Optimizing breeding schemes

Manual / Breeding process assessment

Genetic gain as a high-level key performance indicator

- This manual explains the concept of genetic gain and its relation to key performance indicators, shows computation methods, examples, and provides recommendations for its assessment in breeding programs.
- Published on 01/April/2020

  Excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals

Excellence in Breeding PLATFORM

ExcellenceinBreeding.org

Genetic gain as a high-level key performance indicator

Author.

Giovanny E. Covarrubias-Pazaran // g.covarrubiasl@cgiar.org

Breeding Optimization Lead, Excellence in Breeding Platform

Editors.

Abhishek Rathore // ICRISAT

Bert De Boeck // CIP

Brian Gardunia // Bayer CropScience

Dorcus Gemenet // CIP

Emily Ziemke // Corteva

Ibnou Dieng // IITA

Jerome Bartholome // IRRI

Johannes Martini // CIMMYT

Jose Crossa // CIMMYT

Juan Burgueno // CIMMYT

Sam Storr // Excellence in Breeding EiB

Valentin Wimmer // KWS

BILL & MELINDA GATES foundation

CGIAR | Excellence in Breeding Platform


CIMMYT
International Maize and Wheat Improvement Center


KWS


IRRI
International
Rice Research
Institute


BAYER
Bayer CropScience


ICRISAT
INTERNATIONAL CROPS RESEARCH
INSTITUTE FOR THE SEMI-ARID TROPICS


CORTEVA
agriscience


CIP
INTERNATIONAL
POTATO CENTER
A CGIAR RESEARCH CENTER


IITA
Transforming African Agriculture

Contents

Introduction

One key goal of any crop breeding program is to obtain high rates of genetic gain for traits of interest (i.e. identified in a product profile) while maintaining genetic diversity in order to sustain a high rate of adoption of improved varieties by end users. For genetic gain to result in higher rates of variety turnover, it is necessary for the product profile to be aligned with the needs of stakeholders (e.g. value-chain participants, farmers, consumers, funders, etc.).

It is common best practice to use performance indicators to ensure that activities are aligned towards achieving organizational goals by increasing transparency and accountability for relevant results. In breeding programs, it is possible to design quantifiable indicators such as the rate of genetic gain delivered for traits of interest, resulting in better products.

In the context of breeding programs that are publicly funded or where the development-delivery pipeline involves multiple stakeholders, greater transparency can result in improved ability to attract and maintain funding, coordinate development pipelines and communicate with end-users to improve adoption rates.

The goal of this manual is to demonstrate the use of the concept of “genetic gain” as a quantifiable key performance indicator (KPIs).

Key Performance Indicators

A Key Performance Indicator (KPI) is a measurable value that demonstrates how effectively an institution is achieving key business objectives. Organizations use KPIs at multiple levels to evaluate their success at reaching targets. *High-level* KPIs measure the overall performance of the business, while *low-level* KPIs measure individual processes. This manual will focus on genetic gain as a high-level KPI for assessing the overall breeding process.

In **Figure A**, breeding is represented as a *process*. The major subprocesses of breeding are categorized between design, engineering and scaled-production. At the same time, each of these subprocesses can be broken down into sub-subprocesses (i.e. the engineering component can be split between crossing, evaluation and selection). In this way, KPIs can be assigned to processes at different levels, For example, the engineering process can be evaluated with a *high-level* KPI, such as "rate of genetic gain", or the design process can be evaluated by "sales"; accordingly, subprocesses can be assessed with low-level KPIs, such as using "heritability (h^2) of trials" to measure the evaluation subprocess of the engineering process. Sub-sub processes can be assessed with even lower level KPIs, for example "number of plots planted per day" to evaluate planting, etc.

In this way, it is possible to systematically identify and apply KPIs across the breeding program and to the desired level of granularity. This manual focuses on applying the concept of "genetic gain" as an example of a high-level KPI at the overall process, level among other KPIs. In the next sections, the following topics will be covered:

1) The meaning and interpretation of genetic gain.
2) Different methods of computation, with examples.
3) Challenges to estimating genetic gain.
4) Recommendations for trial design allowing for effective assessment of genetic gain.

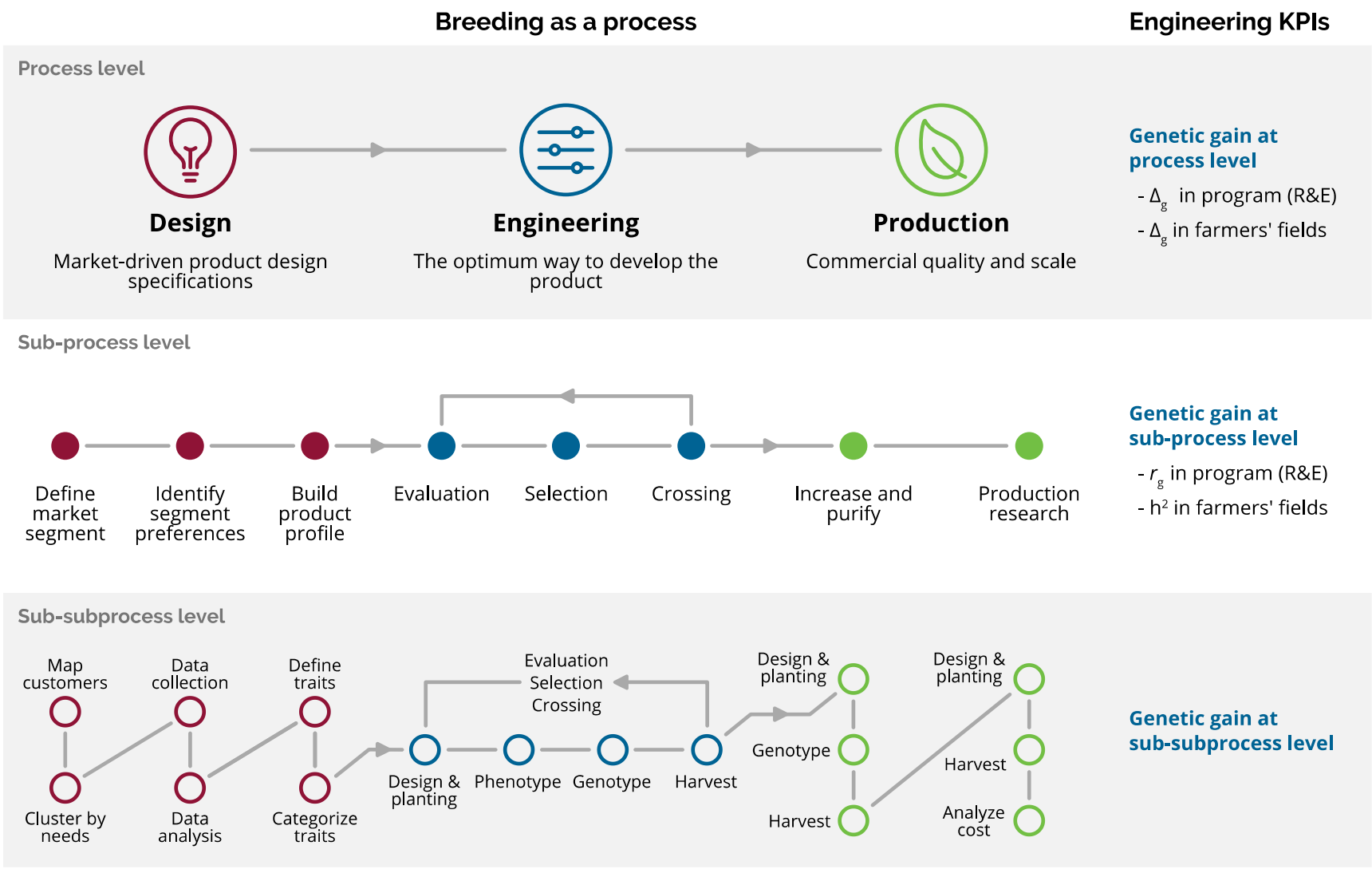

Figure A. Graphic representation of breeding as a process at different levels of detail. The major processes of breeding shown are design, engineering and mass production, each broken into subprocesses and sub-subprocesses. The nature and complexity of the KPIs that should be applied varies between the different levels of processes.

Rate of genetic gain

What is genetic gain?

Genetic gain has been described as the expected or realized change in average breeding value of a population over at least one cycle of selection for a particular trait or index of traits (Rutkoski, 2019a). This change is sometimes referred to as genetic trend and can be estimated by regressing the average breeding value on year or cycle when linearity exists (Eberhart, 1964). Assuming the breeding process remains unchanged and the trait of interest is quantitatively inherited according to the infinitesimal model (Fisher, 1918), this estimate can be used to predict future genetic gain. More extensive revisions of the concept of genetic gain can be found in Rutkoski (2019a, 2019b), Cobb et al. (2019), or classical books like Walsh and Lynch (2018). Here we opt for a high-level interpretation to make the explanation intuitive.

Lush (1942) capitalized on Wright's and Fisher's theoretical developments in quantitative genetics by defining the response to selection (denoted as R), which has become known as genetic gain (Hill, 2014). Lush's most known contribution is the "breeders' equation," which measures the response to selection as a change in average breeding value of a population. In broad terms, the most popular parameterization of response to selection is shown in **Figure B,** where the phenotype (**y**) can either be expressed as a linear combination of an intercept (**μ**), a genotype effect (**g**) and an error (**e**) (eq. 1a), or in terms of the genotype effect as a deviation of the phenotype from the intercept with a slope accounting for the error term (eq. 1b). If we remember that the slope of a regression of the phenotype on the genotype is equal to the heritability (h^2) (eq. 1c) it can be shown that the expectation ($\boldsymbol{\mu}_*$) of a selected individual ($\mathbf{g}_*$) in the parental generation is equal to the mean of the population in the offspring generation (eq. 1d, 1e). The difference between the mean of the parental and offspring generation is called the response to selection (**R**) (eq. 1h). Remembering that the

slope $\mathbf{b = h^2}$ and $\mathbf{S = \mu_* - \mu_p}$ we can see that the response to selection is the selection differential multiplied by the heritability (eq. 1h). Sometimes, the response to selection (R) is expressed in terms of the selection differential that is easily obtained by decomposing the heritability in terms of the genotype ($\boldsymbol{\sigma^2_g = \sigma_g \times \sigma_g}$) and phenotype ($\boldsymbol{\sigma^2_p = \sigma_p \times \sigma_p}$) variances (eq. 1i, 1j) (Walsh and Lynch, 2018).

