

STØTTET AF

# Mælkeafgiftsfonden

## SingleStep report – 2021

Kevin Byskov

The aim of these analyzes are to compare results from the current two-step (TS) procedure with the new single-step (SS) procedure to validate the results of the SS procedure.

### Comparison of trend from TS vs SS for bulls

In November 2016 Nordic Cattle Genetic Evaluation introduced new post processing of genomic enhanced breeding values (GEBVs) to reduce inflation, which caused young genomic tested bull to have to high GEBVs compared to daughter proven bulls. To compare if this is an issue for young genomic tested bulls in the Single Step SS procedure, trends from the SS and TS procedures have been compared for bulls with official NTM in November 2016. This is done by comparing difference in official TS indexes for milk, fat and protein yield between November 2016 and August 2020 with difference in full and reduced (excluding 4 years of phenotypic data) run indexes for the same traits. Bulls have been divided into 3 different groups:

- Proven-proven bull, which are bulls that have at least 50 offspring with data in both full and reduced SS run.
- Genomic-Proven bull, which are bulls with 0 offspring with data in reduced SS run and at least 50 offspring with data in full SS run.
- Genomic-genomic bull, which are bulls with 0 offspring with data in both full and reduced SS run.

Only results for bull categories with at least 25 bulls per year class is shown and only results for bulls with both TS and SS GEBVs in both 2016 and 2020 are used.

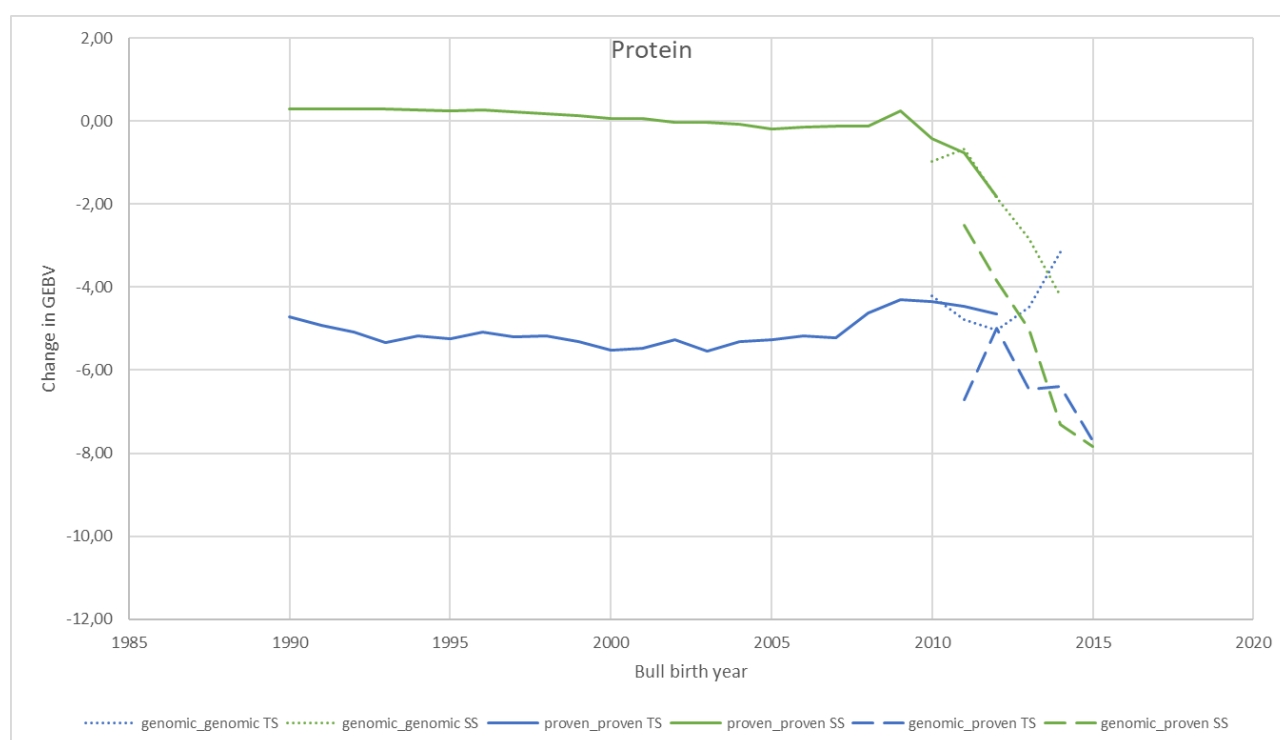


Figure 1 Difference in protein GEBV for protein yield for two-step procedure between November 2016 and August 2020 official run (blue lines) and single-step procedure between full and reduced run (green lines). Full lines are bulls with at least 50 offspring with phenotypic data in both full and reduced single-step run. Dashed lines are bulls with 0 and at least 50 offspring with phenotypic data in reduced and full single-step run, respectively. Dotted lines are bulls with at least 50 offspring with phenotypic data in both full and reduced single-step run.

