with many districts categorized as medium, low, or very low production zones. This distribution highlights the importance of the highlands as the primary potato-producing region in the country, consistent with Peru's historical and economic reliance on potato cultivation.

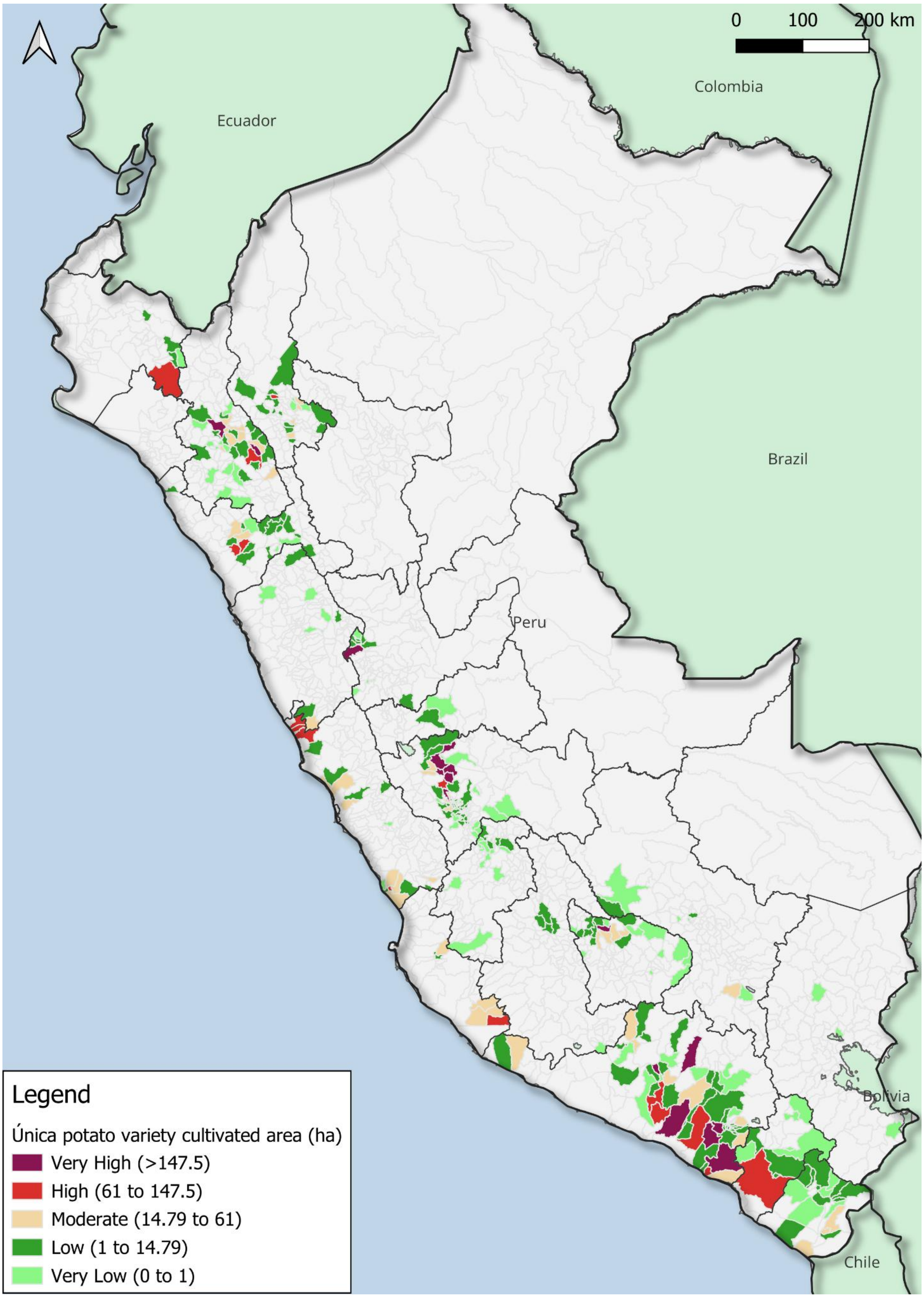
While the exact proportions may vary over time, the current map reflects the spatial distribution of potato cultivation, emphasizing the dominance of the highland regions and the relatively limited production in other areas such as the coast and jungle.

Figure 4 illustrates the distribution of production levels for the ÚNICA variety across regions in Peru. Notably, the variety is cultivated in only 319 out of 1,658 districts where potatoes are grown, which represents approximately 19.2% of the total districts.

Among the districts cultivating ÚNICA, low production areas (1 to 14.79 hectares) are the most prevalent, representing 8.87% of the total (147 districts). In contrast, very high production areas (over 147.5 hectares) are rare, accounting for only 0.90% (15 districts). This distribution underscores that most production occurs at low or very low levels, with few areas reaching moderate or high production levels.

Arequipa stands out as the leading region, with 9,688.04 hectares of ÚNICA under cultivation, representing 56.80% of the national total. Junín follows with 3,732.81 hectares (21.89%), and Cajamarca ranks third with 937.20 hectares (5.49%). Other regions, including Ayacucho, Puno, and Pasco, contribute minimally, each accounting for less than 0.1% of the total cultivated area. It is important to note that while the previous graphs illustrate a significant volume of ÚNICA supplied from the Lima region to the Gran Mercado Mayorista de Lima, in recent years there has been a decline in production from this region. This decrease in the volume of ÚNICA arriving at the market may be related to the decrease in production levels in Lima's region lately.

Figure 4: Distribution of ÚNICA variety at district level in Perú. Source: Own elaboration, based on data extracted from the SISAP - MIDAGRI platform. 2023.