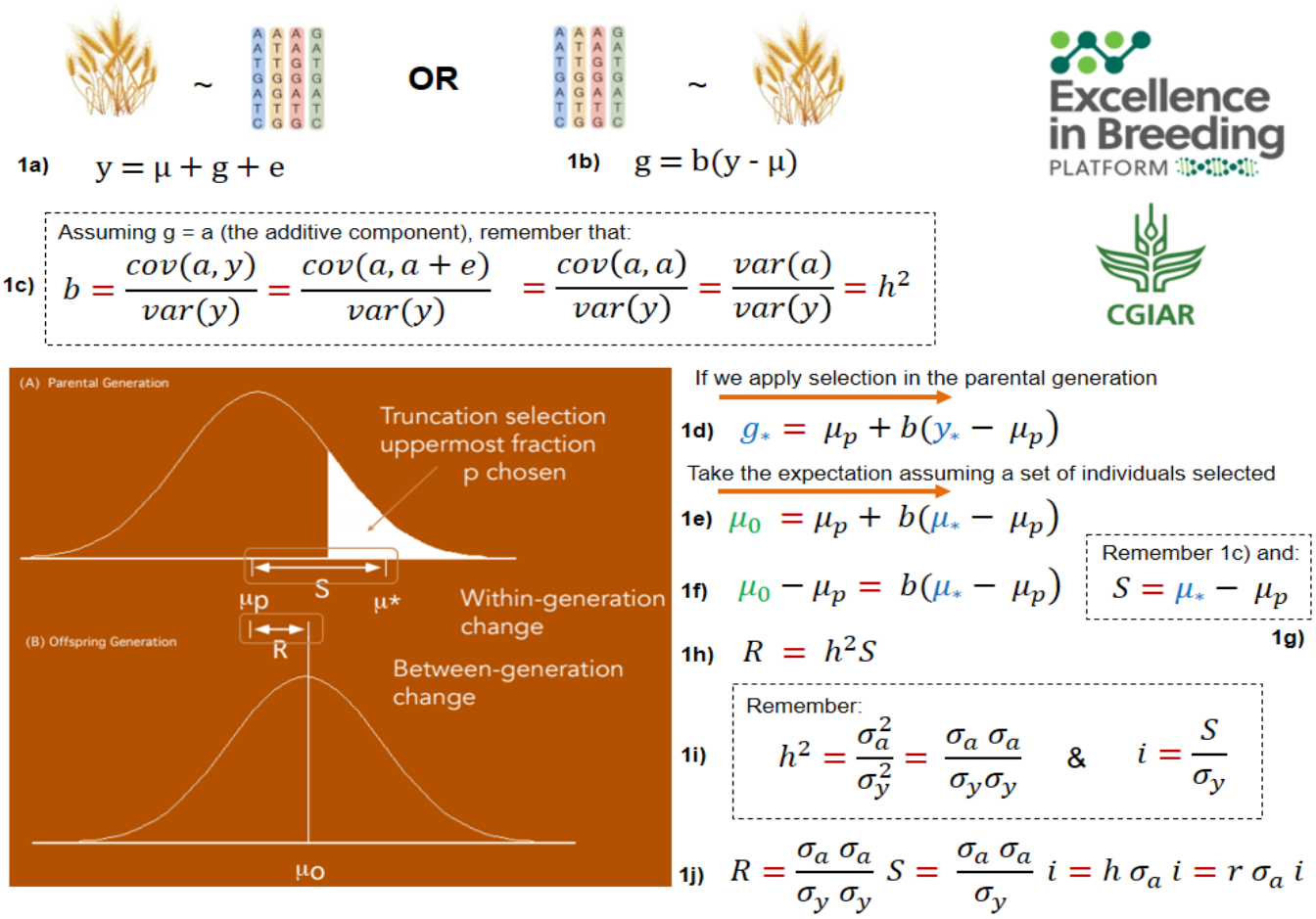

Figure B. Derivation of the breeder's equation in terms of selection differential times heritability and the standardized selection differential, accuracy and genetic variance. See main text for detailed explanation. g_*: genetic value of selected individual, μ_p: mean of the parental generation, μ_*: mean of selected population, S: selection differential, μ_o: mean of the offspring generation. Distribution plot taken from Walsh and Lynch (2018).

There are many other parameterizations for specific or more complex scenarios, such as response for a correlated trait, response when females and males are distinctively selected, and other scenarios (Mrode, 2014; Walsh and Lynch, 2018; Rutkoski, 2019a). In general, the single trait scenario with indistinct selection among males and females and a pure-additive nature clarifies the concept to understand more complex situations.

Methods to estimate the rate of genetic gain

The development of methods to estimate genetic gain is of interest for breeding programs and their stakeholders, as genetic gain constitutes a highly relevant indicator of breeding program performance and a means to compare different crossing, evaluation and selection strategies, either through real experiments or simulations (Cobb, 2019; Walsh and Lynch, 2018; Faux et al., 2016). The expected gain per unit of time (here denoted as L), usually referred to as the rate of genetic gain (Δ_g=R/L), is the most common way to express the gains of programs. In addition, cost is the main constraint applied to this function, being that increasing genetic gain at excessive costs is not optimal (Cobb et al., 2019). The methods to estimate the rate of genetic gain can be separated into *expected* (Falconer, 2005; Walsh and Lynch, 2018; Walsh, 2004) and *realized* genetic gain (Mackay 2011; Rutkoski, 2019a, Walsh, 2004; Laidig et al., 2014; Piepho et al, 2014). In **Table 1** we summarize the features of each method and the recommendations when using each of the different methods.

Table 1. Summary of methods to estimate the response to selection and rate of genetic gain.

Method	Formula used	Data required	Sample	Factors to be considered	Connectivity** / TPE* coverage	Recommendations
Expected	$R = h^2S = i\,r\,\sigma_g$ Lush (1942), Burrows (1972), Walsh (2004)	Any trial information	Any generation material	The heritability used will have an important effect in under- or over-estimating the metric.	Low after first selection cycle.	Use across-trials heritability. Do not use to take complex decisions
Realized	$y = X\beta + Z_d u_d + Z_g u_g + \varepsilon$ β: vector of fixed effects. u_d, u_g, ε: vector of random non-genetic, genetic and error effects. X, Z_d, Z_g: incidence matrices connecting observations with vectors of fixed and random effects Laidig *et al.* (2014), Piepho *et al.* (2014), Mackay (2011), Garrick (2010)	Era trial information	Early generation material	1) TPE* coverage is low (usually some locations & a couple of years). 2) Connectivity** among entries is maximum (cohorts evaluated at the same time). 3) Sample can overestimate the metric.	High / Low	1) Evaluate the material in representative environments for more than one year. 2) Use a replicated design. 3) Take a representative sample from each cohort if an estimate of evolution of genetic variance is required.
			Advanced material			
			Released varieties			
			On-farm			
		Historical trial information	Early generation material	1) TPE* coverage can be low (early), intermediate (advanced) or high (varieties). 2) Connectivity among entries depends on checks and the use of methods like EBV. 3) Sample can overestimate the metric.	Variable / Low	1) Use 4-10 checks depending on the stage*** to increase the connectivity of the data. 2) Use early generation trials for better estimate of evolution of genetic variance and advanced material for better estimates of the rate of genetic gain.
			Advanced material		Variable / Intermediate	
			Released varieties		Variable / High	
			On farm		Variable / High	

* TPE: Target population of environments. The better we cover it the more accurate estimates of the genetic and breeding value we can obtain. *** Connectivity: The degree of overlap of different cohorts in the same year. ** Stage: Refers to the stage of testing, from early to late. The earlier the more representative the sample is from the population. The rate of gain is less susceptible than the evolution of genetic variance to be biased depending on the sample.

Method 1. Expected (predicted) genetic gain

The expected genetic gain method uses the parameters from breeders' equation calculated for a single season to estimate the response to selection and infer the rate of genetic gain (Burrows, 1972). When parameters such as heritability, selection differential, selection intensity or genetic variance are known, these values can be simply applied in the formula shown in **Figure B** to obtain the expected response to selection or expected genetic gain (Falconer, 2005; Walsh and Lynch, 2018). This simplified approach has both advantages and disadvantages: it assumes that selection, evaluation and recombination units are the same breeding materials, selection is one-stage, there is no overlapping and, most importantly, it assumes that the response to selection will be the same in the future because genetic variance is assumed constant (Burrows, 1972). When these conditions are violated, more sophisticated formulas or methodologies are needed (Walsh, 2004).

This method provides a prediction and as such this should only be used as an indication that the program is moving in the right direction, but should not be considered as an accurate estimate of genetic gain. We recommend paying attention and indicate clearly which germplasm sample has been used to calculate this metric (i.e. early or late materials) in order to clarify with respect to which original population the selection differential has been calculated. For example, when calculating this metric using the late generation evaluation trials, the trait-mean of the original and selected populations are different than using early generation evaluation trials. Another recommendation is to obtain cross-environment heritability using a robust method like the one suggested by Cullis et al. (2006) to calculate a more accurate expected response to selection. More detail and examples to calculate heritability can be found in the EiB Heritability Manual.