The full blue line in Figure 1 shows that the trend for protein yield between November 2016 and August 2020 have been approx. 5 index units. The difference in the level between the full blue and the full green line only shows that the cow base for TS GEBVs published November 2016 and August 2020 are different, as a rolling cow base is used for official runs. For the green full line, which is the difference between full and reduced SS run, the change in GEBV is approx. zero until 2009, which reflects that the same cow base is used for both runs. The dotted blue line is at approx. same level as the full blue line, indication that bulls being solely genomic bulls in both November 2016 and August 2020 have a similar trend to bulls being daughter proven in both runs. The dashed blue line is at a bit lower level than the full line indicating, that when genomic bulls get daughter information, they tend to fall a bit more on average than expected even though a new post processing procedure was introduced in November 2016 to handle this problem. But Figure 1 also shows that the problem is much smaller for the TS procedure compared to the SS procedure. For bulls being solely genomic bulls and born in 2015 and having at least 50 daughters with phenotypic data in August 2020, the drop in protein yield index is almost 8 index units compared to bulls born 2009 or earlier, that have daughter information in both the full and reduced run.

In Figure 2 a similar figure as Figure 1 is shown, except that for the TS it is the difference between August 2016 and August 2020 i.e. before the introduction of the new post processing procedure. This shows that the problems with inflation in the TS procedure, which are taken care of with the new post processing, seems to be a problem also for the SS procedure. If it is not inflation, then is at least a problem with same magnitude. Only bulls with both TS and SS GEBVs in both 2016 and 2020 are in the figures, which explains small differences in the SS graphs especially for 2015 birthyear class, where some bulls were not old enough in August 2016 to have official GEBVs.

