Designing a validation application for genetic and genomic evaluation systems in the New Zealand dairy industry

E. Ooi[1], M. Price[2], K. Stachowicz[1], P. Amer[1], R. Handcock[2], and M. Stephen[2]

[1] AbacusBio Limited, 442 Moray Place, PO Box 5585, Dunedin 9058, New Zealand

[2] DairyNZ Limited, Hamilton 3240, New Zealand

Corresponding author: eooi@abacusbio.com

Abstract

New Zealand Animal Evaluation Limited (NZAEL) is considering a transition from their pedigree-based genetic evaluation system to a single-step genomic evaluation system, both of which use BOLT and Helical software. Central to the successful implementation of this system is a robust validation process that ensures the reliability of genomic breeding values (GEBVs) compared to current traditional estimated breeding values (EBVs). To accomplish this task, NZAEL and AbacusBio began a collaborative project to design an automated validation pipeline and accompanying R Shiny application. The objective was to create a tool that efficiently assesses the performance of the new genomic evaluation system across more than 30 traits, focusing on flexibility, minimal user intervention, and applicability to various stakeholder needs. The design process began with a facilitated workshop aimed at defining the project's scope. Key outcomes included the identification of critical validation analyses and metrics, criteria for evaluating (G)EBV performance, and the selection of relevant focal groups for the initial validation. This approach prioritized the needs of preliminary stakeholders, while also considering the broader interests of the New Zealand dairy sector. A significant aspect of the project was differentiating between 'routine' validation analyses, which would be directly integrated into the application, and 'exploratory' analyses, which required additional resources. This distinction allowed for a more focused development effort and a clearer understanding of the project's deliverables. The result of this collaboration was a validation application that streamlines the identification of problems and communication with stakeholders. Our experience underscores the importance of a user-centric design process in developing scientific tools, highlighting the need for clear communication, stakeholder engagement, and flexibility in project management.

Key words: dairy, model validation, genomic evaluation, New Zealand

Introduction

NZAEL is considering transitioning from their pedigree-based genetic evaluation system to a single-step genomic evaluation system, both of which use BOLT and Helical software (Garrick et al., 2018). This includes the introduction of a strict filtering process to ensure that only high-quality data enters the genetic evaluation. Central to the successful implementation of this system is a robust validation process that ensures the reliability of GEBVs compared to current traditional EBVs. To accomplish this task, AbacusBio was asked to develop a system for validating the models.

While the scientific literature extensively covers various validation methodologies such as bootstrapping (Weller et al., 2003), linear regression models (Legarra and Reverter, 2018), and bias estimation (Hickey et al., 2008), less attention has been paid to the practical implementation of these methods. In our experience, validation serves dual purposes: development and communication. From a development perspective, the aim is to refine a genetic evaluation model to ensure it produces the best predictions of genetic merit available

within the constraints of available data and resources. Validation results are used to fine-tune model parameters and the pre-processing of data extracts. On the other hand, communication through validation seeks to gain the approval of key decision-makers, facilitating adoption of the new system and building trust among stakeholders.

In scenarios where a considerable financial commitment has been made and the evaluations are likely to face scrutiny, it becomes especially beneficial to engage an independent third party for the validation. Engaging an independent validator not only supports efficient development, by ensuring comprehensive and unbiased evaluations, but also facilitates high quality communication among stakeholders. Independent expertise is also useful for building in-house capability with fresh perspectives, for validation systems that are complex and extensive, and when an objective confirmation of model performance is crucial for improving stakeholder confidence.

Materials and Methods

Planning workshop

A planning workshop was organized to build a consensus around the intended design. A key objective was to narrow the scope of the project by distinguishing 'routine' validation tests – i.e., those essential for initial screenings of the model – from 'exploratory' analyses, which delve deeper into specific issues as they arise. The primary focus was on the project team's own needs as a key stakeholder, ensuring clarity and relevance in the validation process without being prematurely influenced by broader stakeholder requirements.

Nominal Group Technique is a group process used to explore problems, generate solutions, and assist with decision-making (Delbecq and Van de Ven, 1971). Applying this framework in the validation workshop allowed us to systematically explore diverse opinions and leverage the scientific expertise within the project team. Participants were asked to individually consider key questions before the meeting (e.g., 'How will we know the new EBVs are better than the old EBVs?'), submit their answers anonymously, and then engage in a structured review of all responses during the workshop. This facilitated consensus-building by allowing participants to see both commonalities and outlier opinions. This approach encouraged convergence towards a group norm but also allowed space for discussing and integrating divergent views effectively.

