Continuing genetic improvement and biases in genetic gain estimates revealed in historical UK variety trials data

Joanna Raymond [a,*], Ian Mackay [b], Steven Penfield [c], Andrew Lovett [a], Haidee Philpott [d], Stephen Dorling [a]

[a] *School of Environmental Sciences, University of East Anglia, Norwich, UK*
[b] *Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK*
[c] *John Innes Centre, Norwich Research Park, Norwich, UK*
[d] *National Institute for Agricultural Botany, Lawrence Weaver Road, Cambridge, UK*

ARTICLE INFO

ABSTRACT

Context: The current pace of yield increase for major crops is not fast enough to meet future demand. Crop breeding programmes are under increasing pressure to improve existing crops further. Quantifying the contribution of these programmes to observed yield increases is important for evaluating their success and identifying if crop improvement goals are likely to be met.
Objective: In this paper we explore methods to study the genetic gain of two cereal species, wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.). Specifically, the objective of this research was to identify sources of bias in genetic gain estimates of UK variety trials data.
Methods: Genetic gain was estimated for fungicide-treated and untreated UK winter wheat, winter barley and spring barley for 1982–2018 using UK National List and Recommended List variety trials data. Subsets of the winter wheat variety trials dataset were used to replicate shorter breeding cycles to quantify the impact of the number and choice of long-term check varieties on estimating genetic gain.
Results: While genetic and non-genetic contributions to changes in UK cereal performance are in line with previous estimates, we were able to identify previously undetected changes and biases in estimates of variety performance. Specifically, we observed an increasing yield difference between fungicide treated and untreated variety trials as varieties age, driven by both a breakdown in disease resistance and a previously unobserved long-term increase in yield as varieties age in treated trials. This shows that yields of long-term check varieties cannot be assumed to be stable over time. We found that genetic gain estimates were highly sensitive to the long-term check varieties chosen, whilst the inclusion of multiple checks decreased the standard error of the estimate.
Conclusion: The estimation of genetic gain is highly susceptible to bias. We provide recommendations on how to reduce the risk of bias for estimating genetic gain.
Implications: Accounting for sources of bias in genetic gain calculations is important in any programme of selection to prevent inaccurate quantification of yield progress.

1. Introduction

A major agricultural challenge of the 21st century is to overcome yield stagnation observed in several staple crops globally, to help meet the growing global food demand (Hafner, 2003; Cassman et al., 2011; Grassini et al., 2013). This must be achieved in the context of a rapidly changing climate and increasingly extreme weather events (Parolini, 2022). Significant crop yield fluctuations have been seen in the past

decade in the UK (DEFRA, 2021) which mask recent yield trends and the contribution of plant breeding to any yield increases.

Crop improvement through plant breeding has been widely shown to have contributed to continued increases in yield potential, despite observed national yield stagnation (Peltonen-Sainio et al., 2009; Brisson et al., 2010; Noleppa and Cartsburg, 2021). Genetic gain estimates quantify the increase in performance of crop traits due to selection (Jayaraman, 2000; Xu et al., 2017; Sinha et al., 2021) and provide a

valuable measure of success of a breeding programme (Covarrubias-Pazaran, 2020; Covarrubias-Pazaran et al., 2022).

In the UK, the National List/Recommended List (NL/RL) variety trials test new crop varieties in different growing environments across the country each year. Testing the relative genetic potential of the new varieties ensures the release of only the best proportion for commercial use (Laidig et al., 2008). Varieties are initially tested in single and then multilocation trials by breeders, and these stages can typically last two to three years. If a variety is successful in these trials, it is then entered into the NL trials. These multi-environment trials allow breeders to test their varieties across current climates, soil types and locations within the UK's growing area. After at least two years in the NL trials, the Recommended List committee review variety performance and if successful, a variety will then move into the RL trials until outclassed, which is typically six years, but can be over 20 years (Austin, 1999; Mackay et al., 2011; Berry et al., 2015).

Historical genetic gain analysis has highlighted sources of bias within genetic gain estimates. Reanalyses of NL/RL cereal data showed that from 1982 to 2007, 88% of the improvement in winter wheat (*Triticum aestivum* L.) and winter barley (*Hordeum vulgare* L.) yield was attributable to genetic improvement (Mackay et al., 2011), indicating crop breeding in the UK has been fundamental to increasing the maximum attainable yields. However, genetic gain estimates for untreated variety trials were shown to be biased by the influence of loss of disease resistance. A frequent problem in calculating genetic gain is the confounding genetic and year effects due to a lack of genetic connectivity when breeding materials are tested for just one or two years (Rutkoski, 2019a, 2019b). The effectiveness and accuracy of genetic gain estimated from breeding programmes is not well known.

Therefore, the aims of this paper are to explore the methods of studying the genetic gain of wheat and barley, to model changes in their disease resistance, and to identify sources of bias in these methods to provide recommendations to improve the accuracy of current estimates. Subsets of the UK NL/RL variety trials data are used to mimic breeding programmes of different sizes and with varying number of long-term check (control) varieties.

2. Materials and methods

2.1. Variety trials dataset for wheat and barley

Winter wheat, winter barley and spring barley yield data were extracted from the UK National List (NL) and Recommended List (RL) field variety trials dataset for the period 1982–2018 (1983–2018 for spring barley). The Agriculture and Horticulture Development Board (AHDB) Recommended Lists is managed by a project consortium of AHDB, the British Society of Plant Breeders (BSPB), Maltsters' Association of Great Britain (MAGB) and the United Kingdom Flour Millers (UKFM). Full data for 2002 onwards is available at ahdb.org.uk/rl. Trials were located across England, Wales, Scotland and Northern Ireland, and were focused in relevant growing areas. Since 1982, cereal trials have been split into fungicide untreated and treated trials.

Prior to statistical analysis, we applied several pre-processing steps and quality control of the trials data. This step was required to amalgamate data from different databases (Fig. S1). Variety-year combinations with only one or two sites were also removed to avoid unrepresentative results due to insufficient replication. This included all

varieties introduced in 2011, which were only present for one or two years and/or had just one or two sites per year. Not all years and sites are connected by common cultivars, limiting connectivity upon which to evaluate non-genetic factors. Furthermore, varieties in trial for only one or two years provide little information for trend analyses, therefore these varieties were also removed from the dataset (Mackay et al., 2011; Piepho et al., 2014; Laidig et al., 2021). The summary of the data structure, after quality control, can be seen in Table 1. The distribution of the 2007–2018 trials relative to the respective crop growing area (EDINA, 2022) is graphically shown in Fig. 1. 2007–2018 is the period when all three crops had data on site location.

The number of varieties and trials, and therefore total observations, fluctuate year on year (Table S1). Variety numbers depend on submission by breeders into NL trials and selection by the RL committees to progress through the system. Some trial sites are reviewed on an annual basis, whilst the core trial programme is reviewed every five years. This can cause large fluctuations when trials are moved to accurately reflect the UK crop area. Some trials are also abandoned at an early stage. Given the large inter-annual variation in trial sites and varieties, not all varieties are grown on all sites and the trials data were unbalanced.