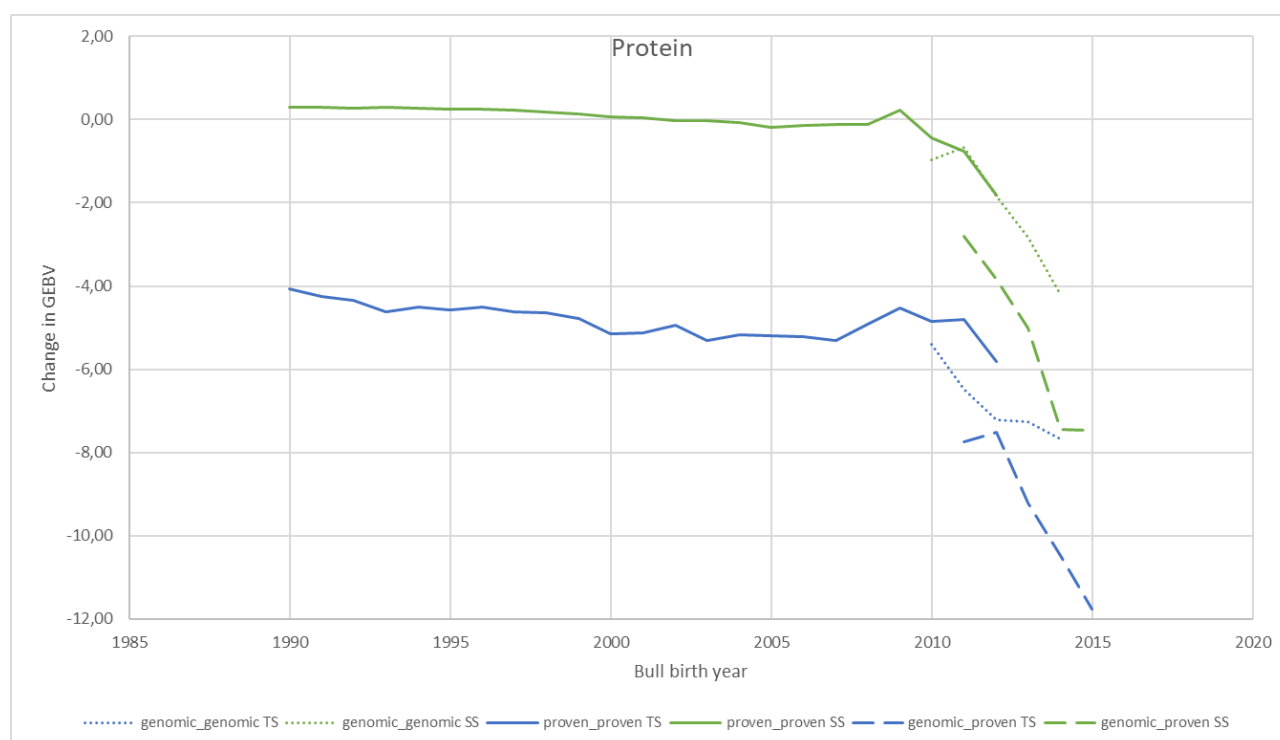


Figure 2 Difference in protein GEBV for protein yield for two-step procedure between August 2016 and August 2020 official run (blue lines) and single-step procedure between full and reduced run (green lines). Full lines are bulls with at least 50 offspring with phenotypic data in both full and reduced single-step run. Dashed lines are bulls with 0 and at least 50 offspring with phenotypic data in reduced and full single-step run, respectively. Dotted lines are bulls with at least 50 offspring with phenotypic data in both full and reduced single-step run.

In Figure 1 and Figure 2 are shown the results for protein yield. In Appendix 1 the results for milk and fat yield are shown.

### MS for offspring of Proven or Genomic bulls, respectively and MS for different groups of EBV levels

It has been shown in 2020 that the Mendelian Sampling term (MS) from the SS GEBVs was larger than zero for genotyped animals without own performance, even though the animals generally were unselected before genotyping. This is still a main issue related to the SS evaluation and is also thought to be at least partly the explanation for the higher GEBV trends found in the reduced datasets compared to the full data. To investigate this problem further, we have investigated if the problem with the higher MS is similar for offspring from genomic bulls and daughter proven bulls, and if EBV level affect the level of MS.

The original hypothesis was that MS bias was less for offspring from proven bulls compared to offspring from genomic bulls, as the distance to phenotypic data is smaller. Definition of groups were based on distance between offspring birthday and sire birthday. If less than 5 years, then the sire was considered a genomic bull. Otherwise, the sire was considered a daughter proven bull. Only female offspring without own phenotype has been used for the comparison of offspring with genomic or daughter proven sires. In the following figures it must be noted that the amount of data until 2018 is limited, as most females born before 2018 did have own production records in August 2020, where the data is from.

