Data collection for genetic improvement from smallholder

S.A Joshi, Y.S Gaundare, N.K Punde, T.V Shirsath, V.V Potdar, A.S Joshi, V.D, Santoshkumar, R. Jadhav, S.V Shinde and K. G Bhave

BAIF Development Research Foundation, Pune, Maharashtra, 412202 India
Corresponding Author: sachin.joshi@baif.org.in

Introduction

Smallholder dairy farmers play a critical role in ensuring rural livelihoods, improving food security, and supporting national economies, especially in developing nations like India. However, the full potential of smallholder dairy systems is often not realized due to various challenges such as limited resources, diverse environmental conditions, insufficient access to advanced breeding technologies etc. Genetic improvement can play a transformative role in enhancing dairy productivity, improving herd efficiency, and ensuring long-term sustainability.

The Genetic improvement program known as Enhanced Genetics Project (EGP), initiated by the BAIF Development Research Foundation, Pune, a significant move aimed at addressing these challenges by combining data-driven decision-making and genomic selection. The project developed a robust framework for systematic data collection system for genetic improvement in India’s smallholder dairy systems. This article explores the key elements of the data collection framework, its impact, and the future directions to further strengthen smallholder genetic improvement in systematic way.

Challenges in genetic improvement program of smallholder dairy systems

Implementing genetic improvement program in smallholder dairy systems poses several challenges due to the complexity and variability of these systems. Some of the key challenges include:

Small herd sizes and diverse breed compositions

Smallholder dairy farmers typically maintain small herds, often consisting of 2-5 animals per household. These herds frequently include a mix of indigenous breeds, crossbreds, and occasionally exotic breeds. Due to this diversity, designing genetic evaluation and breeding program become complicated. Moreover, small herd sizes result in limited data availability for assessing the genetic potential of individual animals, which slows down genetic progress.

Variability in management practices

Management practices vary significantly across different regions, influenced by local traditions, climatic conditions, and available resources. Feeding practices, housing conditions, and disease management systems often differ between regions, leading to genotype-by-environment (G × E) interactions that impact animal performance. These variations make it challenging to identify and implement uniform genetic improvement strategies across diverse agro-climatic zones.

Limited access to performance recording systems

Accurate performance recording is critical for genetic evaluation, yet most smallholder farmers lack access to such systems. Without reliable data on traits such as milk yield, fertility, and health, it becomes difficult to identify superior animals for breeding. Furthermore, inadequate infrastructure and lack of trained personnel to record and validate data often compromise the quality and consistency of the data collected.

Genotype-by-Environment (G × E) interactions

G × E interactions play a significant role in determining the performance of dairy animals. Animals that perform well in one environment may not necessarily excel in another due to differences in climatic conditions, feed availability, and subsequently difference in production system. Not accounting for these interactions may jeopardise genetic evaluation, ranking of bulls and the development of region-specific breeding strategies.

A framework for genetic improvement from smallholder

BAIF addresses these challenges by integrating a structured approach that combines advanced data collection systems, quality assurance protocols, and genomic selection to improve genetic gain in smallholder dairy systems.

Comprehensive data collection system

To overcome data limitations, the EGP has established an online platform-independent, digital data collection system that captures multi-trait phenotypic data. This system collects data on various economically important traits, including:

- **Production Traits**: Milk yield, fat percentage, protein content, SNF and lactose levels.
- **Reproductive Traits**: Calving interval, days open, insemination records, and pregnancy outcomes.
- **Body Weight and Linear Type Traits**: Body weight, structural soundness, and physical measurements.
- **Adaptability Traits**: Panting scores under diverse environmental conditions.
- **Behaviour** Traits: Milking Temperament

The data collection process is conducted at the farmer's doorstep by trained enumerators and artificial insemination (AI) technicians. This approach ensures real-time data recording, minimizes human error, and improves data accuracy.

The following table-1, table-2 and table-3 give an idea about how much data collected by BAIF since inception

Table 1. Data collected from smallholder across the states.