Example

During the second stage of testing a program evaluates 1000 materials in 5 environments. After analyzing the multi-environment data for one trait of interest the best 100 individuals are selected and the following across-environment parameters are obtained:

$\mu_{p1000} = 5\ ; \sigma_p^2 = 2; \sigma_p = 1.41$

(original population phenotypic mean, variance and standard deviation)

$\sigma_g = \sqrt{\sigma_g^2} = 1;\ \sigma_e = \sqrt{\sigma_e^2} = 1$

(original population genotypic and error variance and standard deviation)

$h^2_{Cullis} = 0.5; \quad \sqrt{h^2} = r = 0.7; \quad i = 1.7549$

(heritability, accuracy and selection intensity)

$\mu_{p100} = 7.48; \quad S = 7.48 - 5 = 2.48$

(selection differential between original population mean and selected population mean)

The expected response to selection assuming a single stage selection, indistinctive selection among females and males, among other assumptions is:

$R = h^2 S = i\ r\ \sigma_g$

(response to selection under two parameterizations)

$R = (0.5)(2.48) = (1.7549)(0.7071)(1) = 1.24$

The expected (predicted) response to selection is 1.24 units, which means that the next generation is expected to have a population mean of 6.24 units for the trait of interest. If the response to selection needs to be transformed to a rate (per unit of time) it only needs to be divided by the cycle time of the breeding program which in this example-case, we assume a 5-year cycle:

$$\Delta_g = \frac{i\ r\ \sigma_g}{L} = \frac{1.24}{5} = 0.248\ units/year$$

Method 2. Realized genetic gain

The realized genetic gain method uses phenotypic data from time-representative samples of germplasm from a given stage of testing (early trials, late trials, varieties on registration trials, varieties on farmers' fields) evaluated either in the program across many years as the program evolves (historical data) or evaluated all together in an experiment (era trials). The data is used to fit linear models to infer the realized rate of genetic gain (Mrode, 2014; Mackay 2011; Rutkoski, 2019a; Walsh, 2004; Laidig et al., 2014; Piepho et al, 2014) (**Figure C**, **Table 1**).

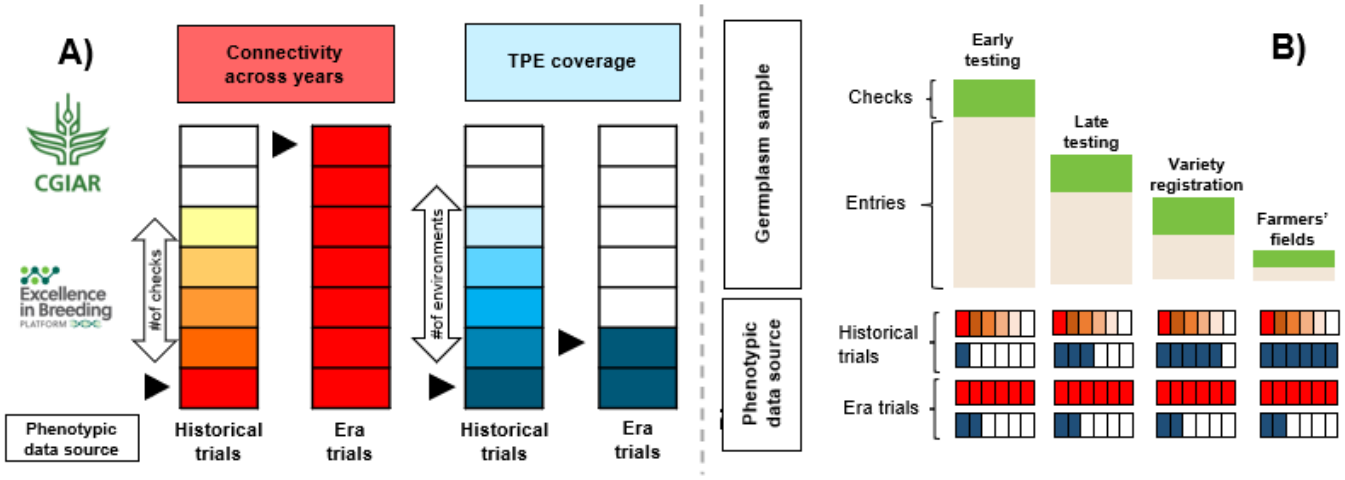

Figure C. Comparison of the target population of environments (TPE) coverage and connectivity between two different phenotypic data generation methods (historical and era) and variance for these parameters between different germplasm samples used for estimating genetic gain. In **A),** the difference in connectivity and TPE coverage for different phenotypic data generation methods (historical and era trials) is shown. Era trial information maximizes connectivity, while historical data depends on checks to have the same power. Era trial information tends to provide less TPE coverage while historical data provides greater TPE coverage. In **B)**, it is shown how these two parameters vary depending on the germplasm sample for a given stage used for the calculation.

Simulations (available with this document through a link) show that the methods used to obtain realized genetic gain that use either era trial or historical phenotypic information can provide an accurate estimate of the true rate of genetic gain, as long as two important features of accurate estimates are carefully considered: connectivity among time-window entries, and TPE coverage (**Figure D**). The decisions behind the calculation of the realized rate of genetic gain and recommendations on how to maximize connectivity and TPE coverage are given next.

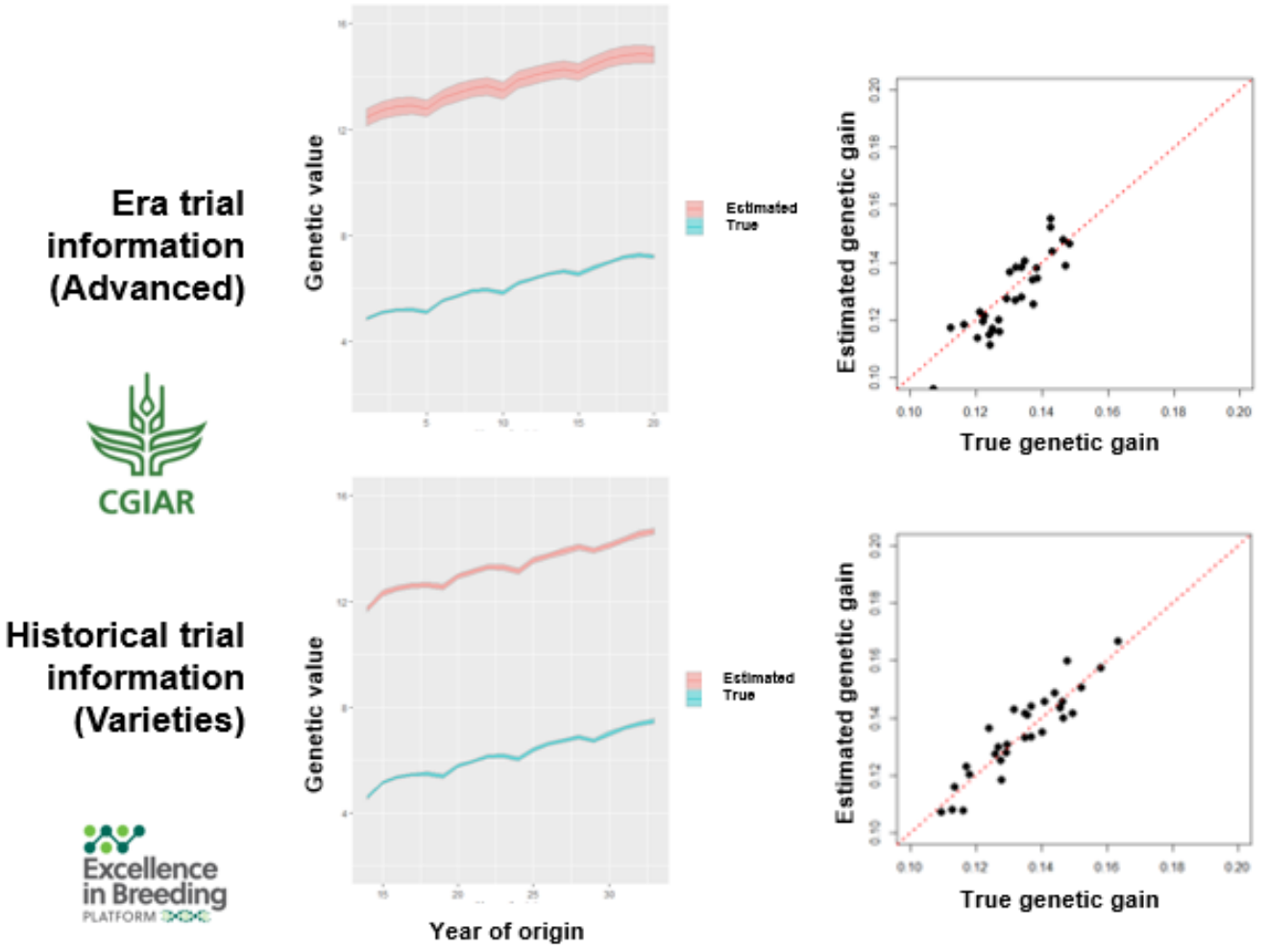

Figure D. Comparison of estimated versus true rate of genetic gain using linear models through different sources of phenotypic information (era and historical trial information). Colored lines (left) represent the estimated (red) and true (blue) rate of gain (genotype means regressed on year of origin), while colored shadows (left) represent the standard error of the values (based on 30 parallel simulations of the same initial population). Scatterplots (right) represent the $\Delta_g$estimates of the 30 parallel simulated programs.

The **first decision** that a program needs to consider when incorporating the rate of genetic gain as a KPI of the breeding process is to decide which **trait or traits** will be monitored. It is common to focus on the rate of genetic gain for productivity, namely yield, given that yield can be seen as an index of many traits of interest. The decision for which trait should be monitored must be driven by the final objective of the breeding program to be delivered to stakeholders.

The **second decision** refers to which **time-window** of materials will be used to estimate the metric. It could be that the program is only interested in demonstrating genetic gains in the last 5, 10 or 20 years. This is a decision that must be driven by the question that management is trying to answer with the calculation of this metric. For example, this is particularly important when comparing different breeding methods going on at different periods of time.

The **third decision** refers to the **sample of germplasm** from a given stage of testing that should be used for the calculation. Some examples of different stage samples include: the use of early testing trials (within program management), late testing trials (within program management), varieties on registration trials (outside program management), and varieties in farmers' fields (outside program management). Any such sample could be used, but each kind will have different properties that affect the accuracy of the Δ_g (connectivity and TPE coverage). It is common that early material (numerous) is only tested in a few environments for one year, while late stage material is tested in many environments for one or two years, material for registration (few) is tested in a wide range of environments for a few years and the material grown by farmers are tested in all the TPE for many years. This means that a Δ_g estimate based on late material will be more representative of TPE, whereas earlier material will be more representative of the breeding population. It is **recommended** that trials from all stages should be considered separately for the estimation of genetic gain KPIs based on several samples to obtain a better indication of the true Δ_g.