2.2. Modelling phenotype trends in UK variety trials data

In order to dissect the multiple factors contributing to yield, we analysed yield trends using linear mixed effects modelling, using a two-part method first described by (Breseghello, Morais and Rangel, 1998) and used widely since (Lange and Federizzi, 2009; Silva Junior et al., 2020; Ayenew, 2021). To analyse changes in variety performance over time and to estimate adjusted means across locations and years for each cultivar, the following linear mixed model was fitted separately to the treated and untreated trials data (Table 1):

$$y_{ijk} = \mu + r_j + v_i + vr_{ij} + s_{jk} + e_{ijk} \tag{1}$$

where y_{ijk} is the yield of variety i in year j at site k, μ is the overall trial series mean, v_i is the effect of variety i, r_j is the effect of year j, vr_{ij} is the effect of the interaction between variety i and year j, s_{jk} is the effect of site k in year j and e_{ijk} is the residual term (Mackay et al., 2011).

Year r_j was fitted as a factor and fixed effect due to the anticipated large non-linear effects of year and to provide consistency with Mackay et al. (2011). Likewise, variety v_i was included as a fixed effect as the NL/RL variety trials data is historical and individual varietal performance was of interest. The interaction terms varieties x years vr_{ij} and sites within years sr_{jk} were fitted as random effects as the data were incomplete.

Estimated variety effects and year effects were calculated using the best linear unbiased estimators (BLUEs). The variety BLUEs were then be regressed on year of first use for each variety to estimate the mean genetic gain, whilst the year BLUEs were regressed on calendar year. This allowed the genetic effects due to breeding efforts to be disentangled from non-genetic effects such as agronomic, climate, and policy changes. The standard error of the regression model was also extracted to quantify the uncertainty in the genetic gain estimate.

2.3. Modelling the changes in disease resistance

Changes in disease resistance of a variety can be observed by

Table 1
Structure of winter wheat (WW), winter barley (WB) and spring barley (SB) variety trials data after quality control and restricting for varieties present for a minimum of three years. Treated-untreated (T-U) pairs is the number of times a variety received both fungicide treatment regimes at the same site, in the same year.

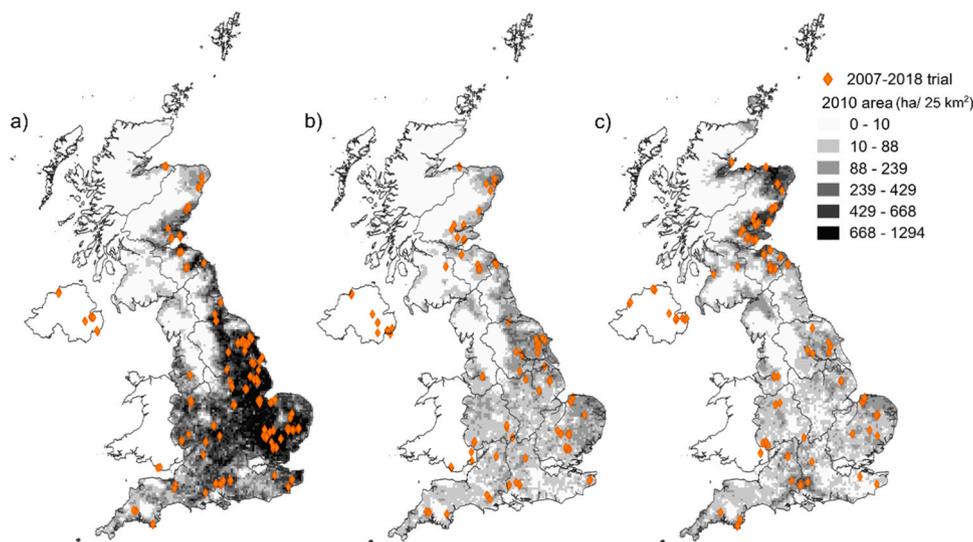| Crop	Number of variety-site-year combinations	Number of varieties	Start year	End year	Varieties per year	Treated trials (T)	Untreated trials (U)	T-U pairs
WW	64,719	274	1982	2018	42	42,472	22,247	17,952
WB	38,786	197	1982	2018	29	21,610	17,176	11,565
SB	34,872	184	1983	2018	32	18,613	16,259	10,468

Fig. 1. Trial locations (orange diamonds) for 2007–2018 for (a) winter wheat, (b) winter barley and (c) spring barley relative to their respective growing areas (ha/ 25 km^2) in the 2010 5 km Agricultural Census (EDINA, 2022). Data on winter and spring barley growing areas in Wales and all three crops in Northern Ireland were not available.

comparing treated and untreated trials grown at the same location. In this analysis, 17,952, 11,565 and 10,468 treated-untreated pairs for winter wheat, winter barley and spring barley, respectively, were available for varieties with a minimum of three years in the trials dataset (Table 1). In calculating the yield difference in treated and untreated pairs, and accounting for disease x environment interactions, it is possible to quantify loss of yield due to disease. The fitted model used is:

$$y_{ijk}^d = \mu^d + a_j + v_i^d + va_{ij} + sr_{jk}^d + e_{ijk}^d \tag{2}$$

y_{ijk}^d is the yield difference in treated and untreated trials for variety i at site k after j years, μ^d is the mean difference, v_i^d is the effect of variety i on yield difference, a_j is the effect of variety age j on yield difference, va_{ij} is the effect of the interaction between variety i and variety age j, sr_{jk}^d is the effect of site k in year j on yield difference and e_{ijk}^d is the residual term.

The variety effect v_i^d and variety age effects a_j have been fitted as fixed effects, whilst the variety x variety age va_{ij} and sites s_{jk}^d terms are fitted as random effects. Estimated variety age effects were then regressed on yield difference to see the extent of disease resistance breakdown as varieties age. Here variety effect v_i^d represents an average resistance of a variety over its lifetime and variety age a_j represents the average reduction in resistance of all varieties over their lifetimes. The variety x variety age interaction term va_{ij} is important for understanding whether different varieties lose resistance at different ages.

Variety age was calculated by subtracting the harvest year from the year of entry into the trials system. Year of entry was calculated as the first year a variety was present in the trials data provided. For varieties present pre-1982, year of entry was found from the trials datasets used in Mackay et al. (2011).

2.4. Estimating uncertainty in genetic gain estimates

To explore the various influences in trials datasets on genetic gain estimates and their uncertainty, we broke the NL/RL winter wheat treated variety trials data down into case study periods. To select these case study periods, varieties present for at least 10 consecutive years were first identified. The 10-year criteria was selected as most varieties are in trial for fewer years than this, but there were still enough meeting the criteria to test the inclusion of multiple long-term "check" varieties in genetic gain estimates within different case study periods. A

connectivity table was then created, to identify case study periods when at least four checks overlapped. These case study periods are summarised in Table 2.

Breeding programmes frequently trial varieties for just one or two years, therefore rather than excluding varieties present for less than three years in the trials data as in 2.2 and 2.3, here only data for the first, second and third year of each variety was included. Given the nature of breeding programmes, there was an additional requirement that the two or three years were consecutive. Missing data for several varieties in 2007 prevented models converging, therefore this requirement was relaxed for 2008. The number of varieties and number of observations for each case study is shown in Table 2.

