MAY 29-31
AIM for Scale Weather
Partnership Convening

Iterative learning via A/B testing

Tomoko Harigya
Chief Economist and Director of Research
Precision Development (PxD)

PxD
PRECISION
DEVELOPMENT

AIM FOR SCALE

Objectives

Introduce the use of experimentation, or A/B testing, in designing and improving weather forecast services.

01 | Why design testing and what is A/B testing?

02 | Examples of A/B testing in large-scale services

03 | Step-by-step process and considerations

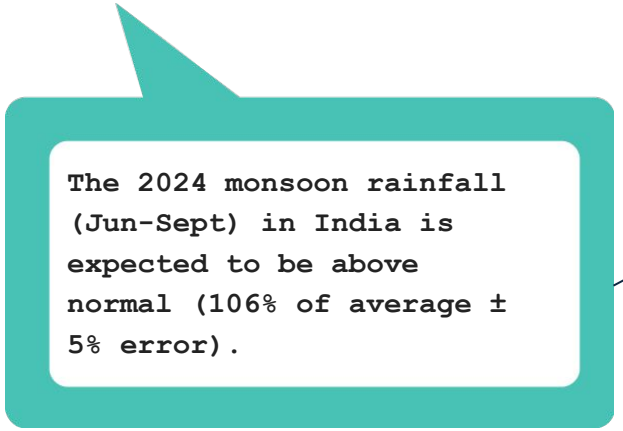
AIM FOR SCALE

Why design testing?

01 | Agricultural or climate information can be difficult to communicate precisely

02 | Small design tweaks matter

03 | Farmer reactions are context-specific



More detailed but not understandable for farmers

Less information, but farmers can understand

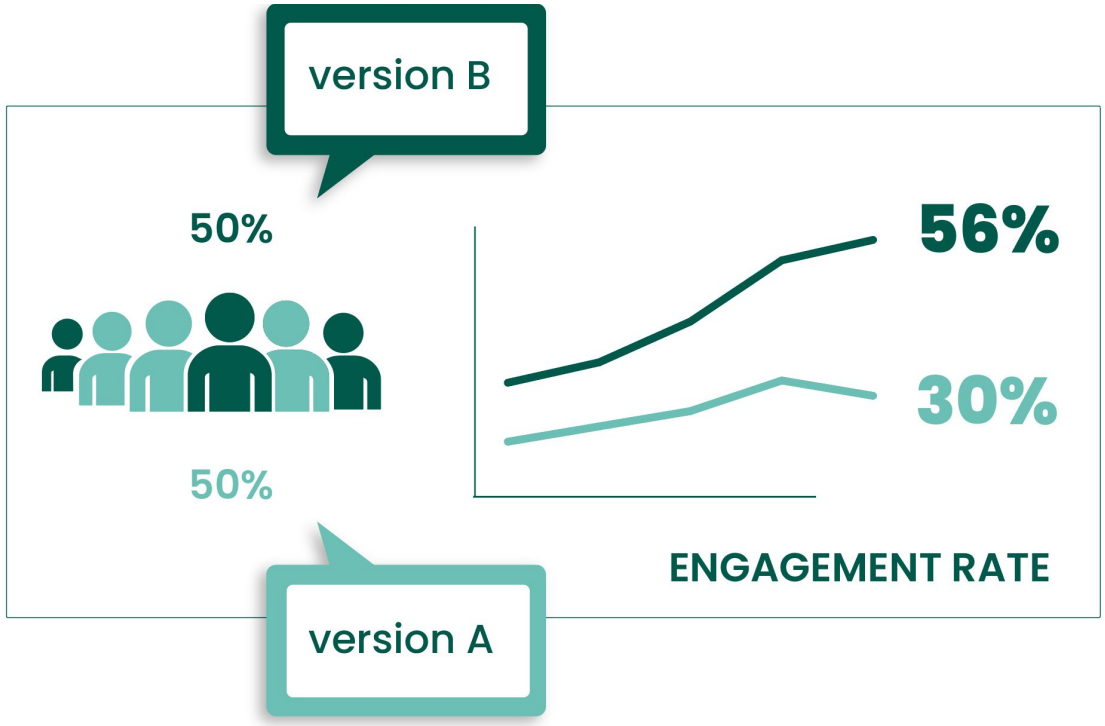
AIM FOR SCALE

A/B Testing: A Smart, Simple Way to Improve Services

What it is
Send two versions (A and B) of a message or service to randomly selected users.