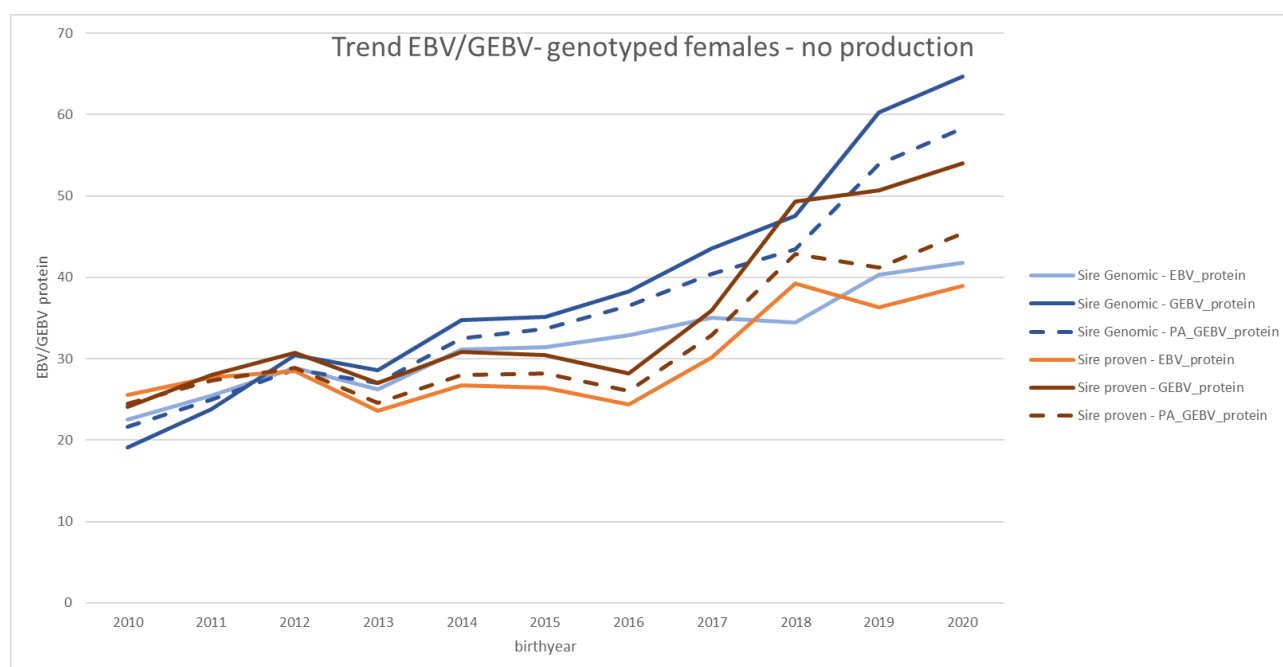


Figure 3 EBV, SS\_GEBV and SS\_GEBV parent average (PA\_GEBV) for protein yield in kg for genomic tested females with no own performance. Bluish lines are offspring from genomic bulls (<5 years), and reddish lines are for offspring from daughter proven bulls. Darker lines are GEBVs and lighter colored lines are EBVs. Full dark lines are SS GEBVs and dashed lines parent average from SS GEBVs.

In Figure 3 it is seen that offspring from genomic bulls tend to have a slightly higher genetic level for protein yield than offspring from daughter proven sires especially when looking at SS\_GEBV. When looking at EBVs the difference is smaller. This also means that difference between SS\_GEBVs and EBVs are larger for offspring for genomic bulls. This is expected, at the EBVs for genomic bulls does not consider the selection of these genomic bulls. But we should expect that these selected bulls on average have a positive MS, and therefore also that their EBVs are underestimated.

Looking at the MS in Figure 4, which is also the difference between the solid and dashed dark lines in Figure 3, offspring from daughter proven sires tend to have higher MS than offspring from genomic sires until the sires get progeny test. This can be recognized from the graphs in Figure 4, where the magnitude of MS is similar until 2017. For offspring from genomic sires born in 2019 and 2020 and partly also in 2018, the sire will in most cases not yet have a progeny test. It is difficult to explain why MS is higher for offspring born in

later birthyear classes from daughter proven sires, even though these offspring are “closer to” phenotypic data. But perhaps it can be caused that genomic sires themselves have slightly overestimated GEBVs which makes the MS look smaller.