Theory of Change

The theory of change for this project posits that the adoption of the ÚNICA variety, due to its high productivity, disease resistance, early maturity, and climatic adaptability, has the potential to improve producers' livelihoods, increase their income, generate employment in the sector, and reduce rural poverty.

Potato Cultivation in Peru

In Peru, potato production is seasonal, which directly impacts market prices. According to SISAP and MIDAGRI (2023), potato production and supply are concentrated in the first six months of the year, especially in the highlands. During this period, oversupply causes price drops, affecting producers' incomes. For instance, in 2020, potato overproduction, including the ÚNICA variety, reduced its minimum price to S/. 0.57 per kilogram (MIDAGRI, 2023). It is important to note that in this context, producers are the only ones negatively affected. Due to the market's structure, buyers and intermediaries benefit from price drops instead.

Despite this phenomenon, ÚNICA producers manage to achieve considerable income due to various factors:

- **Focus on the Industrial Market**: This variety is preferred for processing into French fries and other products. Industrial production ensures a fixed market for ÚNICA producers, facilitating quick sales.
- **Productivity and Yield**: It yields between 25 and 35 tons per hectare, exceeding the national average of 15 to 21 tons (MIDAGRI, 2023). In some districts, yields reach up to 50 tons (Quevedo, M.) per hectare. With water availability, it can be planted year-round (Salas, E.).
- **Early Maturity**: Thanks to its faster sprouting period compared to other varieties, producers benefit from sustained production throughout the year, ensuring higher income. Additionally, it can be grown in both coastal and highland areas, making it a year-round crop, unlike other varieties limited to specific agricultural campaigns.
- **Resistance and Adaptability**: ÚNICA's resistance to diseases and adaptability to altitudes of up to 3,800 meters above sea level make it ideal for diverse regions, enabling cultivation in both the highlands and the coast. However, coastal producers cannot use seeds from their production and must acquire seeds from the highlands (Fonseca, C.).

These features provide ÚNICA with a competitive advantage over other varieties. Consequently, recent years have seen increased adoption of this variety. However, crop substitution has naturally occurred, where older varieties are displaced by newer, more productive ones, including ÚNICA (Salas, E.).

It is crucial to emphasize that the creation of this variety aims to enhance productivity, increase profitability and livelihoods, and ensure sustainable income for farmers. The intention is not to replace existing varieties but to complement the agricultural landscape. Therefore, the substitution of some crops by ÚNICA is a natural process.

Economic and Social Benefits

- **Increased Income**: While prices may fluctuate due to seasonal supply, the high production volume and rapid harvest cycles provide some economic stability. Many farmers have gradually increased the land dedicated to this variety, leveraging "production windows" to maximize earnings during periods of low supply (Salas, E.).
- **High Productivity**: Its year-round high yield allows farmers to profit even under adverse price conditions.
- **Income Diversification**: Stable production and continuous harvests offer additional income, positively impacting farmers' livelihoods.

Notably, small-scale farmers primarily cultivate this crop for sale, favoring native varieties for self-consumption (Castelo, M.).

Value Chain

Understanding ÚNICA's impact requires examining the potato value chain, which includes actors involved in production, processing, distribution, and marketing. Each stage adds value, connecting suppliers with demand.

- **Seed and Input Market**: This market encompasses suppliers of seeds, fertilizers, pesticides, tools, technical assistance, and financing. Seed quality is crucial for determining the final product's quality, prices, commercialization, and competitiveness. While seed prices are similar across potato varieties, ÚNICA producers in the coast cannot reuse seeds from their harvest due to potential long-term disease issues. Instead, they must purchase highland seeds, ensuring quality and better long-term yields (Fonseca, C.).
- **Production Market**: In Peru, approximately 711,313 potato producers exist, 83.9% of whom are small-scale farmers with less than five hectares of land (Castelo, M.). Small producers often sell their harvest to rural aggregators, who consolidate significant volumes (minimum three tons) for local or wholesale markets. In contrast, medium and large producers sell directly to wholesalers. ÚNICA producers are typically medium and large-scale farmers with landholdings exceeding five hectares, driven by industrial market demand (MIDAGRI, 2023).
- **Processing and Marketing Market**: The processing and marketing segment holds the most economic relevance, encompassing high-value transactions. It includes wholesalers and retailers distributing products to local markets and supermarkets, requiring specific quality standards. Transformative activities such as cleaning, sorting, and packaging for supermarkets or industrial processing into chips and purées diversify product uses and increase market value (MIDAGRI, 2023).