The **fourth decision** refers to the **locations and years** representing the TPE, where the sample of germplasm was tested. This introduces two of the most critical factors in Δ_g estimation: TPE-focus estimation and connectivity among time-window entries. The Δ_g needs

to be linked to a specific target represented by the TPE and not for all targets at the same time. It is also common that breeding programs do not have overlapping entries for any given stage of testing (i.e. preliminary yield trial) among different years because of the natural way that breeding programs work by cohorts. The problem with low connectivity is that the estimates of genetic or breeding value from the entries get confounded with the year effect. At the same time, breeding programs will generally have at least a small degree of connectivity insofar as check varieties do not change at the same rate as testing material from the program.

For the above reason, it is **recommended** that a good strategy to maintain connectivity among checks across years is put in place (**Figure 5**). In this check replacement strategy, checks may change over years, but not all at the same time so as to avoid the loss of connectivity. For example, in early testing with a large number of entries, it is possible to maintain eight checks as fixed varieties across years and change only two when new checks are needed in order to maintain the connectivity. A general rule for how many checks can be kept and changed can be based on simulations (Rutkosky, 2019b). It is **recommended** that, as a minimum, the number of checks maintained should be equal to the number of years taken to recycle parents, so that a program with an n-year breeding cycle requires a minimum of n checks to maintain sufficient the connectivity, while 2n and 3n checks is preferable (i.e. in the early testing trials). Is also important to decide how often the check replacement should occur. It is **recommended** that check replacement should happen after the same number of years that it takes to complete a cycle of recurrent selection at a rate of ¼x where x is the number of checks being grown (n, 2n, 3n). For example, a 4-year cycle program should replace ¼ of the checks after 4 years to maintain good connectivity (**Figure E**). In addition, it is always recommended to keep 1 to 2 checks that will never be replaced in order to keep a steady link among all trials from the program.

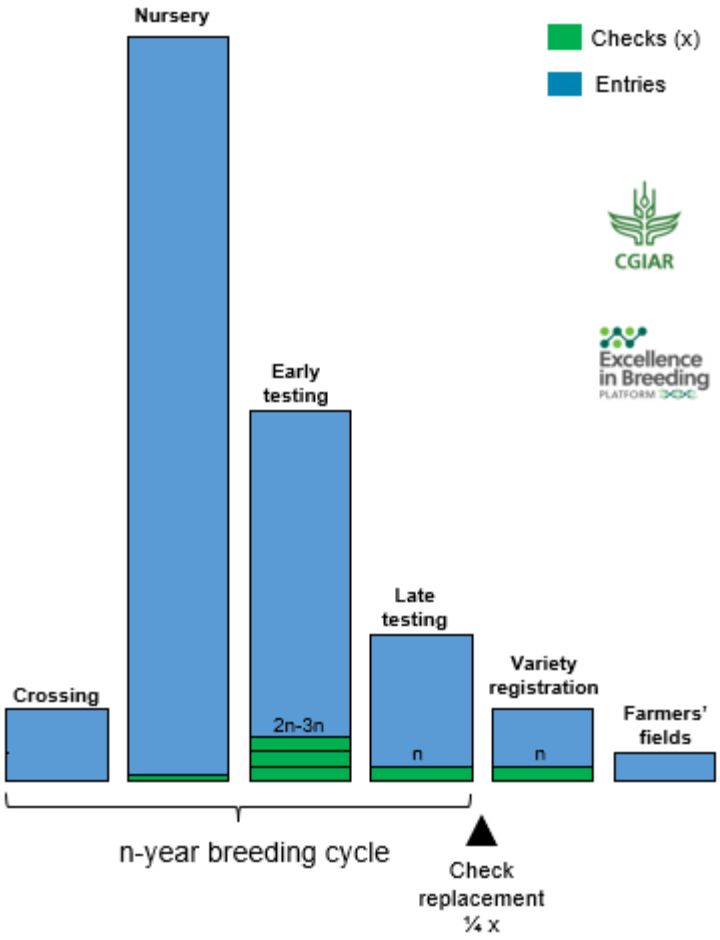

Figure E. Connectivity and check replacement strategy, where n represents the number of years that it takes to complete a breeding cycle. The earlier testing occurs, the more checks (between 2 and 3n checks) can be used to estimate genetic gain; later in the testing phase few checks can be used (i.e. n checks as a minimum). The check replacement strategy can be applied after n years at a rate of ¼ x, where x is the number of checks being grown at a given stage (n, 2n, 3n).

All told, the issue of connectivity can be addressed in three different ways:

1) Following the recommendations outlined above to ensure connectivity through checks.
2) Using the EBV method (Garrick, 2010) to connect the data of a program that didn't follow recommendations.
3) Running an era trial to maximize connectivity.

While it assumed here that the genetic gain KPI is calculated in the ideal way, using historical data from the program, many breeding programs will have not previously adopted the recommendations outlined above to achieve the proper levels of connectivity and TPE coverage required to obtain accurate estimates. In this case, the recommendations above should be immediately adopted, while in the meantime, connectivity can be addressed via options 2 and 3 as outlined above. In option 2, the EBV method can connect time-window data through an additive relationship matrix (pedigree or marker-based), although it should be considered that shrinkage of the estimates dependent on the heritability (h^2) can lead to underestimation of the Δ_g. In option 3, the program can also opt to run an explicit trial with different time-window material – an "era trial" – to remove the year confounding effect and maximize connectivity; the only disadvantage of taking this approach is the additional cost in time and money that the programs have to incur to obtain a baseline.

Example

Assume that a breeding program has followed the recommendations to achieve proper levels of connectivity and TPE coverage for the last 12 years. The program has a breeding cycle of 4 years, with approximately 4 cohorts that have been recycled between 2 to 3 times. The program has been storing information for the different stage materials and obtains trial information for the preliminary yield trials in the following format:

genotype	female parent	male parent	generation	location	phenotype	yearEval	yearOrigin
G232411	G181850	G182485	10	e20	10.79194	2005	2001
G232404	G181850	G182485	10	e20	11.77105	2005	2001
G231572	G172422	G182483	10	e20	9.980104	2005	2002
G232396	G181850	G181765	10	e20	7.884396	2005	2001
...	...	...	...	...	...	...	...

This data is then used to estimate the realized rate of genetic gain by taking across-location and across-year genotype means. These estimates can be accurately estimated despite the difference in years and locations because common checks exist across all these environments (year by location combination). The GxE interaction is also reduced by sampling the TPE as frequently as possible. The cross-environment genotype estimates are as follows:

id	predicted.value	std.error	status	yearOrigin
222243	12.37015	0.380793	Estimable	14
222254	12.02203	0.380793	Estimable	14
222479	11.52694	0.380793	Estimable	14
222567	12.55944	0.380793	Estimable	14
231413	11.28243	0.380792	Estimable	15
...	...	...	...	...

These final estimates can be merged with the year of origin of the material and perform a simple linear regression of the across-environment estimates as a function of the year of origin. Since the regression coefficient represents the rate of genetic gain in the original unit of the trait, using the estimates of genetic value (**y**) and year of origin (**x**) variables, the covariance of both variables and the variance of the x variable is calculated:

$$cov(x, y) = 4.033$$

$$var(x) = 33.585$$

$$\beta = \frac{cov(x, y)}{var(x)} = \frac{4.033}{33.585} = \mathbf{0.12\ units/year} = \Delta_g$$

Detailed examples based on simulated data and scripts are available with this manual in order to recreate the different scenarios of phenotypic data sources [historical information (both by maintaining connectivity using recommendations outlined here or by using the EBV method when connectivity is poor) and era trials]. A detailed explanation of the simulated examples available is provided in **Annex 1**.

Adoption strategy

To design effective KPIs, it is important to clarify the level of the process the KPI applies to, the method to calculate it and the baseline value. This manual has demonstrated the use of genetic gain as a KPI to evaluate the overall performance of the breeding process, how the KPI is derived and different methods to calculate it. In order to define which method should be used and derive the baseline value, it is important to consider the KPI adoption strategy.

Breeding programs should calculate genetic gain using phenotypic information obtained on a yearly basis. Therefore, the logical adoption strategy is to follow the recommendations outlined in this manual to maximize connectivity and TPE coverage of the yearly trials. All stages of testing should maintain the recommended number of checks and apply a check replacement strategy that maintains connectivity (**Figure F**). At the same time, to establish a baseline value for the KPI even if connectivity is poor, a decision can be made to use available historical data and apply the EBV method to connect data (using relationships between germplasm) or, if the available data cannot be used for this purpose, run an era trial to maximize connectivity and obtain an initial value of the Δ_g while the recommendations are adopted (**Figure F**).

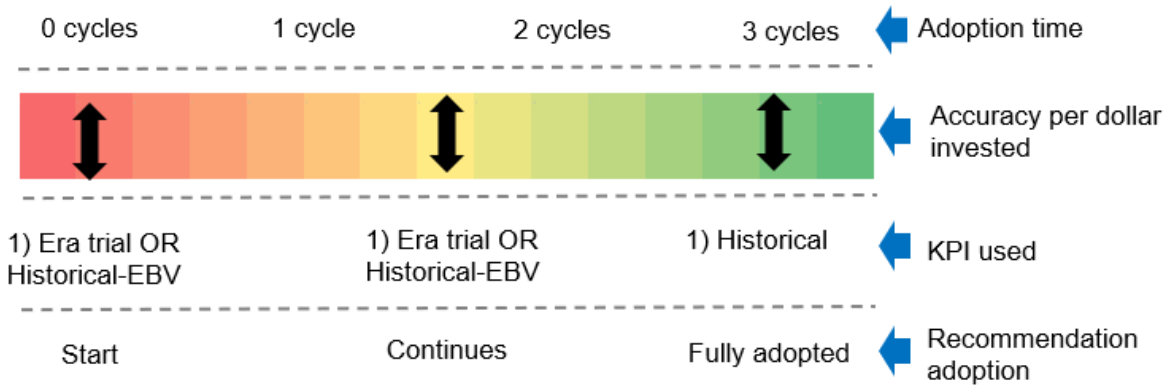

Figure F. Proposed adoption strategy of the genetic gain KPI metric in a breeding program. A phased approach is proposed as to increase the accuracy of the metric by adopting recommendations little by little in a time-bound window.