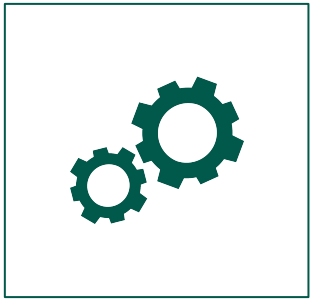
How it works
Track which version leads to better outcomes (e.g. message recall, comprehension).

Why it's useful
- Data-driven insights
- Clear link between design choice and outcomes
- High returns, low cost



AIM FOR SCALE

A toolbox for iterative design improvements



Usability testing

Watch real users interact with the service to identify pain points.

➢ **Quick insights from a small group before wider rollout**

User feedback

Hear directly from farmers about what's working and what's confusing.

➢ **Helps spot usability issues or gaps in understanding**

User data analysis

Track user behavior in the system (e.g., open rates, clicks, response times).

➢ **Uncover trends and drop-off points**

A/B testing

Compare two (or more) versions with real users to see what works better.

➢ **Generates clear, actionable evidence**

AIM FOR SCALE

Theory of Change: Digital information services

Access to service	Engagement	Recall/ Comprehension	Knowledge/ Perception (belief, trust)	Behavior change
# of registered farmers	# of active users Frequency and intensity of **successful** interactions	Immediate recall of content Accurate understanding of advice/information	Correct knowledge of optimal management practices Expectations for weather	Relevant agricultural decisions

AIM FOR SCALE

Examples from Precision Development (PxD)

Global non-profit organization serving 18M+ users in five countries in Africa and Asia

Offices in three countries

Over 100 experiments in partnerships with national and state level governments and other partners

Bangladesh
mPower

Ethiopia
Agricultural Transformation Institute

India
Ministry of Agriculture and Farmer Welfare
Odisha Dept of Agriculture & Farmers Empowerment
Coffee Board of India