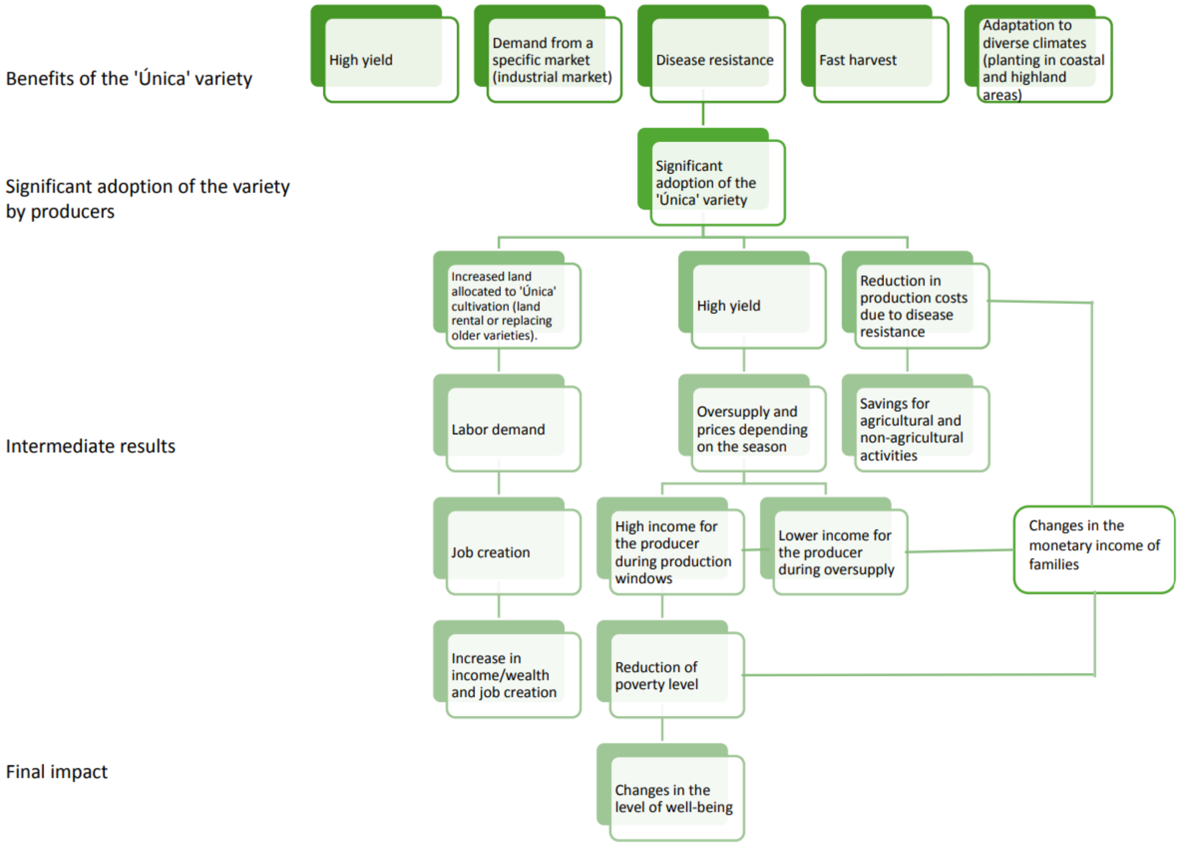
The interdependence of seed, production, and processing markets needs efficient coordination to maximize benefits and minimize losses.

Impact on Employment and Livelihoods

Greater adoption of the ÚNICA variety increases labor demand across markets. However, labor costs vary by region, influencing geographic disparities in job creation. Despite challenges like price drops during oversupply periods, improved livelihoods through increased income could lift small producers out of poverty, significantly enhancing their quality of life.

The Theory of Change presented in this project is based on a comprehensive analysis that combines information gathered from interviews, a review of existing literature, and data from relevant sources. This approach has allowed us to identify the expected outcomes and impacts of the adoption of the ÚNICA variety in the agricultural sector. The following graph illustrates the projected outcomes based on these insights.

Figure 5: Theory of Change. Source: Own elaboration, using information from the interviews conducted.



The Theory of Change outlined above emphasizes the expected transformations resulting from the adoption of the ÚNICA variety. Its high productivity, disease resistance, and adaptability offer significant potential to improve producers' livelihoods by increasing income and generating employment opportunities. The following table presents the project's indicators, outlining the immediate indicators and final outcomes expected from the adoption of this variety. These indicators reflect the changes in crop productivity, land use efficiency, livelihood improvements, and rural employment, which together will contribute to greater economic stability, and an overall improvement in the quality of life for farmers.

Project indicator matrix

The Poverty, Livelihoods, and Jobs (PLJ) indicator framework proposed by CGIAR will be used to evaluate impact indicators related to income, local job creation, and poverty reduction. The data collected will support the construction and adjustment of key indicators to assess the adoption of the ÚNICA variety.

This indicator matrix is organized into two main dimensions: intermediate indicators and final outcomes. The following tables provide the rationale and indicators for the components of our theory of change. First, we present the intermediate indicators, followed by the final outcomes, and lastly, some additional indicators.

Table 1: Project indicator matrix - Intermediate indicators.

Component	Rationale	Indicators
Increase in lands allocated to ÚNICA cultivation	Adoption of the ÚNICA variety due to its higher productivity and market demand encourages farmers to dedicate more farmland to this crop.	- Total area cultivated with ÚNICA (hectares).
Demand for labor	Higher production and faster harvesting cycles create greater labor requirements, increasing opportunities for employment in rural areas. Also, labor intensity can be measured through the average number of hours worked per week.	- Average hours of labor required per hectare. - Number of jobs created (workers/ha). - Hours worked per week in agricultural jobs.
Job creation	Job creation and wages in the agricultural sector are influenced by various factors. While the expansion in production and market participation generates additional employment opportunities across the agricultural value chain, wages in the sector vary seasonally and regionally. The demand for labor does not necessarily correspond to a fixed budget, meaning that an increase in employment does not imply a reduction in wages. Regional differences must also be considered.	- Number of jobs created in production, harvesting, and commercialization stages. - Wage levels (USD/day).
Reduction in production costs	The resistance of ÚNICA to diseases decreases the need for expensive pesticides and other treatments,	- Reduction in input costs per hectare (USD/ha). - Comparison of total production costs (USD/ha).

Component	Rationale	Indicators
	reducing production costs for farmers.	
Savings for other activities	Lower production costs allow farmers to allocate more resources to other agricultural or non-agricultural income-generating activities.	- Proportion of income reinvested in non-agricultural and agricultural activities (%). - Number of farmers diversifying activities.
High income during production windows	Shorter production cycles and higher yields enable farmers to generate significant income during periods of high demand.	- Seasonal income variation (%).
Low income during overproduction	Excess supply during certain seasons can lead to price drops, reducing income levels for farmers despite higher production.	- Average price per ton during overproduction periods. - Change in income during overproduction (%).

Table 2: Project indicator matrix - Final outcomes.