Conclusion

This manual introduced the concept of key performance indicators as a means to evaluate breeding program processes at different levels. The use of genetic gain as an indicator of the overall performance was proposed, and the derivation and calculation of this metric was demonstrated. Approaches for predicted and realized genetic gain were presented alongside recommendations for breeding programs to apply when adopting genetic gain as a KPI based on phenotypic data. Connectivity among years and locations and TPE coverage were presented as the main drivers of the accuracy of this metric. To overcome obstacles to adoption that may arise, different recommendations related to experimental design were proposed. A phased approach to increase the accuracy of this KPI per dollar invested was presented. Simulated data and sample scripts are available to allow breeding programs to recreate analysis and adopt this important metric.

References

1. Allard, R. W. 1961. Principles of plant breeding. Soil Science 91.6 (1961): 414.
2. Allier, Antoine, et al. "Assessment of breeding programs sustainability: application of phenotypic and genomic indicators to a North European grain maize program." Theoretical and Applied Genetics 132.5 (2019): 1321-1334.
3. Baker, L. H., and R. N. Curnow. "Choice of Population Size and Use of Variation Between Replicate Populations in Plant Breeding Selection Programs 1." Crop Science 9.5 (1969): 555-560.
4. Betran, F. J., and A. R. Hallauer. "Characterization of interpopulation genetic variability in three hybrid maize populations." Journal of Heredity 87.4 (1996): 319-328.
5. Burrows, Peter M. 1972. Expected selection differentials for directional selection. Biometrics (1972): 1091-1100.
6. Cobb, Joshua N., et al. "Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation." Theoretical and applied genetics 132.3 (2019): 627-645.
7. Cullis, Brian R., Alison B. Smith, and Neil E. Coombes. "On the design of early generation variety trials with correlated data." Journal of agricultural, biological, and environmental statistics 11.4 (2006): 381.
8. Faux, Anne-Michelle, et al. "AlphaSim: software for breeding program simulation." The plant genome 9.3 (2016): 1-14.
9. Fischer, S., et al. "Trends in genetic variance components during 30 years of hybrid maize breeding at the University of Hohenheim." Plant breeding 127.5 (2008): 446-451.
10. Dudley, J. W., and R. J. Lambert. 2004. 100 generations of selection for oil and protein in corn. Plant breeding reviews 24.1:79-110.
11. Falconer, D.S., and T.F.C. Mackay. 2005. Introduction to quantitative genetics. 4th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

12. Garrick, D. J. "An animal breeding approach to the estimation of genetic and environmental trends from field populations." Journal of animal science 88.suppl_13 (2010): E3-E10.
13. Gogel, Beverley J., Brian R. Cullis, and Arunus P. Verbyla. "REML estimation of multiplicative effects in multienvironment variety trails." Biometrics (1995): 744-749.
14. Hallauer, Arnel R., and Larry L. Darrah. "Compendium of recurrent selection methods and their application." Critical Reviews in Plant Sciences 3.1 (1985): 1-33.
15. Hill, William G. "Estimation of effective population size from data on linkage disequilibrium." Genetics Research 38.3 (1981): 209-216.
16. Knight, Robert L. "Dictionary of genetics, including terms used in cytology, animal breeding and evolution." Dictionary of genetics, including terms used in cytology, animal breeding and evolution. (1948).
17. Laidig, Friedrich, et al. "Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends." Theoretical and Applied Genetics 127.12 (2014): 2599-2617.
18. Legarra, Andres, I. Aguilar, and I. Misztal. "A relationship matrix including full pedigree and genomic information." Journal of dairy science 92.9 (2009): 4656-4663.
19. Lourenço, V. M., et al. "A robust DF-REML framework for variance components estimation in genetic studies." Bioinformatics 33.22 (2017): 3584-3594.
20. Mackay, I., et al. "Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time." Theoretical and Applied Genetics 122.1 (2011): 225-238.
21. Moose, Stephen P., John W. Dudley, and Torbert R. Rocheford. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in plant science 9.7: 358-364.
22. Mrode R., A. 2014 Linear models for the prediction of animal breeding values. 3rd ed. CABI Press, Oxfordshire OX10 8DE UK.
23. Nei, Masatoshi. Molecular population genetics and evolution. North-Holland Publishing Company., 1975.

24. Oakey, Helena, et al. "Joint modeling of additive and non-additive genetic line effects in single field trials." Theoretical and Applied Genetics 113.5 (2006): 809-819.
25. Piepho, Hans-Peter, et al. "Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials." Theoretical and applied genetics 127.5 (2014): 1009-1018.
26. Rutkoski, Jessica. "A practical guide to genetic gain." Advances in agronomy 157 (2019a): 217-249.
27. Rutkoski, J. E. "Estimation of realized rates of genetic gain and indicators for breeding program assessment." Crop Science 59.3 (2019b): 981-993.
28. Searle, Shayle R., and Marvin HJ Gruber. Linear models. John Wiley & Sons, 2016.
29. Schmidt, P., et al. "Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials." Crop Science 59.2 (2019): 525-536.
30. Schmidt, Paul, et al. "Heritability in plant breeding on a genotype-difference basis." Genetics 212.4 (2019): 991-1008.
31. VanRaden, P. M. "Genomic measures of relationship and inbreeding." INTERBULL bulletin 37 (2007): 33-33.
32. Walsh, Bruce. "Population-and quantitative-genetic models of selection limits." Plant breeding reviews 24.1 (2004): 177-226.
33. Walsh, B., and M. Lynch. 2018. Evolution and selection of quantitative traits. 1st ed. Oxford Univ. Press, Oxford, UK. doi:10.1093/ oso/9780198830870.001.0001
34. Wang, J., E. Santiago, and Armando Caballero. "Prediction and estimation of effective population size." Heredity 117.4 (2016): 193-206.
35. Weyhrich, Roger A., Kendall R. Lamkey, and Arnel R. Hallauer. "Effective population size and response to S1-progeny selection in the BS11 maize population." Crop science 38.5 (1998): 1149-1158.
36. Wright, Sewall. "Evolution in Mendelian populations." Genetics 16.2 (1931): 97.
37. Wright, Sewall. "Inbreeding and homozygosis." Proceedings of the National Academy of Sciences of the United States of America 19.4 (1933): 411.

Annex 1

Realized genetic gain methods

Although the expected (predicted) response to selection can be useful, it relies on many assumptions, and as such can only be considered valid for a single generation of response from an unselected base population (Walsh, 2004). The most accurate way to estimate the rate of genetic gain or response to selection is to use phenotypic data, whether from specific experiments to measure genetic gain or from the data generated by the program (this will influence the type of analysis as well). The two-step approach to estimate the rate of genetic gains involves modeling the phenotypic data for a trait or index of genetic merit as a function of time (years; better performance due to non-genetic improvement causes), covariates related to experimental design factors (nuisance parameters), the genotypes, and the interaction of genotypes with other factors such as years and locations, followed by a second model where the adjusted means are fitted as a function of the year of origin of the material (sometimes years of release in the case of varieties) (Mrode, 2014; Mackay 2011; Rutkoski, 2019). The regression coefficient for the time covariate in this second model provides the rate of genetic gain per year avoiding the second model (Piepho et al, 2014)

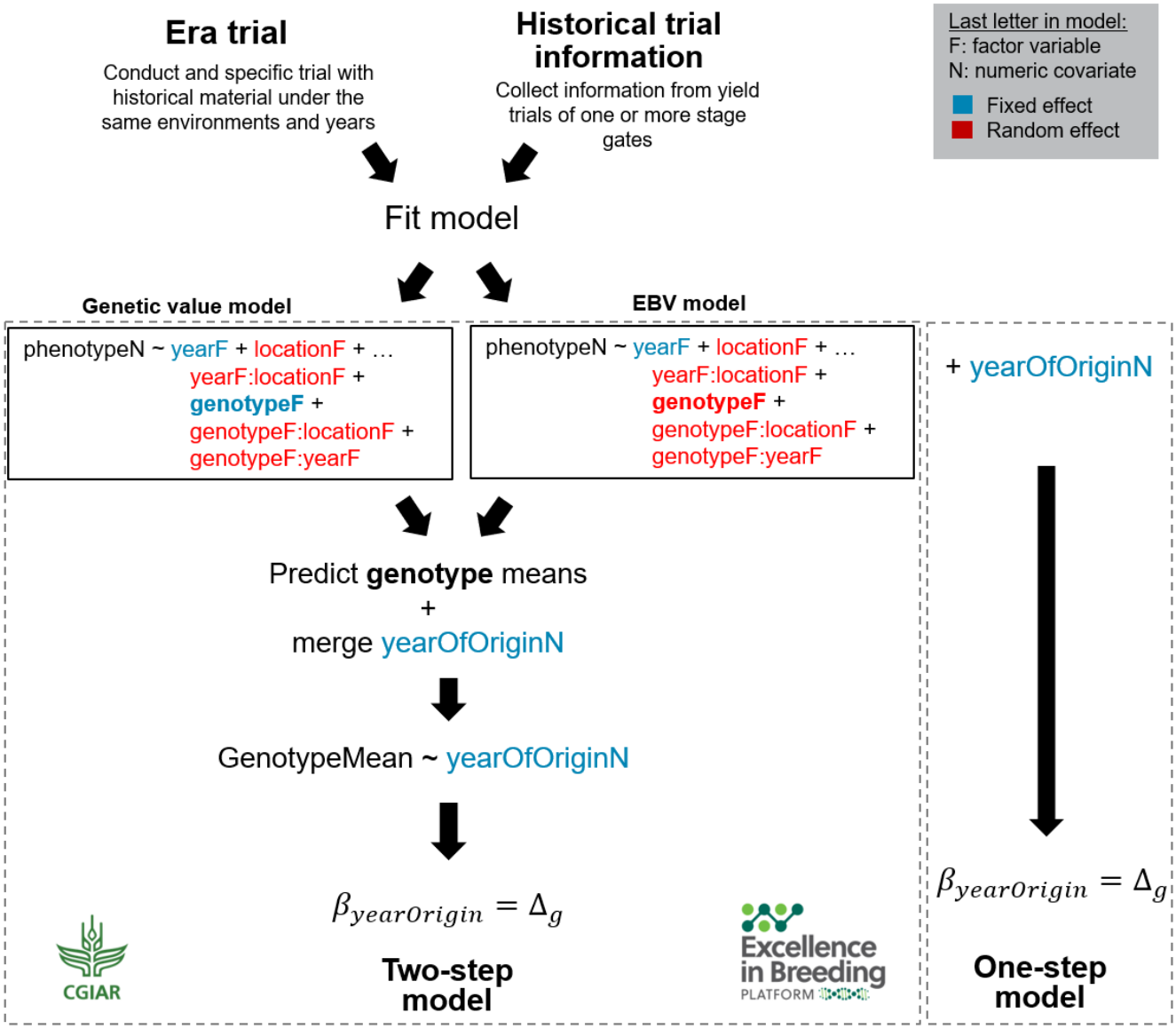

Figure A1 Graphical representation of the computation of the rate of genetic gain using linear models under different experimental situations. The diagram emphasizes the use of era trials and historical data to calculate the rate of genetic gain using a one-step or a two-step modeling approach.