For each case study (Table 2), the genetic gain was calculated as in 2.2, by first estimating the BLUEs for year and variety using Eq. (1), and then regressing adjusted variety means on year of entry but here excluding the check(s). Checks were removed for the regression estimate to avoid bias from including their older years of entry. Initially regressions were calculated for all checks and varieties. For a case study with five checks, each check was then dropped individually to calculate the genetic gain with a combination of four checks. Subsequently each

Table 2
Genetic gain winter wheat case studies for various periods and check varieties from within the UK NL/RL variety trials dataset. Checks refer to varieties present for at least 10 consecutive years in the NL/RL trials dataset. All checks and varieties here have received a full fungicide treatment. Only the first three years of trials for varieties present for less than 10 years are included in each case study dataset. Non-check observations indicates the number of data points within the period for the 0-check model. Adding in checks increases the number of data points.

Case study	Period	Checks	Non-check observations
a	1982–1991	*Norman, Galahad, Fenman, Longbow*	2890
b	1991–1999	*Riband, Hereward, Soissons, Hussar*	4063
c	1996–2006	*Claire, Riband, Hereward, Consort, Soissons*	5982
d	2000–2009	*Claire, Consort, Solstice, Robigus*	5107
e	2005–2015	*Claire, Cordiale, Solstice, JB Diego, Alchemy*	6249
f	2008–2017	*JB Diego, Viscount, Cordiale, Grafton*	7263

combination of two checks were dropped, then three checks, four checks and finally all checks. Hence the effect of the number of checks and the checks chosen on genetic gain and its uncertainty can be investigated.

In addition to the case study periods, the genetic gain for the whole period 1982–2018 was recalculated using the dataset with the only the first 1–3 years for each variety.

3. Results

3.1. Genetic gain in treated and untreated variety trials

Treated variety trial yields increased at a faster rate than the untreated variety trial yields for winter wheat, winter barley and spring barley (Fig. 2). Across all treatments and crops, the positive contribution of variety effects to linear increases in yield (i.e. genetic gain) was significant (Table 3, Fig. 3), confirming that breeding has continued to contribute to the observed yield increases (Fig. 2). Untreated variety trial genetic gain estimates were consistently higher than the treated genetic gain estimates (Table 3), particularly for winter wheat for which genetic gain from 1982 to 2018 was estimated to be 0.063 (SE = 0.002, $p < 0.001$) and 0.109 (SE = 0.003, $p < 0.001$) t ha^{-1} yr^{-1}, for treated and untreated variety trials, respectively.

Table 3

Rate of change in variety and year effects over time for winter wheat, winter barley and spring barley. The standard error (SE) in the linear trend over varieties estimates are also given. All trends are calculated using the method described in Section 2.2 and are significant ($p < 0.05$) unless denoted with *ns*. Yield values are expressed in t ha^{-1}.

| Crop	Period analysed	Fungicide treatment	Linear trend over varieties (SE)	Linear trend over years
Winter wheat	1982–2018	Treated	0.063 (0.002)	-0.0030 *ns*
Winter wheat	1982–2018	Untreated	0.109 (0.003)	-0.076
Winter barley	1982–2018	Treated	0.054 (0.002)	0.035
Winter barley	1982–2018	Untreated	0.057 (0.003)	0.0034 *ns*
Spring barley	1983–2018	Treated	0.058 (0.001)	-0.019
Spring barley	1983–2018	Untreated	0.068 (0.002)	-0.027
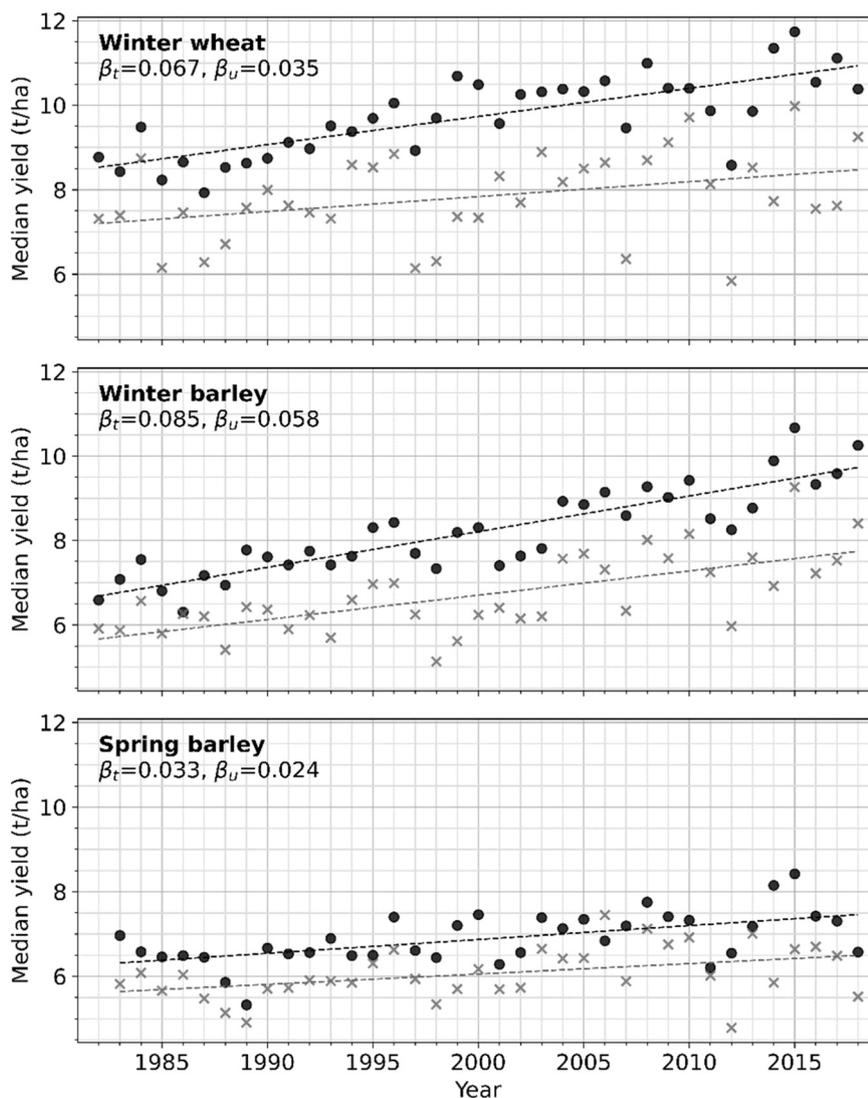
Fig. 2. Median winter wheat, winter barley and spring barley for treated (●) and untreated (x) variety trial yields for 1982–2018 harvest years. The linear increase in median yield was significant ($p < 0.05$) for all three crops and treatments.
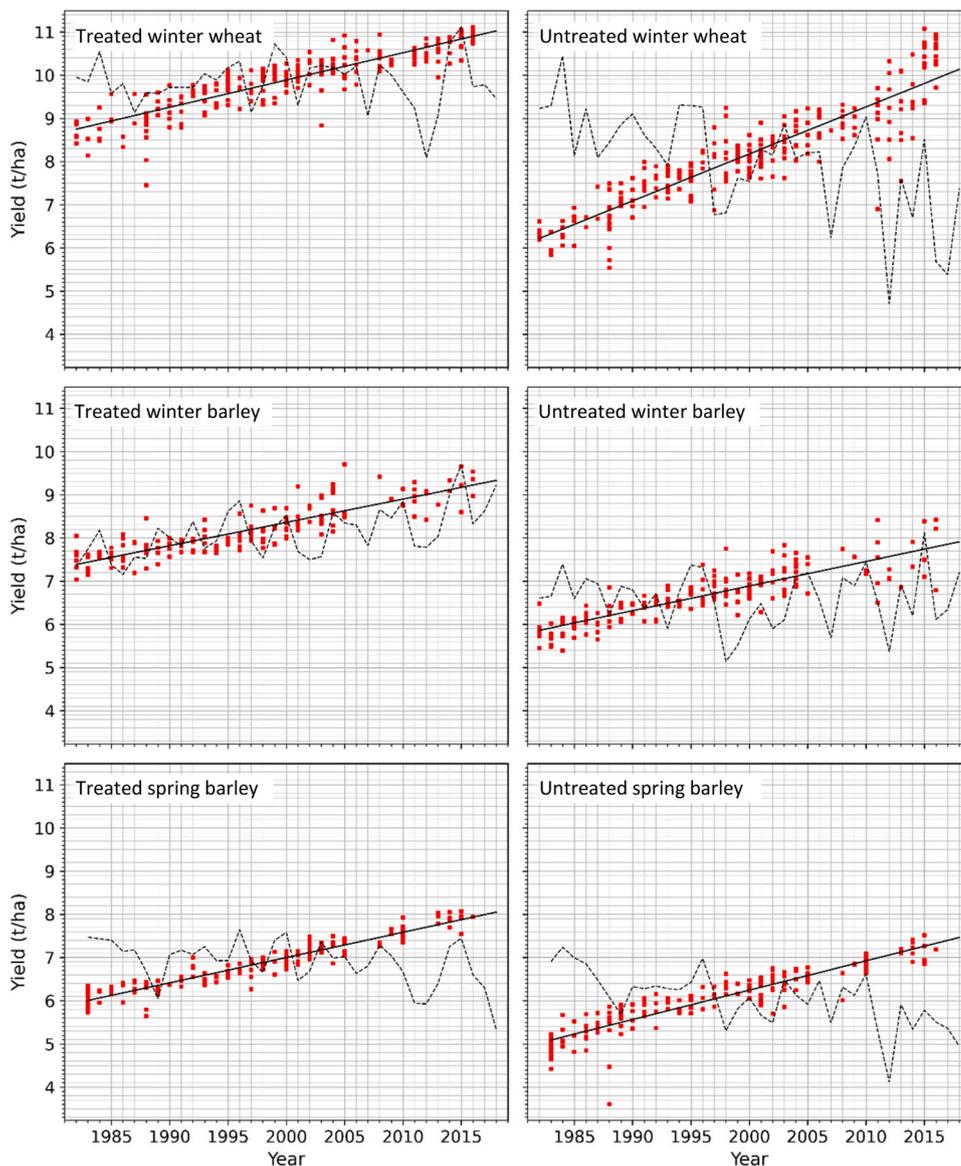
Fig. 3. Trends in variety and year effect for fungicide treated and untreated winter wheat, winter barley and spring barley trial yields from 1982 to 2018 (1983–2018 for spring barley), calculated using the two-step method described in 2.2. Variety effects (red squares) were plotted against the first year they entered the trials. Year effects (dotted line) were plotted against calendar years. In the spring barley and winter wheat data treated and untreated plots there were no variety effect data points for 2011. In the original dataset, varieties introduced this year were only present for just one or two years and/or had just one or two sites per year and were therefore removed prior to analysis.