Kenya
Ministry of Agriculture & Livestock Development
One Acre Fund

Nigeria
Federal Ministry of Agriculture & Food Security

Pakistan
Agriculture Department, Government of Punjab

AIM FOR SCALE

Ethiopia: Improved content access via 8028

Agricultural Transformation Institute

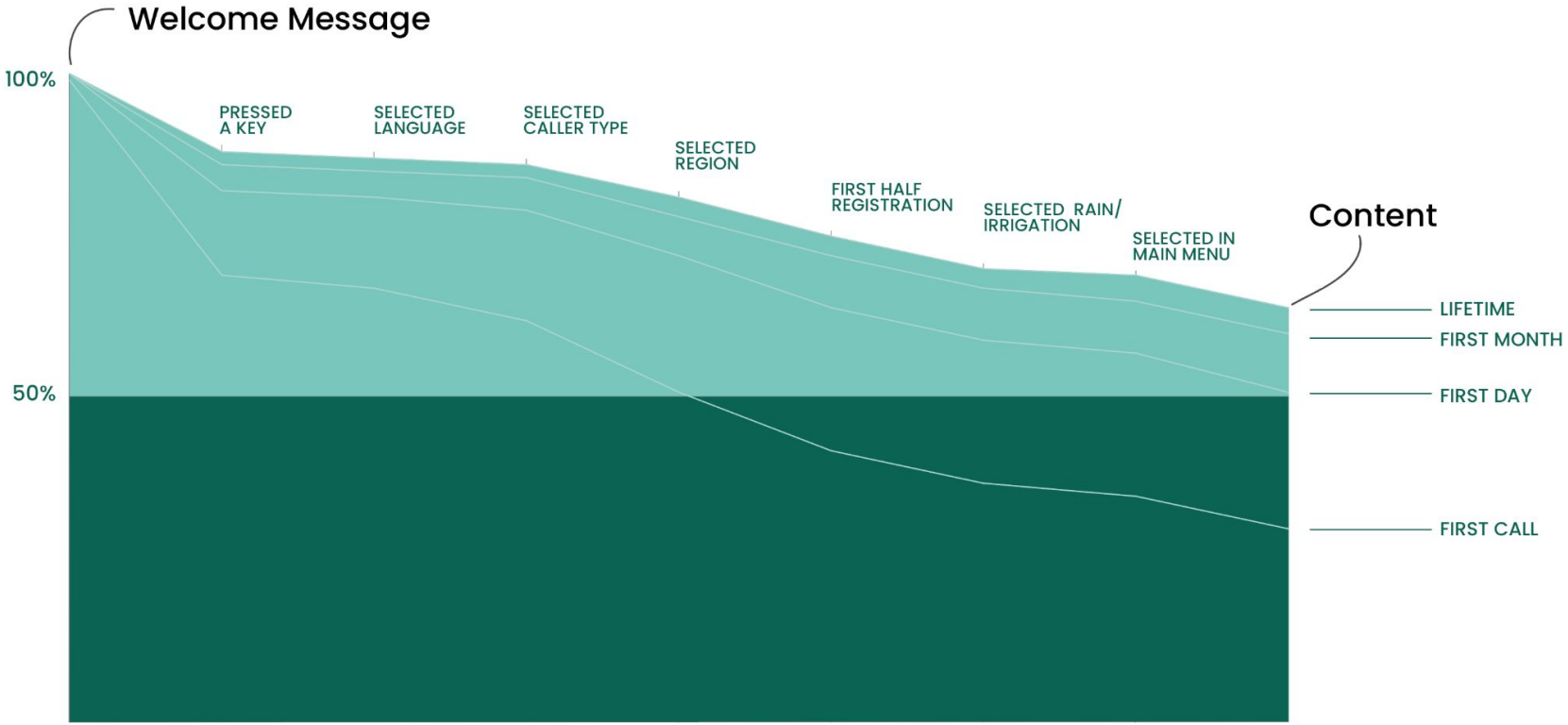
- A public institute within the MoA
- Established to promote agricultural sector transformation in Ethiopia

8028 farmers' hotline

- Launched in 2014
- 21 crops, 4 livestocks, financial literacy
- 67M calls & 6.96M callers
- 1.04M calls from 241K users in 2024