Some of these methods have been compared by Rutkoski (2019b) with the goal of clarifying which methods are the best fit for different breeding program scenarios. Using simulations, Rutkoski compared different types of populations and methods to compare the true and estimated genetic gain, finding that era trials provide a good estimation of genetic gains; other options considered such as the use of control populations are unrealistic scenarios in practice. An important lesson from the simulations executed is that checks play an important role in raising the accuracy of this estimated parameter by connecting data properly.

The use of historical information to estimate genetic gain

These methods have been proposed for use in the scenario that the breeding program has historical information to estimate genetic gain. Historical germplasm samples could include any of the following:

- On-farm trial information
- Variety trial information
- Advanced yield trial information
- Preliminary yield trial information

The use of different sources of information changes the level of coverage of the TPE to properly estimate genetic values for the entries and connectivity between years (**Figure A2**).

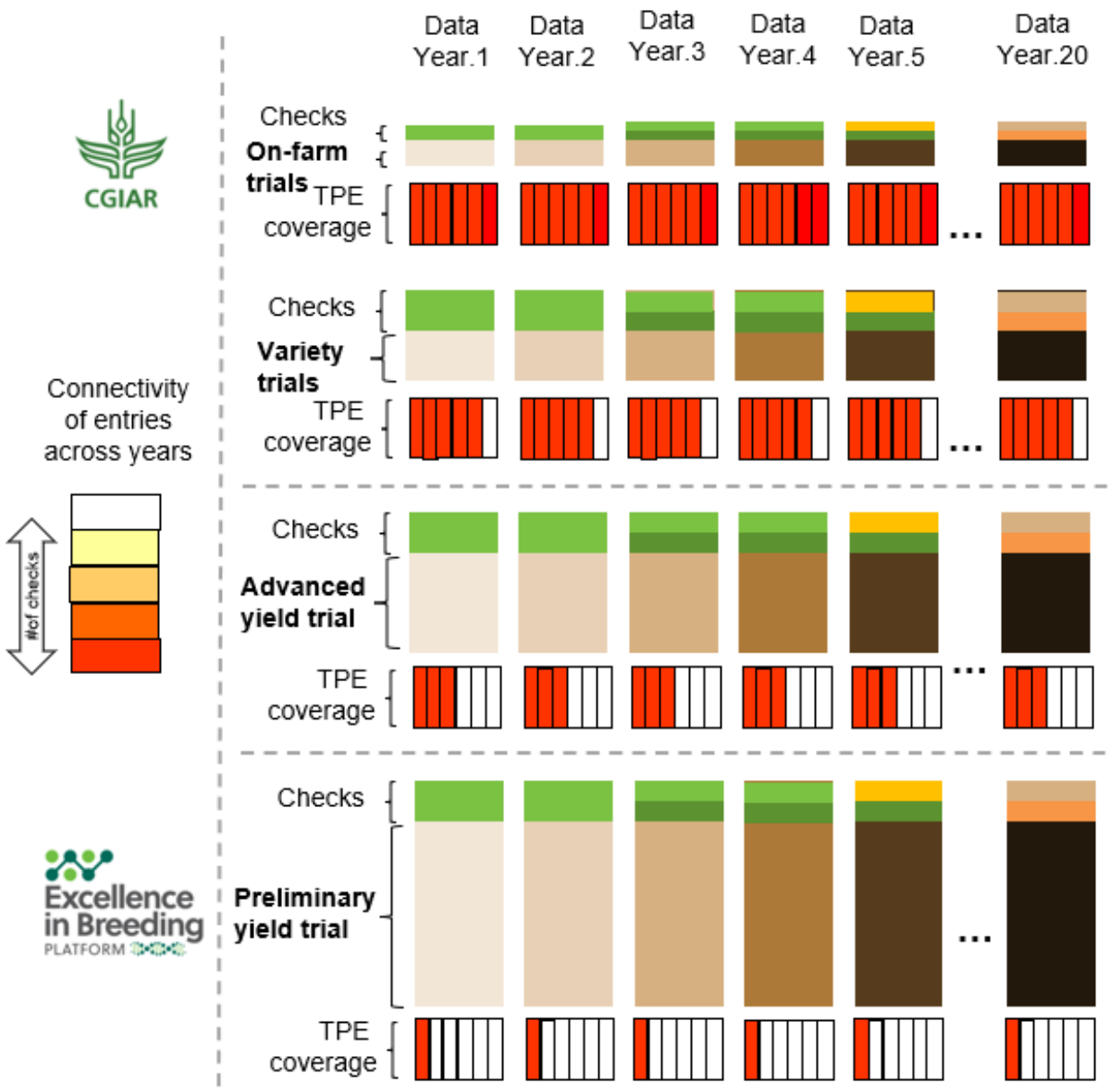

Figure A2. Graphical representation of the use of historical information to estimate genetic gain using different data sources of germplasm samples. The lack of overlapping of the yearly data is typical when using historical datasets (vertical red rectangles changing as #checks increases). On the other hand, the number of entries and their TPE coverage (horizontal red rectangles) varies depending on the stage used (germplasm sample), being the on-farm and variety trials the ones with bigger TPE coverage.

Here we will present an example using variety trial information, in which the material is tested across a large number of locations and years (despite having different checks across years). In this scenario, TPE is assumed to be effectively sampled, thereby decreasing the importance of genotype by year and genotype by location interaction. This makes the adjusted means for genotypes across years and locations accurate and the estimates of genetic gain reliable. The idea behind this model is to fit the main genotype and year effects as fixed, and interactions as random (Mackay et al., 2011; Laidig et al., 2014; Piepho et al., 2014). The model used takes the following form:

$$y = X\beta + Z_d u_d + Z_g u_g + \varepsilon$$

Where the vector β corresponds to the fixed effects of the factor variables for "year" and "genotype", and X is the incidence matrix connecting observations with the vector of fixed effects. The vector u_d refers to the vector of random effects for non-genetic effects such as "year by location", the vector u_g refers to the vector of random effects for "genotype by location" and "genotype by year" and the matrices Z_d and Z_g connect observations with the vectors of random effects u_d and u_g. A graphical representation of the estimation of genetic gain using variety trial information is shown in **Figure A2**.

As an example, the 5-year cycle breeding program from **Figure A3** was simulated for 20 years, assuming that the program has 4 checks per year that change every 3 years. **Step 1** was to gather the phenotypic data that would be used for the calculation of genetic gain. In this simulation, elite materials to be released as varieties (top 5 materials released every year) were picked. The information for the locations, years, genotypes and year of origin for the material (year when the material was created) required is as follows:

genotype	female parent	male parent	generation	location	phenotype	year	yearOrigin
G232411	G181850	G182485	10	e20	10.79194	2005	2001
G232404	G181850	G182485	10	e20	11.77105	2005	2001
G231572	G172422	G182483	10	e20	9.980104	2005	2002
G232396	G181850	G181765	10	e20	7.884396	2005	2001
...	...	...	...	...	...	...	...