3.2. Changes in long-term variety yields

Treated and untreated yield difference had a positive linear relationship with variety age of 0.064 t ha^{-1} yr^{-1} (SE = 0.006, p < 0.001), 0.032 t ha^{-1} yr^{-1} (SE = 0.002, p < 0.001) and 0.015 t ha^{-1} yr^{-1} (SE = 0.004, p < 0.001) (Fig. S2) for winter wheat, winter barley and spring barley, respectively. To investigate the drivers of this trend further, the effects of variety age on treated and untreated trial yields were modelled separately (Fig. 4).

Winter wheat and winter barley yields increased significantly as varieties aged in the treated variety trials, by 0.030 and 0.021 t ha^{-1} yr^{-1}, respectively (Fig. 4). All three crops had significant yield decreases as varieties aged in the untreated variety trials. The standard error in the treated and untreated variety trial yield estimates increased after 10 years, driven by the reduction in the number of varieties contributing to the estimates, and yields began deviating away from the linear trend.

As a result, the analysis was repeated, restricting varieties to their first 10 years of data (Fig. 5). The variety age-yield linear regression models had a much better fit (higher R^2) to this data. The rate of yield loss in untreated variety trials was also greater, and was highest in winter wheat, which showed yield losses of 0.9 t ha^{-1} for a variety present for 10 years in the trials system (Fig. 5). Spring barley untreated

varieties experienced yield losses of 0.6 t ha^{-1} and winter barley yield losses were only 0.3 t ha^{-1} over 10 years. After the first two years in trial, both treated and untreated variety trial yields decreased, coinciding with varieties moving from the NL to RL trials.

This decline is small but statistically significant: for each of the three crops, yield is higher in years one and two than in years three, four and five. Partitioning the data into a contingency table and carrying out a contingency chi-squared test gives a non-significant p-value (1 m permutations) of 0.1 (Table S2a). However, the pattern is identical across crop groups and combining the data (Table S2b) gives a p-value of 0.002. There are 10 possible partitions of the five years of sequential testing into groups of two and three. A Bonferroni adjusted p-value for these data, to protect against the risk of post-hoc testing and cherry picking the first two years to compare against the last three years is therefore 0.02, giving statistical support for a true decline in average performance from NL to RL.