Once the group had agreed on the EBV characteristics which would be targeted for assessment, a similar approach was taken to explore the metrics and analyses which could be investigated. Participants were asked to individually list their preferred analyses, before collating the responses as a group and categorizing the results. This allowed common themes to emerge from the data rather than to be defined *a priori*, reducing the risk of being unduly influenced by the facilitator's personal biases. The group then discussed the 'trigger points' for exploratory analysis, establishing the criteria by which each test result would be considered a failure. This must be done before conducting the validation to avoid the temptation to 'change the goalposts' to suit the results. A similar process was undertaken to explore the population subsets of interest (i.e., focal groups such as young genomic bulls or validation heifers).

Finally, a group prioritization exercise invited participants to cast votes for their preferred metrics and analyses. This was essential due to the scope of the validation; it would not have been possible to implement every desired test or analysis with the limited resources available to the project team, and within the project deadline.

This resulted in a detailed plan outlining the key EBV characteristics, metrics and analyses, and focal groups required to assess the new NZAEL 3.5 models. This was an essential step towards determining whether the new genomic

EBVs would represent an improvement over the current pedigree evaluations.

Validation pipeline development

Once the plan was drafted, we needed a validation pipeline that could undertake the specified analyses. The development of the pipeline centered on the need for robust, automated processes that could handle the high volume of data inherent in a national genomic evaluation including pedigree information for 34.5 million cattle. The pipeline was developed in R and designed to seamlessly integrate with the output files produced by Helical, using the *aws.s3* and *data.table* packages. This ensured that data could be directly fed into the validation processes without manual intervention. The data files required for the validation pipeline are shown in Table 1.

Table 1: Data used in the validation pipeline

Data	Description
Animal information	Pedigree file including animal ID, dam and sire IDs, sex, birth year and breed
Phenotypes	Files containing phenotype and fixed effect data for each trait
(Daughter) yield deviations	Files containing the (D)YD data produced for each trait from models run on all available data
Full EBVs	EBVs and reliabilities for each animal, produced from models run on all available data
Truncated EBVs	EBVs and reliabilities for each animal, produced from models run on training datasets excluding the most recent 4 years of data.

Although the pipeline can ingest high throughput data for analysis, it produces memory-efficient outputs such as plots, tables, and summary statistics. The modular nature of the pipeline ensures that plot generation is separate from the components for analysis, allowing changes to be made to the display of data without needing the entire pipeline to be re-run.

R Shiny application

Comparing three models across five key characteristics, each with approximately five metrics, over eight breed categories and for more than 30 traits would require the project team to assess over 18,000 plots, figures, and tables. This represented a significant mental overhead for the project team, who were split across multiple geographic locations and had differing levels of familiarity with the R programming language.

To facilitate the process, we developed an R Shiny application with a strong focus on user experience, aiming to provide a clear, intuitive interface for users to interact with the validation data. A list of key features is shown in Table 2.

Table 2: Key features of the R Shiny validation app

Feature	Description
Validation-specific tabs	Each key EBV characteristic has a dedicated tab, aligning with the validation design.
Overview summary	Dynamic heatmaps use color gradients (red to blue) to summarize the validation results for each trait, highlighting the best model for each breed and metric.
Dropdown selection	Users can select different traits and focal groups to compare via dropdown boxes.
Version control	A changelog button provides updates on the app's version and recent changes, ensuring that users are informed of modifications.
Security	User credentials are required for login. The app also has two display modes, with a cleaner version for external stakeholders.
Accessibility	The app is globally accessible through a secure server, administrated by AbacusBio.
Performance	The app contains minimal on-the-fly analyses, relying on summarized outputs. This makes it highly responsive, allowing reviewers to make quick comparisons between traits, metrics, and focal groups.

Results and Discussion

Overview

At this stage, the validation pipeline and R Shiny application have been successfully used to compare pedigree and genomic evaluation models across 30 traits. The initial EBV characteristics identified by the project team as essential for assessment are shown in Table 3. The metrics and analyses used to investigate these are also shown.

A detailed explanation of all metrics and analyses is outside the scope of this paper. However, in general, a forward prediction approach was used, where (G)EBVs from data truncated by four years were used to predict daughter performances (Mäntysaari et al., 2010) or (G)EBVs produced from the full dataset (Legarra & Reverter, 2018).

Table 3: Key EBV characteristics assessed, along with the analyses

EBV characteristic	Metrics and analyses
Sense-making	Genetic trends (means and standard deviations)[1] Breed differences (violin plots)[1] Table of summary statistics (count, mean, median, standard deviation)[1]
Predictive ability	Regression of adjusted phenotypes (YDs and DYDs)[2] on truncated EBVs (intercept, slope, and correlation/accuracy) Quintile analysis (difference between the adjusted phenotypic performance of the top and bottom 20% of animals, ranked on their parent average EBVs)
Stability	Regression of full EBVs on truncated EBVs (slope and correlation/accuracy)
Bias	Difference between whole EBVs and truncated EBVs (mean bias)
Interbull suitability	Interbull trend tests 2-4 (DYD trend, EBV trend accounting for new daughters, Mendelian Sampling variance trend)

[1]For each model, separated by breed and sex
[2]Yield deviations (YDs) and daughter yield deviations (DYDs)

This validation process confirmed the superior performance of the NZAEL 3.5 genomic models for most traits, while for others, it identified areas of improvement. In these cases, the process was then used to confirm that adjustments to the model and data processing had the intended positive outcomes.