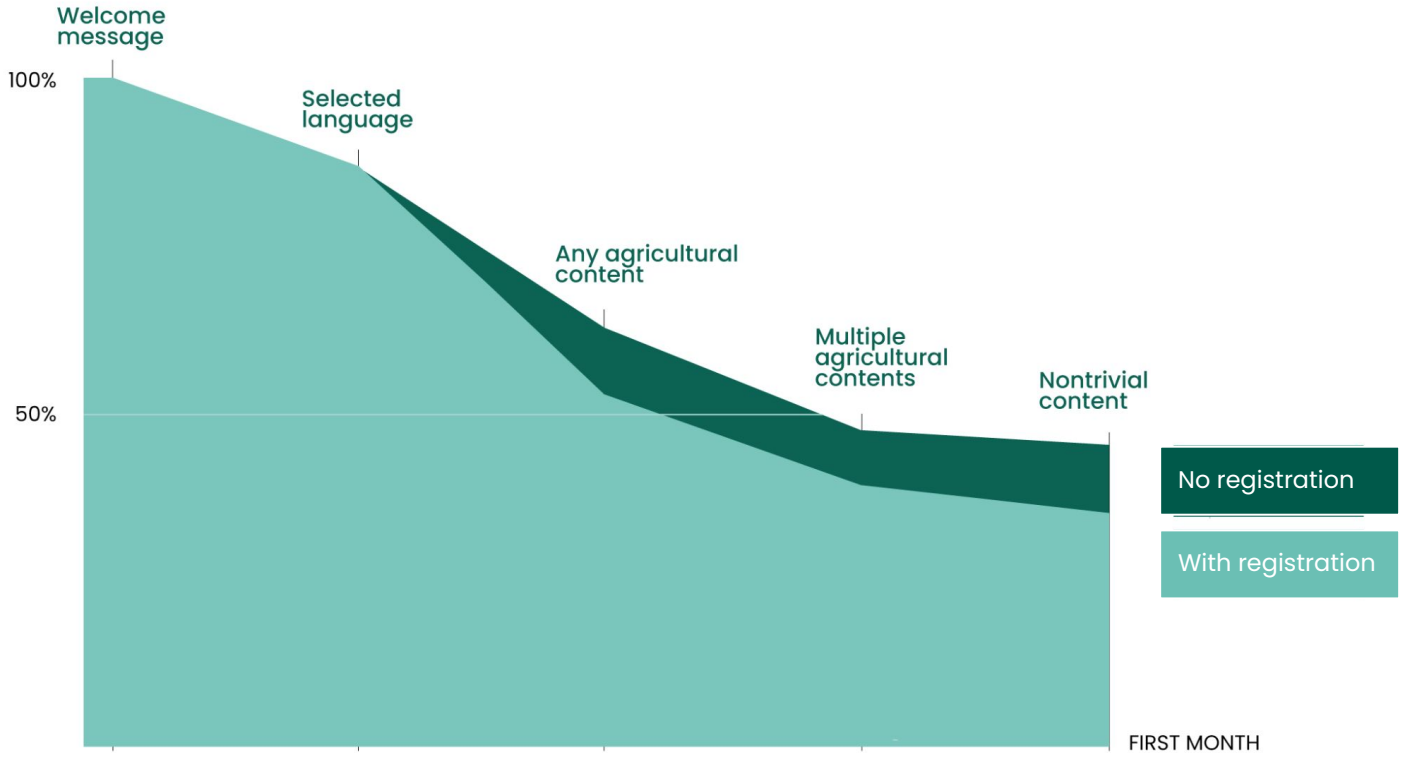
AIM FOR SCALE

User data analysis (1): < 50% of first-time callers reach agricultural content



AIM FOR SCALE

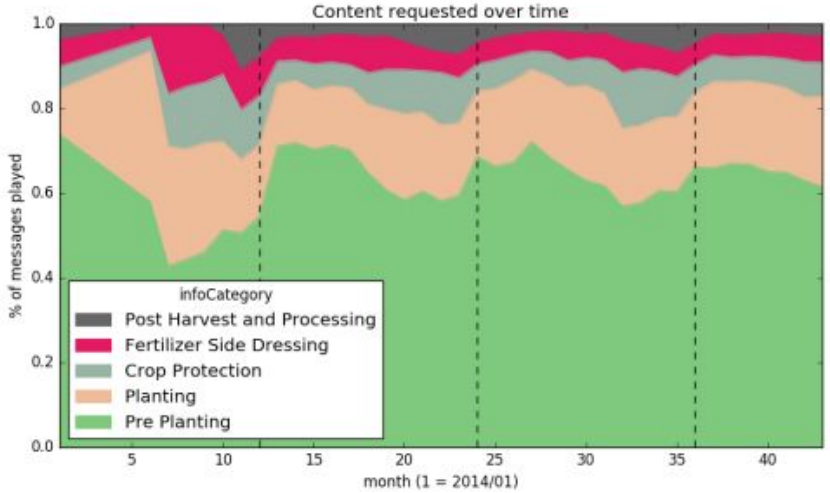
A/B testing (1): Postponing registration increases users' content access from 52% to 63%