An alternative approach to test the significance is a t test of the difference between years one and two, against years three, four and five. P-values from these tests are 0.032, 0.009 and 0.015 for winter wheat, winter barley and spring barley respectively. These values are the smallest (i.e. most significant) among all 10 possible comparisons of two against three years from the five years of NL and RL testing, giving
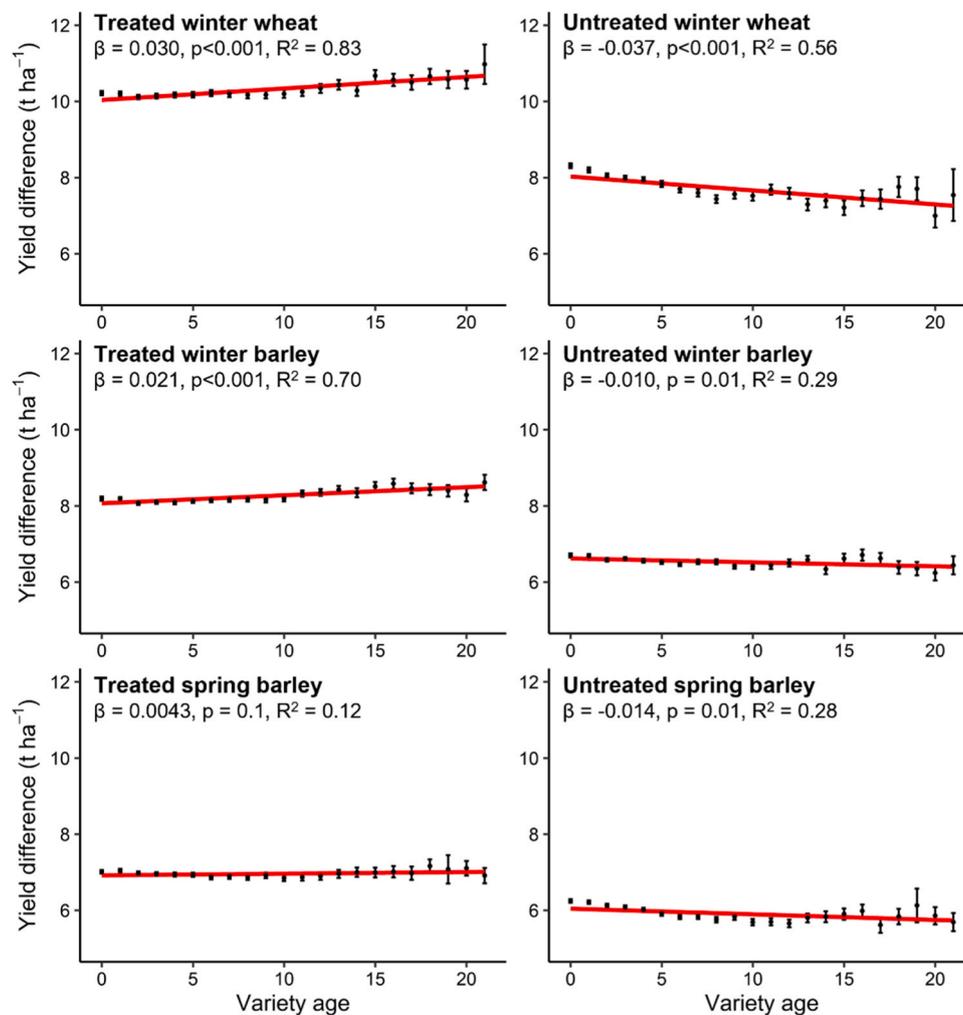
Fig. 4. Variety age against fungicide treated and untreated trial yields for winter wheat, winter barley and spring barley. Variety age indicates the number of years since the variety entered the trials system. The red line shows the linear relationship between the two variables. The regression coefficient (β), p-value and R^2 associated with each linear regression, calculated using methods described in Section 2.3, are given.

statistical support for a true decline in average performance from NL to RL.

3.3. Genetic gain estimates are susceptible to bias

Using subsets of the NL/RL dataset and varying the number of long-term check varieties allowed us to mimic different scenarios that can arise in breeding programmes. Increasing the number of checks was shown to decrease the genetic gain estimate (Fig. 6). This was particularly clear for the first four case study periods (a-d), in which genetic gain estimates were highest when there were no checks included and decreased by 30–40% upon the inclusion of one check, 10–20% when there were two checks, 5–10% when there were three checks. The estimates converged as the number of checks increased and the standard errors decreased. This effect on standard error was expected since the number of data points increased as more checks were added.

For the two most recent time periods (Fig. 6e and f), the association between the genetic gain estimates and number of checks was less clear. This is partly due to the size of genetic gain values, which were much lower for these two periods. For 2005–2015 the values decreased by 20% on average from zero checks to one check, and then by a few percent between subsequent increases in checks. Unfortunately, for 2008–2017 it was also not possible to calculate the genetic gain for zero checks as there were insufficient data.

The 1982–2018 treated genetic gain for winter wheat was recalculated using just the first three years of data for each variety with no checks. The estimated genetic gain was 0.158 t ha^{-1} yr^{-1} (SE = 0.003), which is 2.5 times larger than the original estimate of 0.063 t ha^{-1} yr^{-1} (SE = 0.002). This supports the increased genetic gain values associated with zero checks seen in the case studies (Fig. 6).

The choice of check also influenced the genetic gain estimates (Fig. 6). This is shown by the spread in genetic gain estimates for all case study periods with just one check. For example, the genetic gain estimate for 2005–2015 was 0.050 t ha^{-1} yr^{-1} with check variety *Alchemy*, compared to 0.029 t ha^{-1} yr^{-1} with check variety *Claire*. For 1982–1991, the values ranged from 0.21 t ha^{-1} yr^{-1} (*Galahad*) to nearly 50% larger at 0.29 t ha^{-1} yr^{-1} (*Fenman*).

4. Discussion

There has been much research into genetic gain in both plant and animal breeding, evaluating the success of breeding programmes (Ortiz et al., 2002; Cossani et al., 2022), looking at methods of enhancing genetic gain (Xu et al., 2017; Cobb et al., 2019) and its importance in improving global food security (Tadesse et al., 2019). Much fewer consider sources of bias in their calculation (Mackay et al., 2011; Rutkoski, 2019a, 2019b; Hartung et al., 2023). In this paper, we have focussed on the latter subject, utilising the UK NL/RL dataset to explore how change in disease resistance and use of long-term check varieties can influence estimates.
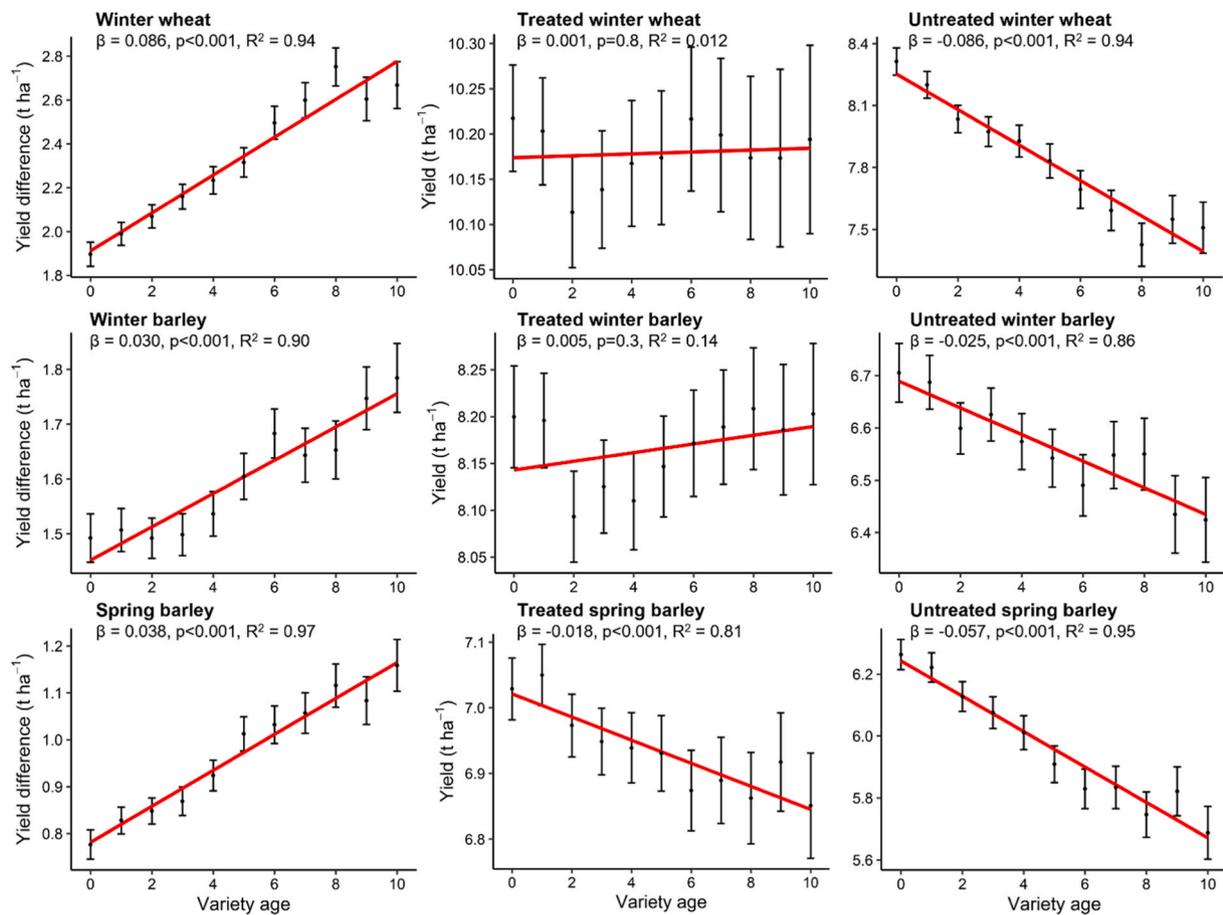
Fig. 5. The effect of variety age on yield difference between paired fungicide treated-untreated variety trials, on treated variety trial yields and on untreated variety trial yields for winter wheat, winter barley and spring barley. Here the analysis has been restricted to the first 10 years each variety is present in the trials system. Variety age indicates the number of years since the variety entered the trials system. The red line shows the linear relationship between the two variables. The regression coefficient (β), p-value and R^2 associated with each linear regression, calculated using methods described in Section 2.3, are given.