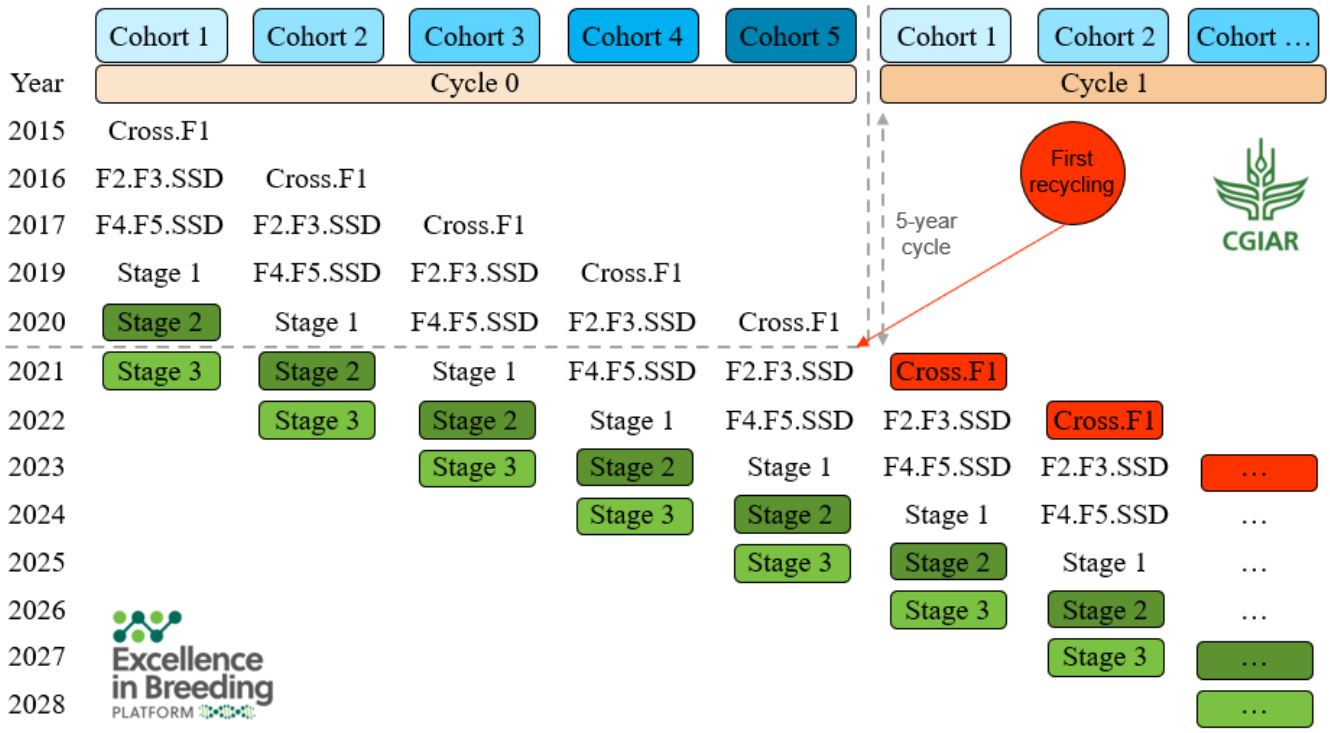

Figure A3. Schematic representation of a breeding program with a 5-year cycle comprising idealized 5 cohorts of crossing. For each cohort, year 1 is used to make the crosses and grow the F1 generation, year 2 is used to grow the F2 and F3's through an SSD process, year 3 the F4 and F5 generations are grown through SSD, year 4 the first stage of testing occurs and the best 10% individuals are grown, year 5 the second stage of testing occurs and the best 10% is selected to become parents of the next generation and be tested in a third stage of testing in year 6 to derive products. Every year the process is repeated and as soon as the pipeline starts to produce new parents these become the new breeding population.

Step 2 was to fit the linear mixed model mentioned above using available software (i.e. ASReml-R, SAS, breedR, sommer). For example, in ASReml-R nomenclature the mixed model fitted is as follows:

```
> myModel <- asreml(fixed= phenotypeN ~ yearF + genotypeF,
                    random= ~ locationF + yearF:locationF +
                                        yearF:genotypeF + locationF:genotypeF,
                    residual=~units,
                    data=myData)
```

The fixed effects for years remove the improvement due to non-genetic reasons, and the genotype effect refers to the improvement due to genetics. The random effects for location, interaction year by location, interaction year by genotype and location by genotype provide the rest of the adjustment, but are considered small given the sampling of the TPE. After this we predicted the genotype means across environments and years:

```
> predictions <- predict(myModel, classify = "genotypeF")
```

After merging the adjusted means with the year of origin of the material, we fitted a second model to calculate the rate of genetic gain:

```
> myModel2 <- lm(AdjustedMeans~yearOrigin, data=predictions)
> summary(myModel2)$coefficients
              Estimate  Std. Error   t value      Pr(>|t|)
(Intercept) 11.6257072 0.098068350 118.54698 1.223554e-107
yearOrigin   0.1377858 0.008186588  16.83068  1.142425e-30
```

As it can be seen the regression coefficient (slope=0.137…) for year of origin provides the increase in the original units of the trait per year, which is fair to call rate of genetic gain (~0.13 units per year; the simulated true rate of genetic gain was 0.12). An R script with a dataset to recreate these results is available in the EiB Toolbox entry for this manual. Here,

for simplicity variables for the experimental design like rows, columns, blocks, etc. were ignored, but they should be considered when fitting the first model.

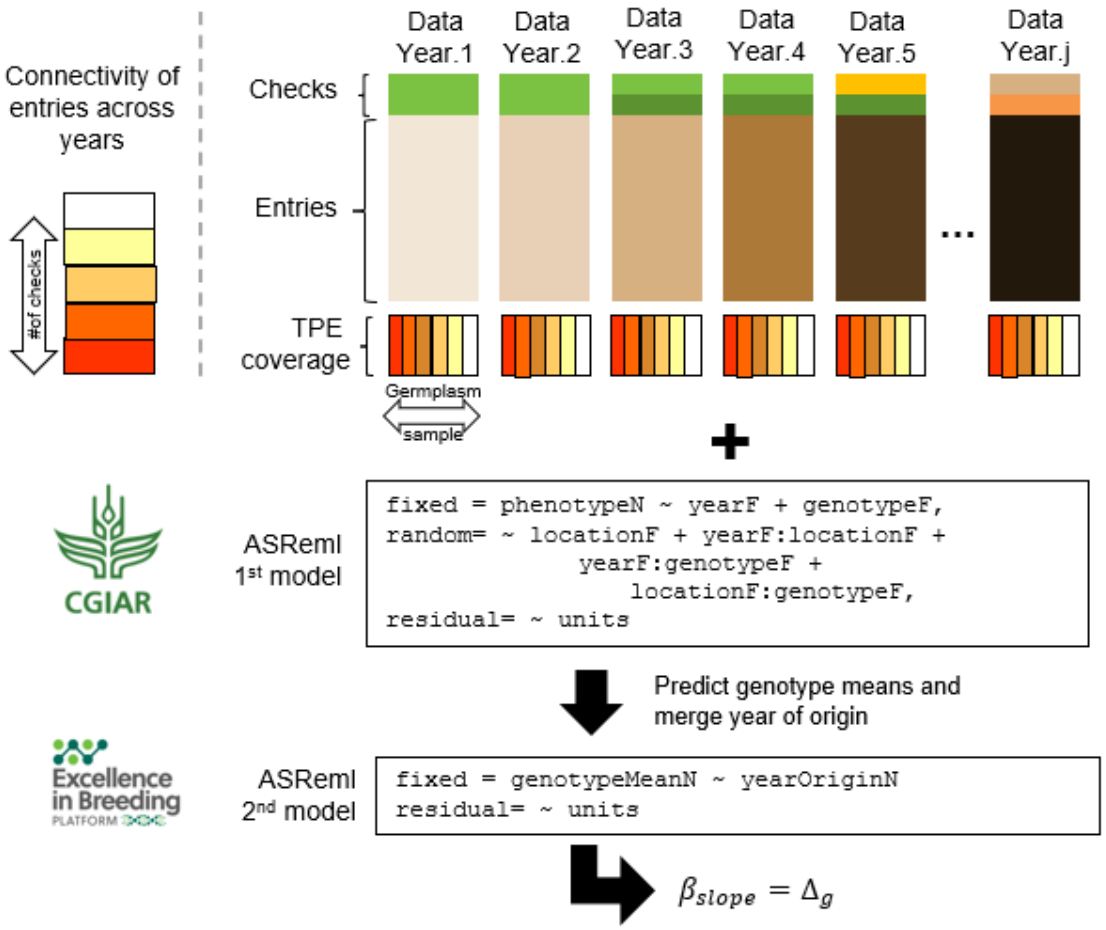

Figure A4. Workflow to analyze historical trial information method using a stage-bound germplasm sample. The lack of overlapping of the yearly data (vertical red rectangles) from the released varieties (same color represents contemporaneous entries) is tackled by using the right number of checks across years or by an effective sampling of the TPE (horizontal red rectangles) making the adjusted means for genotypes across years and locations accurate. Then, a statistical model is fitted to remove all the nuisance of improvement due to non-genetic causes (i.e. years). Finally, adjusted means are merged with and regressed as a function of the year of origin of the genotypes. The regression coefficient from this second model (βslope) is the rate of genetic gain per year.

Challenges and issues: The first issue with the opportunistic approach is that it relies on elite varieties for release being tested in many environments in order to sample the TPE effectively, decreasing the importance of genotype by year and genotype by location interaction, making the adjusted means for genotypes across years and locations accurate and the estimates of genetic gain reliable. If the variety trials do not sample the TPE properly then this method loses effectiveness. The second consideration is that varieties are usually not the best representation of the population mean across cycles but rather of the tails (transgressive events) of certain traits which could lead to some inconsistency of the genetic gain metric. FInally, the use of checks to connect the data is essential and has an important effect in the estimates, independently of the source of historical information.

Using estimated breeding value (EBV) as an alternative to increase connectivity in historical data

This modeling method is an extension of using historical information to estimate genetic gain. The purpose is to tackle low connectivity across years and low sampling of the TPE by connecting the data through the relationships that exist in the breeding material (Garrick, 2010), mainly through the incorporation of the genetic relationship matrix, whether based on pedigree or based on genetic marker data (to calculate a genomic relationship matrix). The idea behind this model is to fit the main genotype, genotype by year and genotype by location effects as random (requirement to use a relationship matrix). The model used takes the following form:

$$y = X\beta + Z_d u_d + Z_g u_g + \varepsilon$$

Where the vector β corresponds to the fixed effects of the factor covariate "year" and the factor and X is the incidence matrix connecting observation with the vector of fixed effects. The vector u_d refers to the vector of random effects for non-genetic effects such as "location", "year by location", the vector u_g refers to the vector of random effects for "genotype", "genotype by location" and "genotype by year" and the matrices Z_d and Z_g connect

observations with the vectors of random effects u_d and u_g. A graphical representation of the EBV method is shown in **Figure A5**.

The main difference with the previous example is the computation of an additive relationship matrix using the information on genotype, female, male names and selfing generation (see Mrode, 2015) for more details on how to calculate it. Many functions and software are already available, and the R script provided shows how to do it. The data required to calculate the relationship matrix based on pedigree could appear as follows, where the columns refer to the identifiers of genotypes, their parents and selfing generation:

genotype	female parent	male parent	generation
G232411	G181850	G182485	10
G232404	G181850	G182485	10
G231572	G172422	G182483	10
G232396	G181850	G181765	10
...	...	...	...

This information is used to create an expected identity by descent matrix among genotypes, known as a relationship matrix (Walsh and Lynch, 2018).

Alternately, if marker data was available, the marker matrix could be placed in a numeric format and subjected to a method to compute the additive genomic relationship matrix, such as Van Raden (2007) or a combination of pedigree and genomic information (i.e. see Legarra et al., 2009). A relationship matrix for genotypes was obtained in this step (**Figure A5**).