State	Milk recorders	CDC	Villages	Farmers	Animals
Andhra Pradesh	6	10	60	1,011	3,980
Bihar	7	15	120	2,637	6,827
Gujarat	17	21	86	2,337	8,586
Haryana	3	3	7	95	292
Jharkhand	6	37	187	1,708	7,482
Karnataka	3	3	16	415	1,272
Madhya Pradesh	7	7	22	408	2,284
Maharashtra	8	18	98	1,991	10,216
Odisha	8	15	152	3,192	6,655
Punjab	5	14	165	1,464	6,089
Rajasthan	9	11	55	782	3,021
Telangana	2	2	9	37	600
Uttar Pradesh	15	32	377	4,186	12,166
Total	**97**	**188**	**1,354**	**20,263**	**69,470**

Table 2. Phenotype data.

Trait-wise	No. of Animals	No. of Records
Test-day milk yield	69,470	11,42,032
Milk composition (fat, protein, SNF)	65,324	6,97,376
Body weight	51,056	3,78,403
Panting score (heat tolerance)	36,114	2,07,545
Linear Confirmation traits	7,450	7,567
Fertility traits	35,987	64,654

Table 3. Genotype data.

Genetic group	Number of animals genotyped
Purebreds Cattle	653
Crossbreds Cattle	15,575
Indigenous Cattle	4,502
Non-Descript Cattle	1,587
Buffaloes	3,744
Total	**26,061**

Ensuring data accuracy and quality

Data accuracy is a cornerstone of effective genetic evaluation. The program has implemented multiple layers of quality assurance to ensure the reliability of collected data. These measures include:

- Validation Checks at Multiple Levels: Real-time validation of collected data ensures that errors are identified and corrected promptly.
- GPS-Tracked Data Collection: Real-time GPS tracking of enumerators ensures that data collection is conducted at the designated locations, reducing the risk of false data entry.
- Use of Bluetooth-Enabled Devices: Bluetooth-enabled milk analyzers and weighing scales ensure consistent and accurate recording of milk production and animal weight data.
- Hierarchical Monitoring Systems: A multi-level monitoring system supervises data collection at regional and central levels to maintain data integrity and consistency.

Genomic evaluation

Phenotypic data collected through the data collection system is integrated with genotypic data from animals sampled across diverse management conditions. The project uses customized *GAU* and *MAHISH* SNP chips to genotype indigenous cattle and buffalo breeds. Key aspects of the genomic selection process include:

Efforts towards establishment of Genomic Reference Populations: A reference population comprising diverse breeds and traits has been established to facilitate accurate genomic predictions.

Computation of Genomic Estimated Breeding Values (GEBVs): A suitable genetic evaluation models are used to compute GEBVs, allowing for more accurate and faster identification of genetically superior animals.

Validation and Refinement of Genomic Models: Continuous validation and refinement of genomic prediction models ensure that the models remain effective across changing environmental and management conditions.

Key findings and achievements

The genetic improvement program has demonstrated remarkable success in applying genomic selection to smallholder dairy systems, resulting in significant achievements such as:

Extensive data collection and farmer participation

Over 74,000 animal performance records have been collected from 21,000 farmers across multiple regions. This large-scale data collection effort has enabled the development of high-quality phenotypic and genotypic datasets.

Development of high-quality genomic datasets

The combination of phenotypic and genomic data has enabled the development of multi-breed genomic datasets that facilitate accurate genetic evaluations. These datasets serve as a foundation for implementing genomic selection strategies across diverse agro-climatic regions.

Enhanced farmer engagement and adoption

The program has adopted a participatory approach by involving farmers in data collection and decision-making processes. Digital and in-person feedback mechanisms have been established to provide farmers with insights on their animals' performance and genetic potential. This approach has increased trust and motivation among farmers to adopt genomic technologies.

Future directions and policy implications

To build on the success of the program and further strengthen genetic improvement in smallholder dairy systems, future efforts should focus on the following areas:

Expansion of genomic reference populations

Expanding genomic reference populations to cover a wider range of breeds and agro-climatic conditions will enhance the accuracy of genomic predictions. This expansion will ensure that genomic selection models remain relevant and effective in diverse environments.