We have shown that plant breeding has continued to contribute to yield increases in winter wheat, winter barley and spring barley, shown by the positive genetic gain estimates for all NL/RL variety trials analysed for 1982–2018 (Fig. 3, Table 3). We found untreated genetic gain estimates to be higher than treated genetic gain estimates for all three crops, in particular winter wheat (Fig. 3, Table 3). Comparing these higher genetic gain values with the slower rate of yield increase found in the untreated variety trial yields (Fig. 2) suggests that these genetic gain estimates are overestimated. Indeed, (Mackay et al., 2011) found genetic gain to be overestimated for untreated variety trials. They attributed this to a reduction in disease resistance of varieties as they age, due to the increase in the yield difference observed between treated and untreated variety trials over time. We also observed this significant ageing trend on treated-untreated yield difference in the UK NL/RL variety trials for 1982–2018 (Fig. S2), which agreed with the previous finding by Mackay et al. (2011).

In France, genetic gain was estimated to be higher in untreated wheat trials, but was instead attributed to improvement in resistance to fungal disease (Brisson et al., 2010). This was also suggested as an explanation by (Shorinola et al., 2022) for the UK. However, by modelling the effect of variety age on treated and untreated variety trial yields separately, we have shown that untreated variety trial yields do significantly decrease as they age up to 10 years, supporting the theory on increased disease susceptibility (Fig. 4) (Laidig et al., 2021). The greatest yield loss in untreated variety trials was for winter wheat, in agreement with (Laidig et al., 2022), whilst winter barley was less affected by untreated yield loss as varieties aged. This highlights the

continued need for new improved varieties to combat yields losses as varieties age.

Unlike the decline over the first 10 years, the upward deviation in yield trends after 10 years affected both treated and untreated trials (Fig. 4). This previously undocumented trends suggests a cause unaffected by fungicide treatment. Long-term varieties in this dataset were introduced at different points in the period of interest, therefore it does not appear to be the effect of increasing yields at some point in the timeseries which is independent of genotype. Some varieties may have become more resistant to disease as different disease races come to dominate. For example, a variety is normally more susceptible to one or more races of yellow rust rather than all races of yellow rust. If, over time, the dominant race isn't the one it's susceptible to, its resistance could improve and this may explain the observed increase seen. It may also be possible that these longer standing varieties end up benefitting from the effect of being surrounded by newer resistant varieties, so they get less disease than they would if older, less resistant varieties were nearby. A more sophisticated analysis at the plot level, taking into account the effects of neighbouring plots, could be a way of testing this.

4.1. Genetic gain estimates can be biased by long-term check variety usage

Estimates of genetic gain for different case study periods within the 1982–2018 NL/RL trials data were dependent on the number of check varieties included in the dataset and the specific checks chosen. Specifically, increasing the number of long-running check varieties lowered the genetic gain estimates (Fig. 6). Absence of check varieties results in
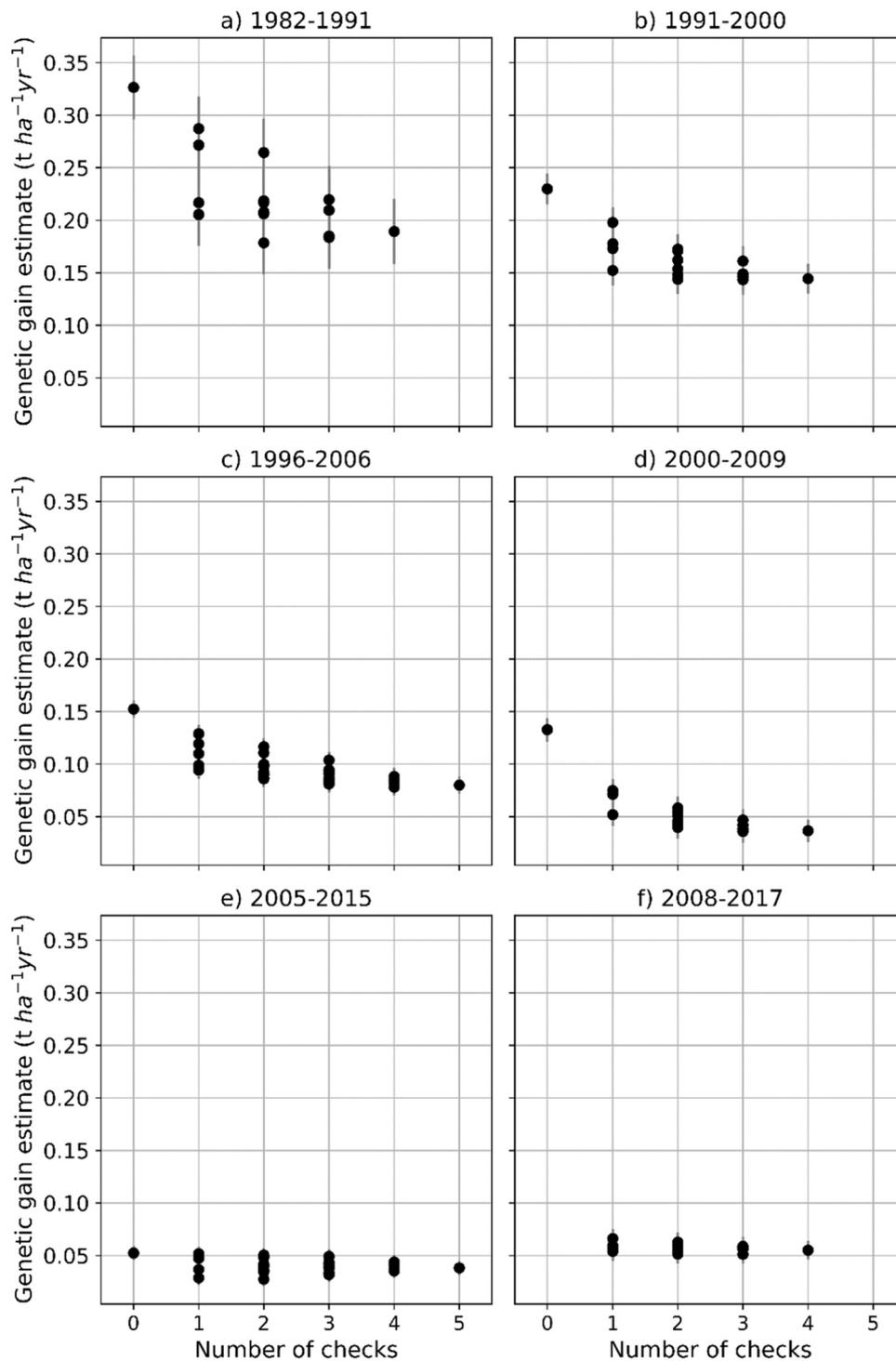
Fig. 6. Winter wheat genetic gain estimates for six case study periods and varying numbers of checks extracted from the 1982–2018 NL/RL fungicide treated variety trials dataset. Checks refer to varieties present in the trials system for a minimum of 10 consecutive years. Varieties with more than three years of data were restricted to their first three years in trial.