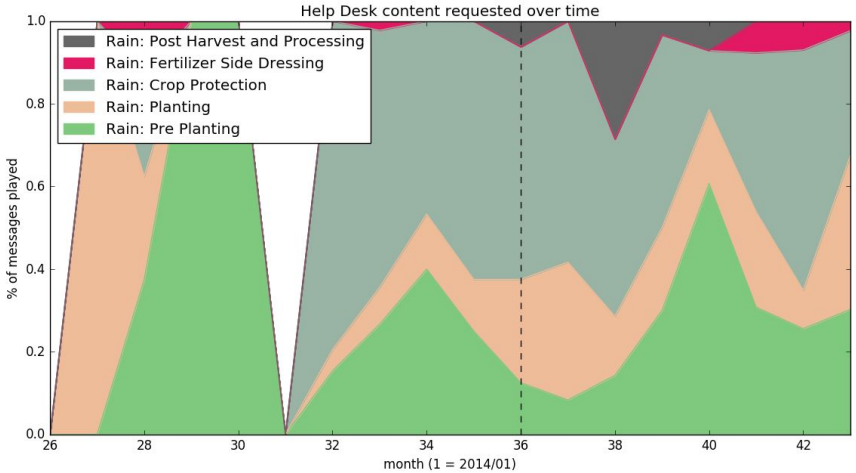
AIM FOR SCALE

User data analysis (2): Majority access pre-planting advice throughout the season

Users mostly choose preplanting (press 1) regardless of the season



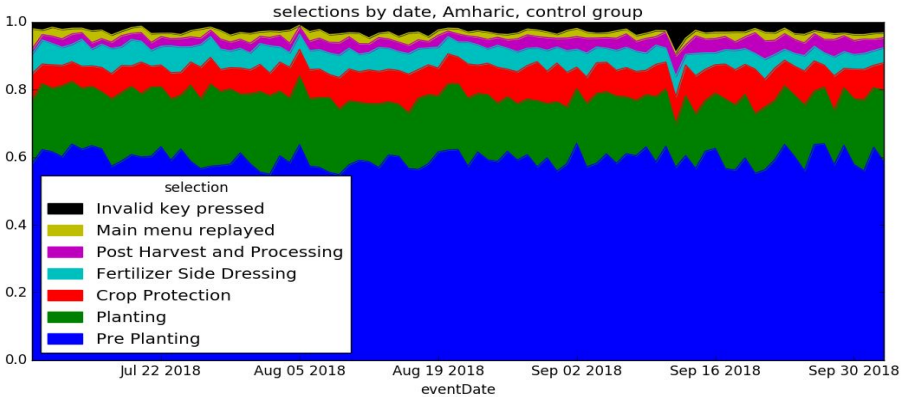
Help desk content illustrates what seasonal variation should look like



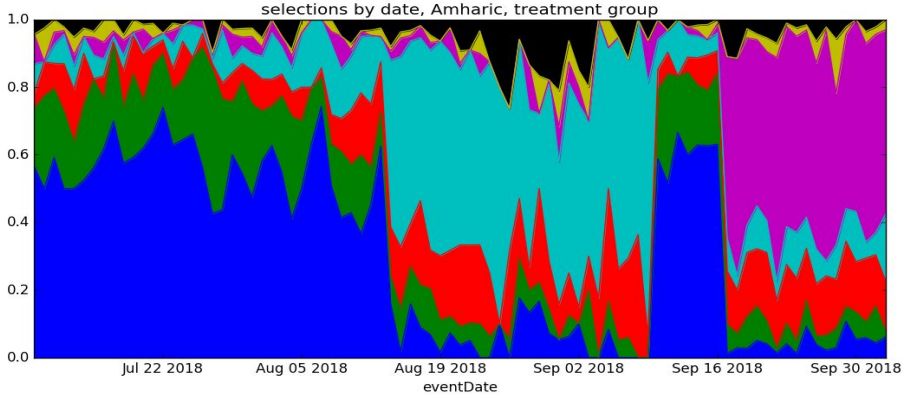
AIM FOR SCALE

A/B testing (2): Rotating menus increase access to seasonally relevant content

No rotation



Seasonal menu rotation



AIM FOR SCALE

India: Testing weather forecast messages

Context

- Digital advisory service by the Coffee Board of India
- Nearly 50K coffee farmers in Karnataka
- Weekly forecasts for 5-day cumulative rainfall