Development of AI-driven genetic evaluation models

Integrating artificial intelligence (AI) into genetic evaluation models can enable real-time prediction of breeding values and facilitate faster decision-making. AI-driven models can also identify novel traits that contribute to improved animal performance and resilience.

Strengthening farmer outreach and capacity building

Wider adoption of genomic technologies requires sustained efforts in farmer education and capacity building. Developing training programs and extension services that focus on genomic selection principles and data interpretation will empower farmers to make informed breeding decisions.

Policy support and institutional integration

Government and institutional support are crucial for scaling up genomic selection initiatives. Integrating genomic selection into national breeding programs and providing financial incentives for data recording and genetic improvement can accelerate the adoption of genomic technologies.

Conclusion

The genetic improvement program provides a scalable and replicable model for genetic improvement in smallholder dairy systems. By integrating advanced data collection technologies with genomic selection, the project has established a strong foundation for sustainable dairy development. Moving forward, collaborative efforts between research institutions, policymakers, and farmers will be essential to realizing the full potential of genomic selection in improving productivity, sustainability, and rural livelihoods.

Acknowledgements

We acknowledged the BAIF Senior Management’s support of Dr. B. K. Kakade, Dr. A. B. Pande and Dr. J. R. Khadse. We also acknowledge the extensive contributions of field level Data recorders, District level supervisory staff, and State level monitoring officers, for their tireless works and dedication in BAIF’s Enhanced Genetics Project.

We are thankful to the Bill and Melinda Gate Foundation for this grant under investment name: Scaling Genomic Selection Across the Indian Smallholder Dairy Sector.

List of references

BAIF Experience in Field Data Collection: Vinod V Potdar*, Bhave K, Gaundare YS, Khadse JR and Pande AB Department of Genetics, BAIF Development and Research Foundation, India. August 2017, 10.36876/ijas.1004

FAO. 2022. “The Role of Smallholder Dairy Farmers in Global Food Security.” Food and Agriculture Organization, Rome.

Mohammad Al Kalaldeh, Marimuthu Swaminathan, Yuvraj Gaundare, Sachin Joshi, Hassan Aliloo, Eva M. Strucken, Vincent Ducrocq and John P. Gibson . Genomic evaluation of milk yield in a smallholder crossbred dairy production system in India: . December 2021, Genetics Selection Evolution 53(1), 10.1186/s12711-021-00667-6

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. 2001. “Prediction of total genetic value using genome-wide dense marker maps.” Genetics, 157(4), 1819-1829.

Pryce, J. E., and Daetwyler, H. D. 2012. “Designing reference populations for genomic selection.” Journal of Animal Science, 90(6), 1643-1651.

VanRaden, P. M. 2008. “Efficient Methods to Compute Genomic Predictions.” Journal of Dairy Science, 91(11), 4414-4423.

Wiggans, G. R., VanRaden, P. M., and Cooper, T. A. 2011. “The genomic evaluation system in the United States: Past, present, and future.” Journal of Dairy Science, 94(5), 2529-2536.

Schrag, N. F. D., S. M. Godden, M. D. Apley, R. S. Singer, and B. V. Lubbers. 2020. Antimicrobial use quantification in adult dairy cows – Part 3 – Use measured by standardized regimens and grams on 29 dairies in the United States. Zoon. Pub. Health 67:82-93. (https://doi.org/10.1111/zph.12773).

Swets, J. 1988. Measuring the accuracy of diagnostic systems. Science 240:1285-1293. (https://doi.org/10.1126/science.3287615).

Timms, L. L. and L. H. Schultz. 1987. Dynamics and significance of coagulase-negative staphylococcal intramammary infections. J. Dairy Sci. 70:2648-2657. (https://doi.org/10.3168/jds.s0022-0302(87)80335-1)

Whyte, D. S., R. G. Orchard, P. S. Cross, T. Frietsch, R. W. Claycomb, and G. A. Mein. 2004. An on-line somatic cell count sensor. Automatic Milking, a better understanding. Wageningen Academic Publishers, Wageningen, The Netherlands. (https://doi.org/10.3920/9789086865253_001).