low connectivity which means the estimates of genetic gain can be confounded with the year effect, hence it is recommended that checks are used to improve estimates (Rutkoski, 2019a, 2019b; Covarrubias-Pazaran, 2020). Having multiple checks makes it easier to identify the effects of years (Fig. 6). However, it also means an increased proportion of the estimate of genetic gain comes from the difference in age and yield between the checks themselves. If these yields are increasing at a lower rate than the new varieties, this can drag estimates down.

Genetic gain estimates were also highly dependent on the checks chosen, particularly when only one check was included. We suggest this is because the checks were not stable and behaved differently within the case study periods: they had different mean yields and some showed slight increases over time whilst others did not. The effect of smaller trial datasets was also demonstrated by the larger standard errors in genetic gain estimates with fewer checks and overall data points (Fig. 6). Breeding programmes with more sites per variety and trials overall can reduce this standard error, as well as sampling error (Carena et al., 2010;

Rutkoski, 2019a, 2019b).

Additional model runs that included checks in the regression estimate showed that this lowers the genetic gain estimate further. The extent to which the estimate is lowered is dependent on the mean yield of that check. For example, a check with a higher adjusted mean yield (c2 in Fig. S3) will lower the genetic gain estimate (G2) compared to a check (c1) that behaves the same across the period but with a lower adjusted mean yield introduced in the same year, as newer varieties with higher adjusted mean yield will not show as large relative increases in comparison. When checks are included in the final regression estimate of genetic gain, it could therefore be possible to, knowingly or not, bias a genetic gain estimate upwards by using a consistently low yielding check in a breeding programme (Fig. S3). Evidently this method of calculating genetic gain needs refining to reduce the vulnerability of the estimate to the choice and number of checks. This could include excluding the check from the regression estimate and using checks with stable yields over time.

A large inflation in the genetic gain estimate was also found in the 1982–2018 treated winter wheat variety trial dataset, when varieties were restricted to their first three years in trial (0.158 t ha^{-1} yr^{-1} vs. 0.062 t ha^{-1} yr^{-1}). In this particular winter wheat dataset it was found that variety yields decreased in the first three years (Fig. 5), before showing overall long-term increases (Fig. 4). This fall in post-registration wheat performance has also been observed by wheat breeders (Joe He, *pers. comm.*). Therefore, it is possible that the treated genetic gain value here is biased in a similar way to that of the untreated variety trial yields, such that the year effects are underestimated and variety effects overestimated.

It is noteworthy that the size of the NL/RL dataset has enabled the detection of the small drop in yield from NL (years one and two) to the RL system (years three to five) and is found in all three crops. Contemporary experiments on a scale required to detect effects this small would be impractical and uneconomic. Historical data supports hypothesis generation and testing which would not otherwise be possible.

There are two possible and non-exclusive causes of the reduction in yield. Firstly, selection bias: as varieties are advanced from NL to RL trials, the highest performing varieties in the NL trials are both higher yielding genetically and also experience positive yield deviations due to the effect of a favourable growing environment, such that their average yields are overestimated. In the subsequent years of RL testing, selection bias is greatly reduced as the effect of the environment averages out. Secondly, it may be due to a difference in seed quality between National List and Recommended list trials. The former requires smaller quantities of seed provided directly from the breeder. Better seed lots can therefore be selected and seed processing can be to a higher standard. For Recommended List trials, seed quantities are increased and the opportunity to grade and select seed is reduced. Ultimately, seed is sampled from certified seed as sold to growers, which will meet the high statutory requirements for quality, but cannot be graded to extreme high standards possible for NL trials.

A significant long-term increase in yield was found in long-lasting winter wheat and winter barley varieties (Fig. 4). Untreated variety trial yields also stopped declining at 10 years. This means in the full dataset analysis (Fig. 4) the adjusted mean yields for long lasting varieties were higher than in the analysis restricted to the first three years during which yields generally decline, resulting in a lower genetic gain estimate, as explained in Fig. S3.

4.2. Recommendations

With these findings in mind, we make the following recommendations to reduce the risk of bias in future genetic gain calculations:

- At least two stable long-term check varieties must be included to increase connectivity between varieties

- Checks must be used to calculate best linear unbiased estimators (BLUEs) for variety and year effects but are then removed for the regression estimate
- The yield effect of factors such as variety age or seed source must be considered prior to estimating genetic gain, and if there is a distinct yield drop as seen in the NL/RL data (Fig. 5), a term should be included in Eq. (1) to account for this

If genetic gain is to be continued to be used as a high-level key performance indicator for public and philanthropically funded breeding programmes (Covarrubias-Pazaran, 2020; Williamson and Leonelli, 2022), significant research is required to achieve a better understanding on the causes of variation in the estimates seen here and minimise bias in future genetic gain estimates.

5. Conclusion

Breeding is still contributing to increases in yield in UK winter wheat, winter barley and spring barley. The increase in yield difference between fungicide treated and untreated variety trials as varieties age is driven by both a breakdown in disease resistance of untreated varieties and previously unobserved long-term yield increases of varieties in treated trials.

Use of NL/RL trials data enabled us to explore potential sources of uncertainty in genetic gain estimates. Varying the number of long-term check varieties in the data showed that inclusion of checks leads to a less biased estimate of year effects. However, the genetic gain estimate is highly sensitive to the check chosen and is influenced by the initial drop in yield associated with moving from NL to RL. This raises important questions about how best to calculate genetic gain.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The datasets used in this paper are available for academic research upon request to AHDB-BSPB, subject to AHDB-BSPB's Materials Transfer Agreement.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fcr.2023.109086.

References

Austin, R.B., 1999. Yield of wheat in the United Kingdom: recent advances and prospects (Available at:). Crop Sci. 39 (6), 1604–1610. https://doi.org/10.2135/cropsci1999.3961604x.