The robust validation design and wide range of analyses performed helped improve the project team's confidence in the performance of the NZAEL 3.5 genomic models. This increased confidence informed communications when seeking internal funding approval and presenting the project to external stakeholders.

The R Shiny application was also shared with international reviewers, providing an additional layer of objective and scientific expertise to the validation. Positive feedback on the application was provided by the reviewers, who commented on its ease of use, the inclusion of heatmap summaries, and the convenience of switching between different traits and focal groups. Where appropriate, we incorporated several reviewer suggestions directly into the application design, further improving the validation process.

Key learnings and challenges

It was essential to start the validation project with a plan. However, as with any complex task, the team made early decisions which were then reassessed after improving our understanding of the process. By reporting these decisions here, we hope to assist other readers in their own validations.

For example, the validation initially focused upon the use of yield deviations (YDs) and daughter yield deviations (DYDs) rather than raw phenotypes to assess the predictive ability of the EBVs. This was a practical decision to avoid the need to incorporate trait-specific fixed effects into the validation pipeline and worked well for most traits. However, due to the differences in data pre-processing between the models, and the fact that YDs are products of the models that we were assessing, it was

difficult to know which set of YDs was to be used for validating three different models. For this reason, it became necessary to undertake exploratory analysis to validate the YDs for some traits, which may have been avoided by focusing on the raw phenotypes from the start of the project.

We also needed a highly disciplined approach to development, to avoid incorporating unnecessary features into the application or pipeline. It was essential to keep referring to the plan and to remind the project team of the distinction between 'routine' and 'exploratory' analysis, to avoid a continually expanding codebase and overly complicated user interface. In some cases, findings from the exploratory analysis needed to be incorporated into the core pipeline; these two concepts lie upon a continuum, and it can be difficult to know where one ends and the other begins. However, gentle resistance to design suggestions originating from stakeholders who are not part of the target audience is almost always a useful general guideline.

Finally, careful specification of file names, missing values, and column names was also needed to ensure that model results would be compatible with the pipeline. This required clear communication between the modelling and validation groups. A good understanding of data pre-processing and model specifications was essential, both to focus our attention on trait-specific areas of interest, and to interpret anomalies in the results.

Conclusions

This project demonstrates the utility of a comprehensive, independent validation application developed by NZAEL and AbacusBio, aimed at enhancing the credibility and acceptance of genomic breeding values in the New Zealand dairy industry. By integrating an automated validation pipeline with an R Shiny application, the project exemplifies a more structured and transparent approach to evaluating genomic predictions.

The project also highlights the importance of clear communication and collaborative planning in validating genomic models. By distinguishing between routine and exploratory analyses and defining the target audience for the validation, our approach concentrates resources on areas of critical importance to the project.

Acknowledgments

We thank our colleagues and collaborators on the project team for their contributions, including Daniel and Dorian Garrick from The Helical Company. This work was supported by funding from New Zealand dairy farmers through the Dairy Levy, and the DairyNZ Resilient Dairy project, part of the Innovative Breeding for a Sustainable Future Programme funded by LIC, DairyNZ, and MPI.

References

Delbecq, A.L and Van De Ven, A.H. 1971. A group process model for problem identification and program planning. *J Appl Behav Sci*. 7 (4), 466-492.

Garrick, D.J, Garrick, D.P. and Golden, B.L. 2018. An introduction to BOLT software for genetic and genomic evaluations. *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand*. 973.

Hickey, J.M., Keane, M.G., Kenny, D.A., Cromie, A. R., Mulder, H.A., and Veerkamp, R.F. 2008. Estimation of accuracy and bias in genetic evaluations with genetic groups using sampling. *J Anim Sci.* 86, 1047–1056.

Legarra, A. and Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Set Evol*. 50 (1), 53.

Mäntysaari, E., Zengting, L. and VanRaden, P. 2010. Interbull validation test for genomic evaluations. *Interbull Bulletin.* 41, 17-22.

Weller, J.I., Emanuelson, U., and Ezra, E. 2003. Validation of Genetic Evaluation Methodology Using the Nonparametric Bootstrap Method. *Interbull Bulletin*. 31, 26-29.