Component	Rationale	Indicators
Increase in household monetary income	The adoption of the ÚNICA variety contributes to a steady and higher income for households, improving their overall economic status.	- Annual increase in household income (USD/year). - Percentage of households above the poverty line.
Reduction in poverty levels	Improved income levels and expanded economic opportunities reduce poverty, lifting farming families out of extreme poverty.	- Number of individuals surpassing the poverty threshold ($3.65/day). - Number of people moving above the extreme poverty line of $2.15 per day.
Improvements in overall well-being	Higher incomes and job creation lead to better access to education, healthcare, and basic services, improving families' quality of life. In this context, purchasing power refers to the ability to afford essential goods and services, such as nutritious food, quality education, and healthcare. An increase in income allows families to make better financial decisions, reducing economic vulnerability.	- Proportion of household spending on health and education (% of income).

Table 3: Project indicator matrix - Additional indicators.

Component	Rationale	Indicators
Productivity: Crop yield (kg/ha) (crop-specific)	Increased crop yield can lead to increased food security for smallholder farmers, an increase in market surplus, and enhanced national food security.	- Crop yield (kg/ha) (crop-specific).
Productivity: Crop yield (rainfed vs irrigated systems)	Higher yields are generally seen in irrigated systems than in rainfed systems.	- Crop yield (kg/ha) (crop- and water regime-specific) (e.g., rainfed and irrigated).
Productivity: Yield gap	Reducing the crop yield gap can lead to increased food security and market surplus.	- Absolute or relative yield gap (kg/ha or %) (crop- and water regime-specific) (e.g., rainfed and irrigated).
Water productivity (WP)	Increased crop productivity with the same or less water input can improve water productivity, increasing food security.	- Water productivity (WP) (crop- and water regime-specific).
Profit and labor productivity	Increased labor productivity leads to higher profitability and more resources for investment.	- Value of production per capita (USD) (crop-specific).
Jobs: Agricultural employment	Employment in agriculture can increase rural economic activity.	- Employment rate in agricultural activities.
Jobs: Seasonal agricultural work	Employment during peak agricultural seasons (e.g., planting or harvesting).	- Proportion of the rural population engaged in seasonal agricultural work.
Jobs: Agricultural wage	Wages represent the purchasing power of agricultural workers.	- Daily agricultural wage (adjusted for purchasing power).
Jobs: Female agricultural employment	Proportion of women employed in agricultural activities, disaggregated by task.	- Female share of agricultural employment.
Labor productivity in agriculture	The value of agricultural production per worker shows labor productivity.	- Labor productivity in agriculture (USD per worker).

Methodology and evaluation design

Methodology

To evaluate the impact of adopting the ÚNICA variety on key socioeconomic outcomes such as household income, employment, and poverty reduction, this study employs Propensity Score Matching (PSM) and Difference-in-Differences (DID) methodologies. These methods are particularly suited for this context as they address challenges like selection bias, unobservable factors, and the absence of baseline data. Below, we detail the rationale and implementation to ensure robust causal inference.

Propensity Score Matching (PSM)

Propensity Score Matching (PSM) is a statistical method used to construct treatment and control groups with similar observable characteristics. It addresses selection bias by balancing these groups based on a propensity score, which is the estimated probability of adopting the treatment—in this case, the ÚNICA variety—given a set of observable characteristics (Rosenbaum & Rubin, 1983). This approach is particularly valuable in non-randomized studies, where selection into treatment is not random. In this study, PSM ensures that farmers adopting the ÚNICA variety (treatment group) are similar to non-adopting farmers (control group) across variables such as:

- **Socioeconomic Characteristics**: Household income, household assets.
- **Productive Characteristics**: Cultivated area (particularly potatoes), land tenure situation, access to technical services or markets/type of organization the household member belongs to or belonged to.
- **Demographic Characteristics**: Household size (number of members), age, education of the household head, and primary occupation.
- **Spatial Characteristics**: Distance to market, altitude, distance to irrigation source, distance to the nearest lakes, rivers, etc.

The estimated model is as follows:

$$logit(\pi_j) = log\left(\frac{\pi_j}{1-\pi_j}\right) = X_j\beta + Z_j\gamma$$

Where:

- π_j: The probability that a household in village j adopts the ÚNICA variety (1 if adopted, 0 otherwise).
- X_j: A vector of household-level observable characteristics, such as household size, age, and education level of the household head.
- Z_j: A vector of village-level characteristics, such as distance to the market, distance to the water source, or other spatial factors influencing adoption.
- β, γ: Coefficients to be estimated.

This equation estimates the propensity score π_j= 1, which is the predicted probability of a household adopting the ÚNICA variety based on its characteristics.

Once the propensity scores are calculated, households in the treatment group (adopters) are matched with those in the control group (non-adopters) using a 1:1 nearest-neighbor matching algorithm with a specified caliper. This ensures that the matched pairs are sufficiently similar in their observable characteristics, improving the validity of the causal inference by isolating the effect of adopting the ÚNICA variety.

Difference-in-Differences (DID)

Difference-in-Differences (DID) is a quasi-experimental method that uses temporal variations in treatment status to estimate causal effects. This approach compares changes in outcomes (e.g., household income, employment rates) over time between the treatment and control groups[4], accounting for unobservable characteristics that remain constant over time (Angrist & Pischke, 2009). The DID estimator is particularly relevant in this study because it allows us to isolate the effect of adopting the ÚNICA variety from other factors. This will directly answer questions such as: ***What changes in farmers' incomes can be attributed to the use of the ÚNICA variety***?