Ayenew, B.Z., 2021. Genetic gain in yield and yield attributing traits of rice under upland ecosystem of Fogera, Northwest Ethiopia (Available at:). Black Sea J. Agric.. https://doi.org/10.47115/bsagriculture.639422.

Berry, P.M., et al., 2015. Historical analysis of the effects of breeding on the height of winter wheat (Triticum aestivum) and consequences for lodging (Available at:). Euphytica 203 (2), 375–383. https://doi.org/10.1007/s10681-014-1286-y.

Breseghello, F., Morais, O.P. de, Rangel, P.H.N., 1998. A new method to estimate genetic gain in annual crops (Available at:). Genet. Mol. Biol. 21 (4), 551–555. https://doi.org/10.1590/S1415-47571998000400024.

Brisson, N., et al., 2010. Why are wheat yields stagnating in Europe ? A comprehensive data analysis for France (Available at:). Field Crops Res. 119, 201–212. https://doi.org/10.1016/j.fcr.2010.07.012.

Carena, M.J., Hallauer, A.R., Miranda Filho, J.B., 2010. Quantitative Genetics in Maize Breeding. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4419-0766-0.

Cassman, K.G., Grassini, P., van Wart, J., 2011. Crop Yield Potential, Yield Trends, and Global Food Security in a Changing Climate. Handbook Of Climate Change And Agroecosystems: Impacts, Adaptation, And Mitigation. Imperial College Press, London, pp. 37–51. https://doi.org/10.1142/9781848166561_0004.

Cobb, J.N., et al., 2019. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation (Available at:). Theor. Appl. Genet. 132 (3), 627–645. https://doi.org/10.1007/s00122-019-03317-0.

Cossani, C.M., Palta, J., Sadras, V.O., 2022. Genetic yield gain between 1942 and 2013 and associated changes in phenology, yield components and root traits of Australian barley (Available at:). Plant Soil 480 (1), 151–163. https://doi.org/10.1007/s11104-022-05570-7.

Covarrubias-Pazaran, G. (2020) Genetic gain as a high-level key performance indicator.

Covarrubias-Pazaran, G., et al., 2022. Breeding schemes: what are they, how to formalize them, and how to improve them? (Available at:). Front. Plant Sci. 12 (January), 1–15. https://doi.org/10.3389/fpls.2021.791859.

DEFRA (2021) 2020 Agriculture in the United Kingdom'.

EDINA (2022) Agricultural Census 2010, AgCensus. Available at: https://agcensus.edina.ac.uk/.

Grassini, P., Eskridge, K.M., Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends (Available at:). Nat. Commun. 4 (1), 11. https://doi.org/10.1038/ncomms3918.

Hafner, S., 2003. Trends in maize, rice, and wheat yields for 188 nations over the past 40 years: a prevalence of linear growth (Available at:). Agric., Ecosyst. Environ. 97 (1–3), 275–283. https://doi.org/10.1016/S0167-8809(03)00019-7.

Hartung, J., Laidig, F., Piepho, H.-P., 2023. Effects of systematic data reduction on trend estimation from German registration trials (Available at:). Theor. Appl. Genet. 136 (1), 1–12. https://doi.org/10.1007/s00122-023-04266-5.

Jayaraman, K., 2000. 6.1 Genetics and plant breeding. A Stat. Man. For. Res.

Laidig, F., et al., 2021. Long-term breeding progress of yield, yield-related, and disease resistance traits in five cereal crops of German variety trials (Available at:). Theor. Appl. Genet.. https://doi.org/10.1007/s00122-021-03929-5.

Laidig, F., et al., 2022. Yield reduction due to diseases and lodging and impact of input intensity on yield in variety trials in five cereal crops (Available at:). Euphytica 218 (10), 1–29. https://doi.org/10.1007/s10681-022-03094-w.

Laidig, F., Drobek, T., Meyer, U., 2008. Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials (Available at:). Plant Breed. 127 (6), 541–547. https://doi.org/10.1111/j.1439-0523.2008.01564.x.

Lange, C.E., Federizzi, L.C., 2009. Estimation of soybean genetic progress in the south of Brazil using multi-environmental yield trials (Available at:). Sci. Agric. 66 (3), 309–316. https://doi.org/10.1590/s0103-90162009000300005.

Mackay, I., et al., 2011. 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 (Available at:). Theor. Appl. Genet. 122 (1), 225–238. https://doi.org/10.1007/s00122-010-1438-y.

Noleppa, S., Cartsburg, M., 2021. The socio-economic and environmental values of plant breeding in the EU and selected EU member states. hffa Res.

Ortiz, R., et al., 2002. Genetic gains in Nordic spring barley breeding over sixty years (Available at:). Euphytica 126 (2), 283–289. https://doi.org/10.1023/A:1016302626527.

Parolini, G., 2022. Weather, climate, and agriculture: historical contributions and perspectives from agricultural meteorology (Available at:). WIREs Clim. Change 13 (3), 1–19. https://doi.org/10.1002/wcc.766.

Peltonen-Sainio, P., Jauhiainen, L., Hakala, K., 2009. Are there indications of climate change induced increases in variability of major field crops in the northernmost European conditions? Agric. Food Sci. 18, 206–222.

Piepho, H.P., et al., 2014. Dissecting genetic and non-genetic sources of long-term yield trend in german official variety trials (Available at:). Theor. Appl. Genet. 127 (5), 1009–1018. https://doi.org/10.1007/s00122-014-2275-1.

Rutkoski, J.E., 2019a. A practical guide to genetic gain. in Advances in Agronomy. Elsevier Inc, pp. 217–249. https://doi.org/10.1016/bs.agron.2019.05.001.

Rutkoski, J.E., 2019b. Estimation of realized rates of genetic gain and indicators for breeding program assessment (Available at:). Crop Sci. 59 (3), 981–993. https://doi.org/10.2135/cropsci2018.09.0537.

Shorinola, O., et al., 2022. Trend, population structure, and trait mapping from 15 years of national varietal trials of UK winter wheat (Available at:). G3 Genes|Genomes|Genet. 12 (2). https://doi.org/10.1093/g3journal/jkab415.

Silva Junior, A.C. da, et al., 2020. Genetic progress over twenty-three years of irrigated rice breeding in southeastern Brazil (Available at:). Acta Sci. Agron. 43 (1), e45767. https://doi.org/10.4025/actasciagron.v43i1.45767.

Sinha, P., et al., 2021. Genomics and breeding innovations for enhancing genetic gain for climate resilience and nutrition traits (Available at:). Theor. Appl. Genet. 134 (6), 1829–1843. https://doi.org/10.1007/s00122-021-03847-6.

Tadesse, W., et al., 2019. Genetic gains in wheat breeding and its role in feeding the world (Available at:). Crop Breed., Genet. Genom.. https://doi.org/10.20900/cbgg20190005.

Williamson, H.F., Leonelli, S., 2022. Accelerating agriculture: data-intensive plant breeding and the use of genetic gain as an indicator for agricultural research and development (Available at:). Stud. Hist. Philos. Sci. 95, 167–176. https://doi.org/10.1016/j.shpsa.2022.08.006.

Xu, Y., et al., 2017. Enhancing genetic gain in the era of molecular breeding (Available at:). J. Exp. Bot. 68 (11), 2641–2666. https://doi.org/10.1093/jxb/erx135.