The other feature of the estimated breeding value method is the use of software to add relationship matrices for random effects (i.e. ASReml-R, SAS, breedR, sommer) to fit the linear mixed model mentioned above.

For example, in ASReml-R nomenclature the mixed model fitted is as follows:

```
> myModel <- asreml(fixed= phenotypeN ~ yearN + yearF,
                    random= ~ locationF + yearF:locationF +
                                vm(genotypeF, source=A) +
                                  diag(yearF):vm(genotypeF, source=A),
                    residual=~units,
                    data=myData)
```

The important component here is the fit of genotype and interaction of genotypes with years and locations as random to connect the data through the pedigree (in ASReml-R i.e. this is done using the *vm* function to add relationship matrices). Genotype means across environments and years are predicted as before, and the adjusted means are merged with the material year of origin. We then fit a second model to calculate the rate of genetic gain. An R script with a dataset to recreate these results is available in the EiB Toolbox entry for this manual. Here, for simplicity, the variables for the experimental design such as rows, columns, blocks, etc. were ignored, but this should be considered when fitting the first model.

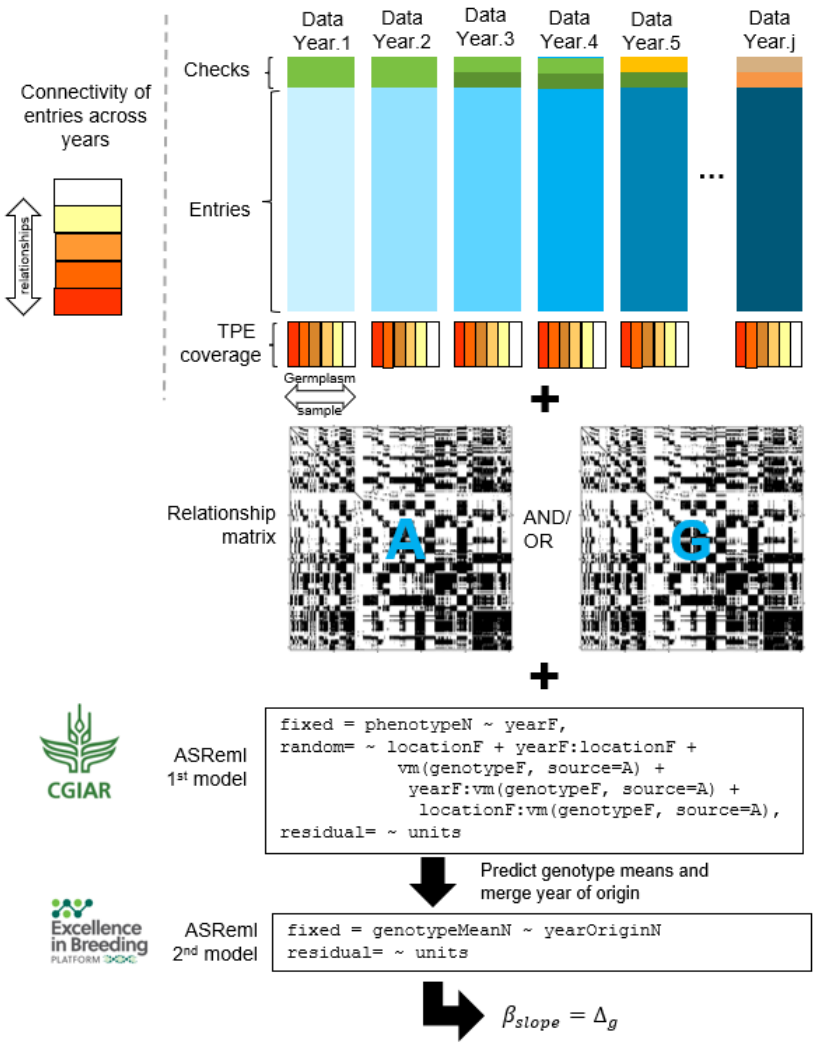

Figure A5. Workflow to analyze historical information where checks have not been sufficiently maintained to provide genetic gain estimates, requiring the application of the "EBV" method to increase connectivity. The lack of overlapping year data (vertical red rectangles) from the yield trials entries and checks (same color represents contemporaneous entries) is tackled by the use of a relationship matrix for genotypes to leverage from the population structure created by breeding programs. Then, a statistical model is fitted to remove all the nuisance of improvement due to non-genetic causes (i.e. years). Finally, adjusted means are merged and regressed as a function of the year of origin of the genotypes. The regression coefficient from this second model (β_1) is the rate of genetic gain per year.

Challenges and issues: The first challenge with the opportunistic approach and model are is the availability of pedigree information to fit models, as pedigree information is not consistently tracked by all breeding programs. In practice, the pedigree information requirement only necessitates that programs record parental information in two additional columns alongside the selfing generation. The second challenge is the issue of shrinkage when genotype effects are considered random. Fitting genotypes and interactions as random are requirements to use the pedigree information (to connect the data), but the amount of shrinkage will be dictated by the heritability of the trials. If the h^2 is so low that it will create a large shrinkage that will result in underestimates of the rate of genetic gain. This is confirmed in simulations that show that shrinkage could lead to the rate of genetic gain being underestimated by a factor of 5 to 6.

Conducting "era trials" as an alternative to increase connectivity in historical data

This method to estimate the rate of genetic gain aims to remove all the nuisance as a result of from under-sampling the TPE, year effects, and low connectivity of entries across years. The idea is very simple and consists in sampling material from all years of the breeding program from a given stage or germplasm sample (i.e. preliminary yield trial) and running a specific trial with all entries together in the same locations and years to estimate this metric.

The model is exactly the same as the one used for the "historical trial" method:

$$y = X\beta + Z_d u_d + Z_g u_g + \varepsilon$$

Where the vector β corresponds to the fixed effects of the factor variables for "year" and "genotype", and X is the incidence matrix connecting observations with the vector of fixed effects. The vector u_d refers to the vector of random effects for non-genetic effects such as "year by location", the vector u_g refers to the vector of random effects for "genotype by location" and "genotype by year" and the matrices Z_d and Z_g connect observations with the vectors of random effects u_d and u_g.

A graphical representation of the estimation of genetic gain using era trial information is shown in **Figure A6**.

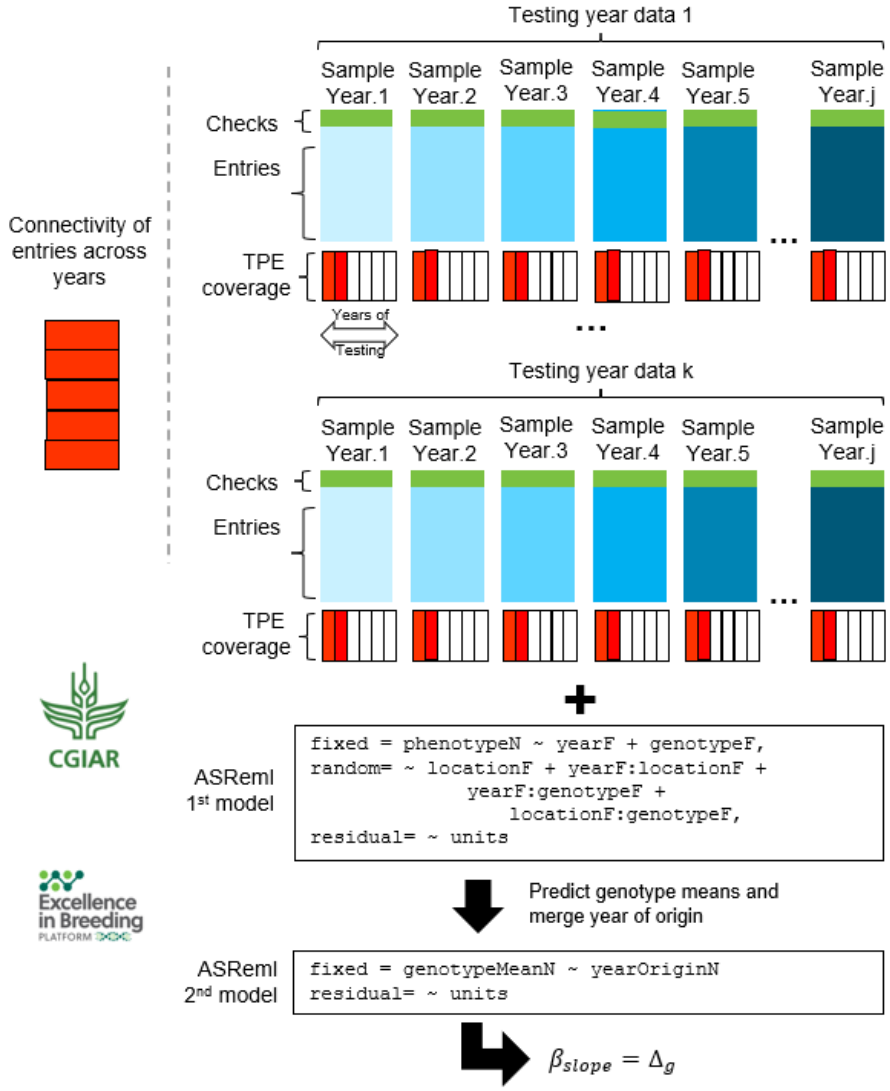

Figure A6. Workflow to analyze era trial information. The design of a trial with a sample of historical material allows for a full overlap of entries and minimize the connectivity issue and maximizing the accuracy of the estimate of the rate of genetic gain. On the other hand, the TPE is under-sampled leading to inaccuracy of the population mean which is of little relevance in these studies. A statistical model is fitted to remove all the nuisance of improvement due to non-genetic causes (i.e. years, locations, etc.). Finally, adjusted means are merged with and regressed as a function of the year of origin of the genotypes. The regression coefficient from this second model (β_1) is the rate of genetic gain per year.

Considerations for hybrid crops

Hybrid crops where the final product is not improved population (pool) but rather the hybrid between pools present a special case. In order to adjust to this scenario, instead of looking at the *per se* performance of the within-pool material, the program should monitor the increase in performance of the actual hybrids, or otherwise the testers used. Though it has been suggested to consider the general combining ability (GCA), this is not as important as considering the hybrids.