The general DID model is expressed as:

$$y_{it} = \alpha_i + \lambda_t + \rho D_{it} + \varepsilon_{it}$$

Where:

- y_{it}: The outcome variable (e.g., household income) for household i at time t.
- α_i: Household fixed effects, capturing time-invariant characteristics specific to each household.
- λ_t: Time fixed effects, capturing common shocks or trends affecting all households at time t (e.g., market-wide changes in potato prices).
- ρ: The treatment effect, representing the impact of adopting the ÚNICA variety
- D_{it}: A binary variable equal to 1 if household i adopted ÚNICA by time t, and 0 otherwise.
- ε_{it}: The error term.

The integration of PSM and DID provides a robust methodological framework. PSM ensures that treatment and control groups are comparable on observable characteristics, thereby reducing selection bias. DID complements this by accounting for unobservable, time-invariant factors and leveraging temporal changes to isolate the causal effect of adoption. Together, these methods address both observable and unobservable sources of bias, enhancing the validity and reliability of the results (Caliendo & Kopeinig, 2008).

[4] It is important to note that the information will be collected through surveys; however, the exact timeframe is yet to be determined. The database will be built based on the collected data.

Methodological limitations

1. Challenges in Ensuring the Validity of Comparability Assumptions

A central limitation is the potential challenge of ensuring the validity of the parallel trend's assumption, particularly for the DID method. This assumption requires that, in the absence of the adoption of the ÚNICA variety, the treatment and control groups would have followed similar trajectories in their outcomes over time (Rosenbaum & Rubin, 1983). However, this assumption may be compromised if producers in different districts or villages exhibit unobservable structural differences.

- **Inherent advantages**: Some districts may have better infrastructure, easier access to markets, or stronger social networks, which independently influence socioeconomic outcomes such as household income or productivity.
- **Environmental factors**: Differences in soil quality, climatic conditions, or irrigation access may affect productivity and income but remain unmeasured in the analysis.

If these unobservable differences are correlated with both the likelihood of adopting the ÚNICA variety and the outcomes of interest, the estimated treatment effects could be biased, reflecting these underlying characteristics rather than the true causal impact of adoption.

Mitigation Strategies:

- **Inclusion of Covariates**: Incorporate observable variables related to infrastructure, environmental conditions to account for potential differences.

2. Potential Spillover Effects

Another limitation arises from spillover effects, where the impact of the intervention extends beyond the treatment group to the control group. This is particularly plausible in rural, interconnected communities, where informal networks or seed exchanges may allow non-adopters to benefit indirectly from the ÚNICA variety. For example, farmers in geographically close control villages might gain access to seeds, techniques, or knowledge through informal interactions with adopters.

Such spillovers can dilute the estimated treatment effect by reducing the contrast between treatment and control groups, potentially underestimating the true impact of adoption.

Mitigation Strategies:

- **Geographic Separation**: Select treatment and control villages that are not geographically contiguous, establishing a minimum distance threshold to minimize interactions between groups.
- **Spatial Covariates**: Include variables capturing geographic characteristics, such as distance to other villages or seed distribution hubs, to account for potential spillover influences.

3. Unobservable Factors in Propensity Score Matching

Propensity Score Matching (PSM) relies on observable characteristics to balance treatment and control groups. However, it cannot account for unmeasured variables, such as individual farmer skills, risk tolerance, or informal market access, which may influence both adoption and outcomes. This introduces the risk of omitted variable bias, where unobservable factors confound the estimated treatment effect.

Mitigation Strategies:

- Comprehensive Covariate Selection: Include a wide range of socioeconomic, productive, and spatial variables to minimize the influence of unobservable factors.

Sample design

The objective of the sample design is to evaluate the impact of adopting the ÚNICA variety on the indicators selected previously. This will be achieved by comparing producer households that have adopted the variety (treatment group) with those that have not (control group), using the methodologies previously mentioned.

Unit of Analysis

A stratified design by adoption groups is proposed:

- **Adopters (treatment group)**: Producers who have adopted the new variety. Households that cultivate the ÚNICA variety along with other varieties.
- **Non-adopters (control group)**: Similar producers who have not adopted the variety. Households that do not cultivate the ÚNICA variety, but cultivate another potato varieties

Stratification and Sampling

For this study, stratification occurs before sampling. Villages are first divided into two groups based on potato adoption: adopters and non-adopters of the ÚNICA variety. This ensures that the sampling process captures variation between these two groups.

Sampling Process

This study follows a two-stage clustered sampling design at the village level:

1. **Village Selection (Primary Sampling Unit - PSU)**: From each adoption group (adopters and non-adopters), we randomly select a set number of villages.
2. **Farmer Selection within Villages**: Within each selected village, we randomly sample 10 producers. This sample size is based on previous studies and practical considerations, as it is uncommon to find villages with more than 10 producers cultivating the ÚNICA variety.

Randomization

Villages were first stratified into adopters and non-adopters, ensuring representativeness within each category. Then, villages were randomly selected from each group. This process helps minimize selection bias and improves comparability across groups.

Importantly, while the design involves treatment and control groups, this is not an experimental study with assigned treatments, but rather an observational approach to compare naturally occurring adoption patterns.

Operative sampling design

Data Collection and Targeting

First, we collected the microdata from the Profile of the Agricultural Producers Registry to identify the specific villages where producers cultivating ÚNICA and non-adopters are located. This data will help us determine how many districts are needed to reach our target.

Matching Process

After completing the baseline survey, where control villages are oversampled, the data will be aggregated at the village level. It is relevant to mention that the matching will be conducted at the village level, not at the individual producer level. This involves calculating the averages of socioeconomic, productive, and demographic characteristics, as well as including spatial variables like altitude, distance to main markets, etc.

Working at the village level justifies the inclusion of aggregated variables at this level. Based on this aggregated information, the matching is performed, and the final set of villages to be included in the study is determined.