Design choice

- Deterministic forecasts: easier to communicate but doesn't indicate uncertainty levels
- Probabilistic forecasts: more complex, but useful information on uncertainty levels

A/B testing

Welcome to the CKT weather forecast service! For the next 5 days, in your village…

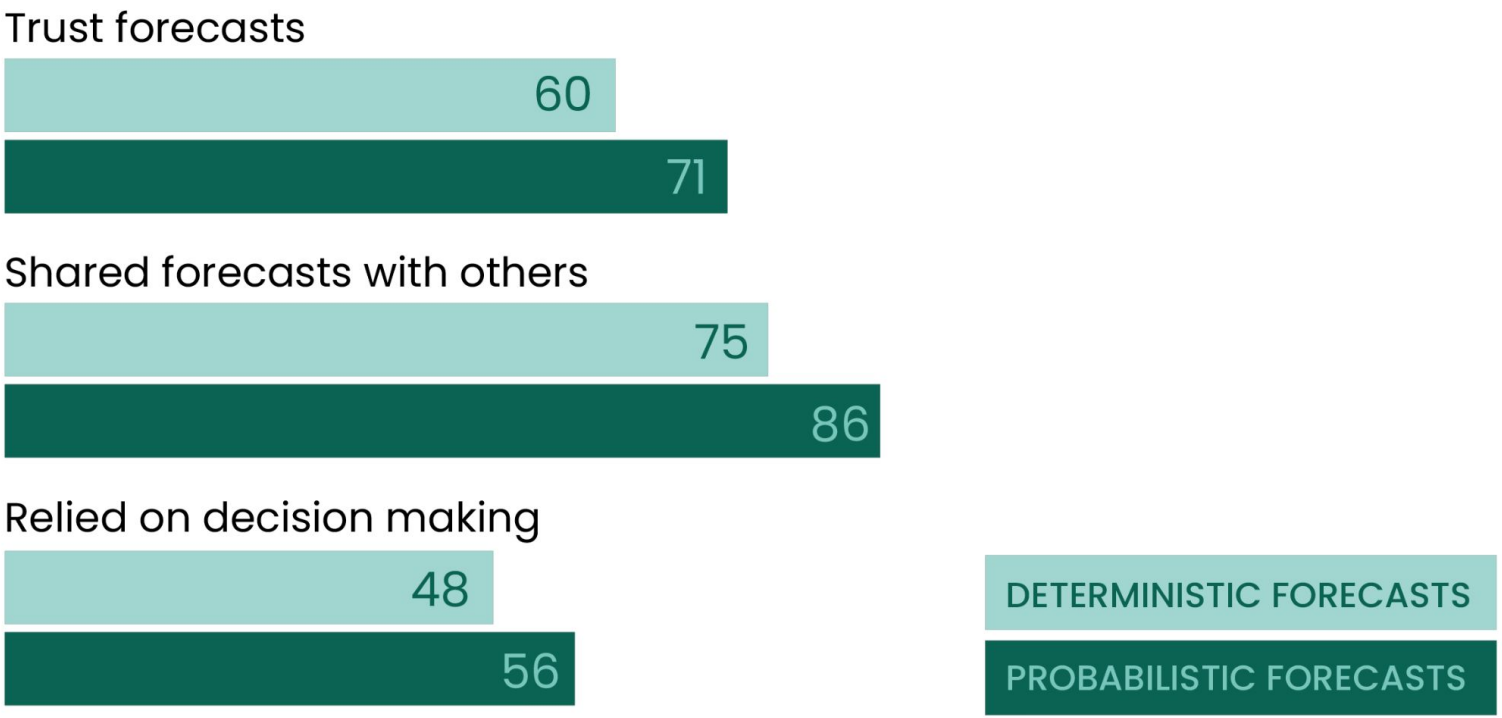
Version B. There is a 60% chance of 1 inch or more of rain. On average, rainfall of 1.2 inches is expected…