Matching Variables

Household characteristics (observable), to be collected via a survey, including:

- Socioeconomic Characteristics: Household income, household assets.
- Productive Characteristics: Cultivated area (particularly potatoes), land tenure situation, access to technical services or markets/type of organization the household member belongs to or belonged to.
- Demographic Characteristics: Household size (number of members), age, education of the household head, and primary occupation.
- Spatial Characteristics: Distance to market, altitude, distance to irrigation source, distance to the nearest water source.

Sample

This section details the parameters and assumptions used to estimate the required sample size for the study. Key parameters include a significance level (α) of 0.05 and a statistical power of 0.8, both standard in the social sciences. Sensitivity analyses are conducted to account for variable parameters. Fixed assumptions include having 10 producers per village. Variable parameters include the intracluster correlation coefficient (ICC) and the minimum detectable effect (MDE).

Sample Size

To calculate the required sample size, the following parameters are considered:

- **Fixed Parameters:**
  - Producers per Village: Assumed to be 10 in both treatment and control villages.
  - Significance Level (α): 0.05, consistent with standard practice.
  - Statistical Power: 0.8, ensuring an 80% probability of detecting significant effects.
- **Variable Parameters**:
  - Intracluster Correlation Coefficient (ICC): We account for ICC, which measures the similarity between households within the same village. A higher ICC indicates greater homogeneity, requiring a larger sample size. For this analysis, we are relying on a previous study by CIP (2017), which used an ICC value of 0.08. To assess the impact of variations in this parameter, we test a range of values both above and below 0.08, specifically 0, 0.05, 0.1, 0.15, and 0.2. This approach allows us to observe how changes in ICC influence the sample size requirements.
  - Minimum Detectable Effect (MDE): Ranges from 1% to 10%, reflecting different effect sizes.
- **Output**:
  - The analysis estimates the required number of villages and households needed to achieve the desired power and significance level, identifying the most efficient sampling design to detect significant changes under different scenarios.

Sensitivity analysis

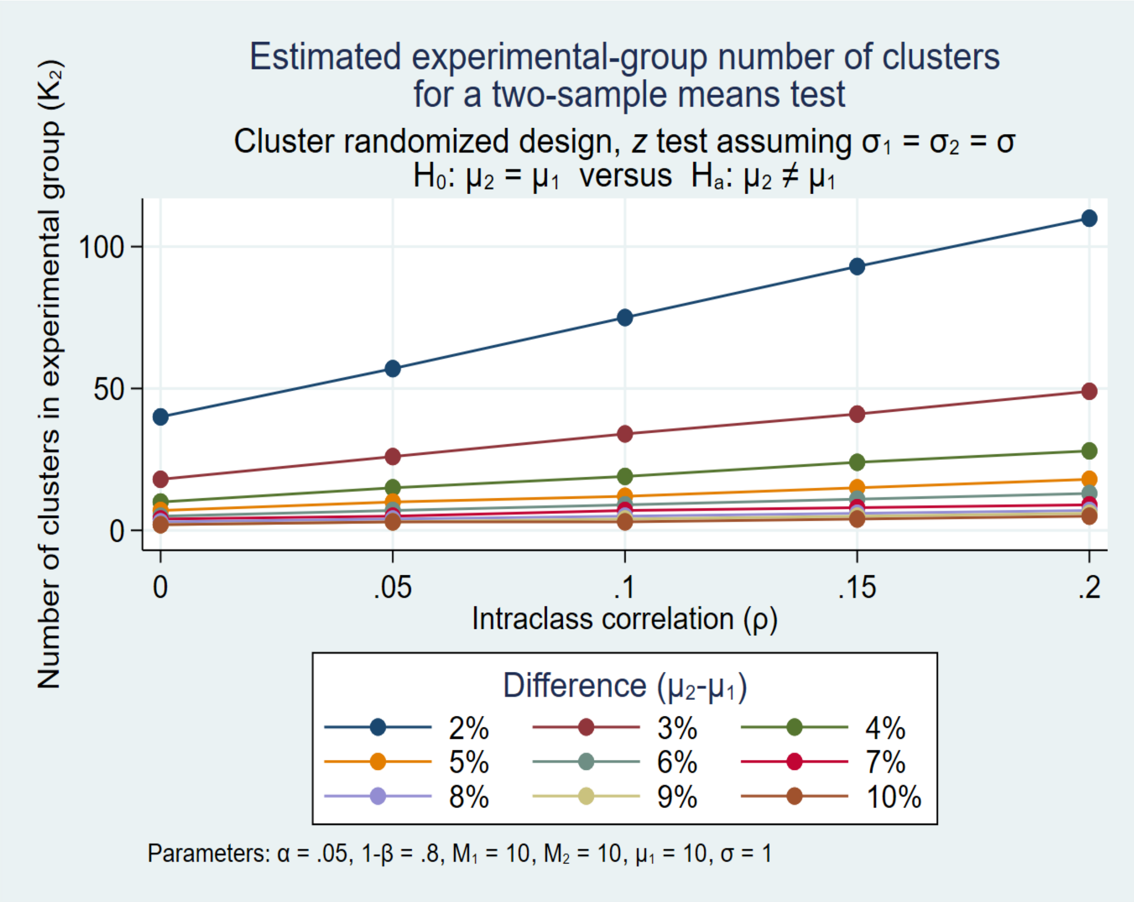
Figure 6 illustrates the estimated number of clusters required as the ICC value varies. This information is also summarized in a table located in the appendix section.

We then conduct a sensitivity analysis, assuming an ICC of 0.1, which is slightly more conservative than the previously mentioned value of 0.08. With a minimum detectable effect (MDE) of 3% and 10 producers per cluster, the analysis indicates that 34 clusters are required for each group.

It is important to highlight that in these types of evaluations, matching is conducted at the village (cluster) level, which can present certain challenges. A baseline survey is first conducted with 34 treatment villages, collecting data from 340 producers. However, for the control group, instead of using 34 villages, around 50 villages would be included.

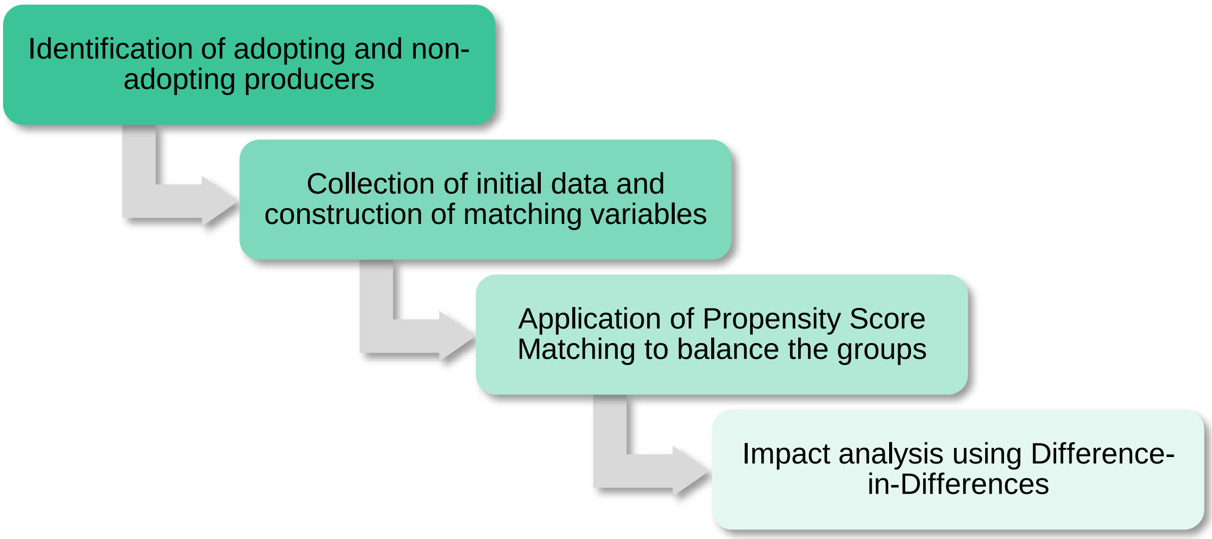
From this data, average values for key variables—such as the mean age of producers per village or the average land size per village—are calculated to perform the matching process. It is worth noting that some treated villages may not have a corresponding control village with similar characteristics, which is why a larger number of control villages is needed.

Figure 6: Estimated experimental number of clusters for a two-sample means test. Note: A statistical power of 80% was assumed, with a significance level (alpha) set at 0.05.



As a result, it is advisable to survey more producers in the first round to ensure robust matching. In subsequent rounds, once the matching has been refined, a more optimal number of producers can be surveyed.

Figure 7: Summary of the steps proposed for the study.



Alternative methodologies

While the proposed methodologies (PSM and DID) provide a strong framework for evaluating the impact of the ÚNICA variety, it is important to consider other potential methods. However, these alternatives face severe limitations that make them impractical to implement in this study. Below, we analyze three widely used causal inference approaches and explain why they are not feasible.

1. Randomized Controlled Trial (RCT)

Pros ✓	• Strong internal validity: Ensures that treatment and control groups are directly comparable due to randomization, eliminating selection bias. • Causal inference: Guarantees a valid counterfactual, allowing measurement of the true Average Treatment Effect (ATE). • Clear implementation: Results are easy to interpret as they minimize confounding factors.
Cons ✗	RCTs require randomly assigning the treatment (ÚNICA adoption) to one group and withholding it from another. However, this is unfeasible for the following reasons: • ÚNICA is already widely adopted: The variety has been available in Peru for over 30 years, meaning many farmers have already decided whether to adopt it or not. It is not possible to retroactively randomize adoption in areas where farmers are already growing it. • Contamination risk: Spillover effects (e.g., informal seed exchange) can reduce the contrast between treatment and control groups. • High cost: Implementing the intervention (e.g., distributing seeds) adds financial and operational burdens. • Limited adoption assurance: Unlike medical RCTs, where researchers can control who receives treatment, here, farmers decide on their own whether to adopt the variety. Even if seeds were provided randomly, there is no way to ensure farmers will plant them, meaning the study would measure intention to treat (ITT) rather than the actual adoption effect. Simply offering the variety does not guarantee adoption, which would result in measuring ITT instead of ATE.

Conclusion:

An RCT would not be a viable option because ÚNICA is already available, farmers cannot be randomly assigned to use or avoid it, and even if randomization were attempted, informal markets and personal choices would invalidate the results.

2. Spatial Regression Discontinuity Design (RDD)

Pros ✓	• Localized comparisons: Uses geographic boundaries to compare similar contexts, reducing biases from unobservable characteristics. • Natural experiment: Identifies causal effects by exploiting geographic thresholds where ÚNICA adoption differs.
Cons ✗	RDD relies on a well-defined geographic cutoff where some farmers are exposed to ÚNICA while others are not, creating a sharp discontinuity in adoption rates. This study does not meet that requirement for the following reasons: • Adoption is gradual, not abrupt: There is no clear geographic point where ÚNICA adoption suddenly changes from low to high. Instead, farmers adopt the variety incrementally based on individual decisions, market conditions, and access to seeds. Without a well-defined adoption threshold, RDD cannot be applied. • No external policy-driven boundary: RDD often works in cases where a government policy or external factor artificially limits access to treatment in some regions (e.g., subsidies granted only in certain districts). However, ÚNICA adoption has been entirely market-driven, with no external rule restricting or enforcing its use in specific locations. • Nearby villages have similar adoption levels: For RDD to work, adoption must be significantly different on either side of the threshold. In reality, neighboring villages tend to have similar levels of adoption, as farmers in close proximity share access to markets, seeds, and information. This lack of contrast makes RDD methodologically invalid. • Requires high-quality geographic data that may not be available: RDD demands detailed spatial data to define precise boundaries and ensure that other confounding factors do not drive differences in outcomes. Given that ÚNICA adoption data is primarily based on farmer self-reports and broad agricultural surveys, such granular data is unavailable, making implementation highly challenging.