Version A. Average rainfall of 1.2 inches is expected…

Note: This A/B test was run In collaboration with V.Surendra

AIM FOR SCALE

India: Testing weather forecast messages (preliminary results)



Note: This A/B test was run In collaboration with V.Surendra

AIM FOR SCALE

How do we incorporate A/B testing?

Understand user patterns and feedback	What parts of the current service work well? Where do farmers have challenges?
Identify what to test and measure	Choose design elements (e.g., message format) and outcomes (e.g., comprehension)
Deliver the test	Identify the testing sample and randomize, and send out different message versions
Collect and analyze data	Compare outcomes across groups Assess the reliability of the results

Assess technical feasibility

Technological infrastructure: *Does the platform support your dissemination plan?*

Backend data systems: *Can the backend generate your engagement metrics?*

Data collection mechanisms: *Can the improvement be scaled across system?*

→ **Use insights for design decisions**

AIM FOR SCALE

Key messages

01

Design choices matter — Small tweaks can have big impacts

Even minor changes in wording, timing, or delivery can shift engagement and behavior significantly.

02

Our priors can be wrong — Test, don’t guess

What “should” work doesn’t always; Experiment helps identify what actually works

03

A/B testing can be low-cost, high-return

Large samples and use of administrative data allow cost-effective learning

04

Platform, backend and monitoring systems need a lot of attention!

Sustainability of data-driven learning depends on the capabilities of backend and monitoring systems

AIM FOR SCALE

Design elements for service improvements: examples

Dissemination channels

- SMS vs IVR vs WhatsApp
- Push messages vs. on-demand service
- Targeting extension agents, agro-dealers and/or engaged users

Communication design

- Frequency of forecast updates
- Message repetition
- Promotional messages
- SMS heads-up for incoming calls
- Voice of narrator

Forecast design

- Probabilistic vs. deterministic forecasts
- Lead time for forecasted events

Forecast communication

- Forecasts only vs. Forecasts + associated advice
- Accuracy of forecasts
- Annotations about how to interpret forecasts
- Mitigation messages in cases of wrong forecasts

Theory of change

AIM FOR SCALE

Example Worksheet 1: probabilistic vs. deterministic forecasts

Dissemination channel: voice messages

Specific forecast: 5-day cumulative rainfall

Design Question	Hypothesis	A/B Variants	Outcomes of Interest
How does the forecast format affect the use of forecasts?	Probabilistic forecasts are more effective than deterministic forecasts in building farmer trust and influencing farmer decisions.	Probabilistic vs. deterministic forecasts	Likelihood of listening to the message Trust in forecast Use of forecast

AIM FOR SCALE

Example Worksheet 1: probabilistic vs. deterministic forecasts

Access to service	Engagement	Recall/ Comprehension	Knowledge/ Perception (belief, trust)	Behavior change
NA	Proportion of farmers who listen to the message	NA	Trust in forecasts	Timing of fertilizer application Timing of harvest
NA	Farmer-level engagement data from the platform	NA	A phone survey	A phone survey

AIM FOR SCALE

Example Worksheet 2: probabilistic vs. deterministic forecasts

Technical/ operational feasibility	Feasibility Considerations	Notes/Concern
Technological infrastructure	The system needs to be able to send two different message versions to different groups of farmers at once	
Backend data systems	Need a call log that tracks details of the interaction	
Data collection mechanism	A survey to collect outcome data needs to be designed with the call center capacity constraints in mind.	

AIM FOR SCALE