Conclusion:

RDD is not feasible because ÚNICA adoption does not exhibit a sharp geographic discontinuity, there is no external rule creating a clear threshold, and the available data lacks the spatial precision required to apply this method.

3. Instrumental Variables (IV)

Pros ✓	• Addresses endogeneity: Helps identify causal effects even when adoption is correlated with unobservable factors. • Flexibility: Can be applied to observational data, provided a valid instrument is available (e.g., distance to ÚNICA seed distribution centers).
Cons ✗	Instrumental Variables (IV) estimation requires identifying an external factor (instrument) that influences ÚNICA adoption but has no direct effect on the outcomes of interest (e.g., income, employment). In this study, no such instrument exists for the following reasons: • Most potential instruments violate the exogeneity assumption: A common choice for IV could be distance to ÚNICA seed distribution centers. However, this variable is highly correlated with market access, infrastructure, and economic conditions, all of which directly impact income and livelihoods. Since the instrument must only affect adoption and not the outcome itself, this violates the fundamental assumption of IV estimation. • No external shocks or policy-driven variation to exploit: IV often works when a government intervention randomly assigns access to treatment (e.g., a sudden policy change that makes ÚNICA available in some regions but not others). No such event has occurred in the case of ÚNICA, meaning there is no natural instrument available. • IV estimates the effect only for a subset of farmers: Even if a valid instrument existed, IV would only estimate the Local Average Treatment Effect (LATE)—the effect on farmers whose adoption decisions were influenced by the instrument. This may not be generalizable to all farmers, limiting the study's broader applicability. • Results are highly sensitive to model specification: The validity of IV results depends entirely on how the instrument is chosen and modeled. If the instrument is weak or not truly exogenous, the estimates can be highly misleading. Given the challenges in identifying a suitable instrument, IV would introduce more uncertainty rather than clarifying the impact of ÚNICA.

Conclusion:

IV cannot be used because there is no external shock or policy that creates an exogenous variation in adoption, any potential instrument would be correlated with key outcome variables, and the results would only apply to a limited subgroup of farmers.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Centro Internacional de la Papa. (2018). UNICA, una papa adaptable y productiva para los agricultores de todo el mundo. https://cipotato.org/es/blog-es/unica-papa-adaptable-productiva-agricultores-mundo/

Gutiérrez-Rosales, R. O., Espinoza-Trelles, J. A., & Bonierbale, M. E. R. I. D. E. T. H. (2007). UNICA: variedad peruana para mercado fresco y papa frita con tolerancia y resistencia para condiciones climáticas adversas. *Revista Latinoamericana de la Papa, 14*(1), 41-50.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86. https://doi.org/10.1257/jel.47.1.5

MIDAGRI. (2023). *Análisis de brechas de rendimiento en el cultivo de papa en el Perú, 1997-2023.* Ministerio de Agricultura y Riego. https://cdn.www.gob.pe/uploads/document/file/6305660/5543457-analisis-brecha_rendimiento-en-el-cultivo-de-papa.pdf

Pradel, W., Hareau, G., Quintanilla, L., & Suárez, V. (2017). *Adopción e impacto de variedades mejoradas de papa en el Perú: Resultado de una encuesta a nivel nacional (2013).*

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-21. https://doi.org/10.1214/09-STS313

World Bank Group. (2017). *Gaining momentum in Peruvian agriculture: Opportunities to increase productivity and enhance competitiveness*. Peru Agriculture Opportunities ASA. Agriculture Global Practice; Environment Global Practice. https://documents1.worldbank.org/curated/es/107451498513689693/pdf/P162084-06-26-2017-1498513685623.pdf

Annexes

Annex 1: Sensitivity Analysis

Clusters for treatment group	Clusters for control group	Minimum Detectable Effect	ICC	Clusters for treatment group	Clusters for control group	Minimum Detectable Effect	ICC
157	157	1 %	0	40	40	2 %	0
18	18	3 %	0	10	10	4 %	0
7	7	5 %	0	5	5	6 %	0
4	4	7 %	0	3	3	8 %	0
2	2	9 %	0	2	2	10 %	0
228	228	1 %	0.05	57	57	2 %	0.05
26	26	3 %	0.05	15	15	4 %	0.05
10	10	5 %	0.05	7	7	6 %	0.05
5	5	7 %	0.05	4	4	8 %	0.05
3	3	9 %	0.05	3	3	10 %	0.05
299	299	1 %	0.1	75	75	2 %	0.1
34	34	3 %	0.1	19	19	4 %	0.1
12	12	5 %	0.1	9	9	6 %	0.1
7	7	7 %	0.1	5	5	8 %	0.1
4	4	9 %	0.1	3	3	10 %	0.1
369	369	1 %	0.15	93	93	2 %	0.15
41	41	3 %	0.15	24	24	4 %	0.15
15	15	5 %	0.15	11	11	6 %	0.15
8	8	7 %	0.15	6	6	8 %	0.15
5	5	9 %	0.15	4	4	10 %	0.15
440	440	1 %	0.2	110	110	2 %	0.2
49	49	3 %	0.2	28	28	4 %	0.2
18	18	5 %	0.2	13	13	6 %	0.2
9	9	7 %	0.2	7	7	8 %	0.2
6	6	9 %	0.2	5	5	10 %	0.2

laterite

DATA | RESEARCH | ANALYTICS

From data to policy