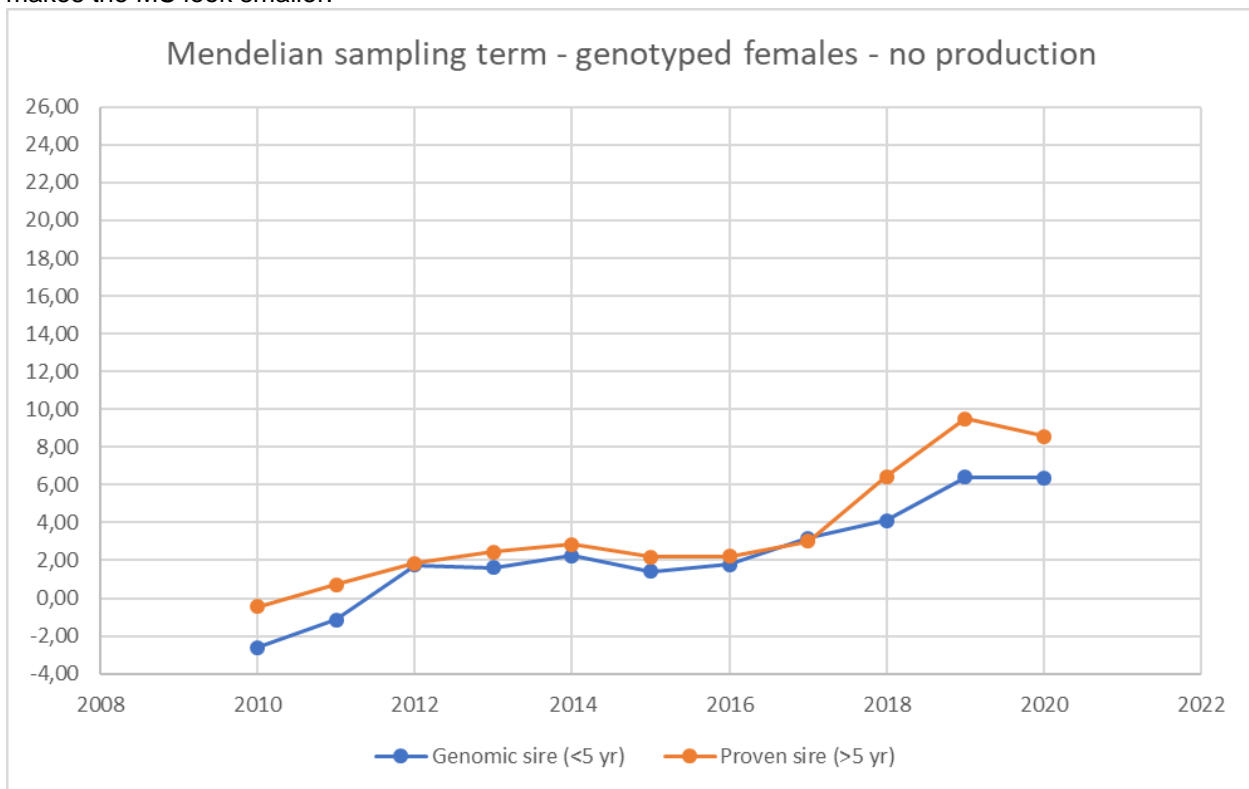


Figure 4 Mendelian Sampling term in kg protein for female offspring with no own performance data from either genomic or daughter proven sires.

Another thing we investigated was MS for different levels of EBVs. Based on females EBVs they were grouped into 3 groups, and MS was calculated as  $SS\_GEBV$  minus parent average  $SS\_GEBV$ . In Figure 5 it can be seen that the group with the lowest EBVs also have the lowest MS for birthyear classes before 2019. In these years own performance of the cows affects the EBVs, which thereby also contain information on MS. Therefore, the figures looks as expected, in years where most of the animals have own performance. Also, the middle group have a MS close to zero is as expected.

For the latest birthyear classes the MS is of similar magnitude for all 3 EBV groups, but it was expected that this MS should be close to zero. It can be concluded that the problems we see with positive MS is not related to genetic level for females without own phenotype.

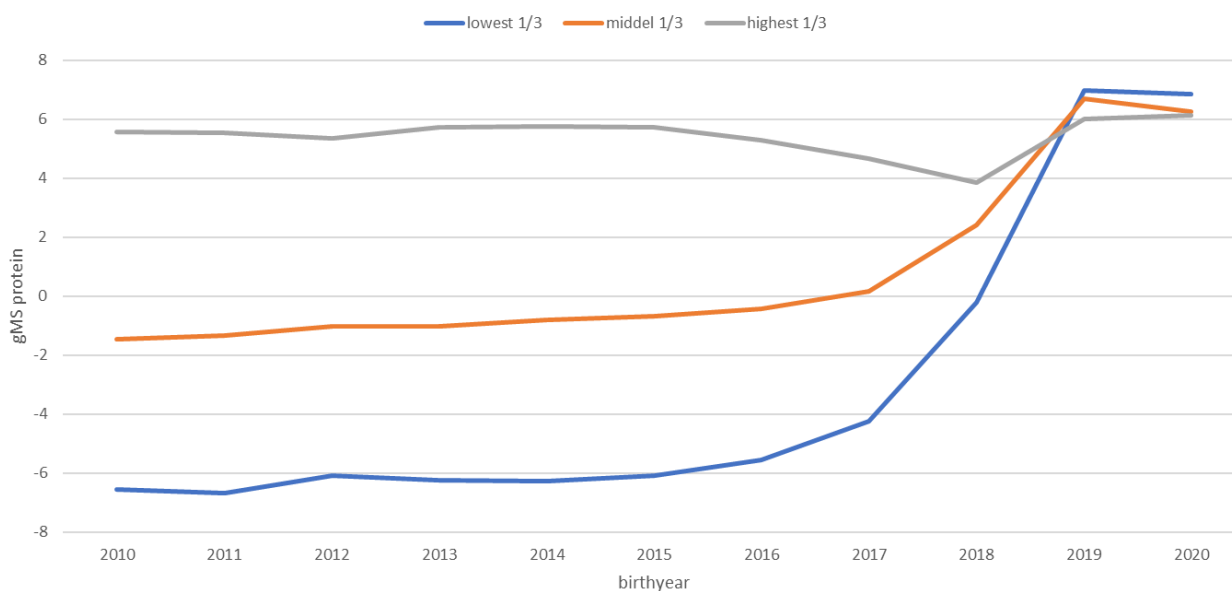


Figure 5 Mendelian Sampling term in kg protein for female grouped by EBVs per birth year class.

### Conclusion

Post processing have reduced the problem with inflation in the TS procedure. Inflation or at least a problem with similar magnitude seems to occur for SS procedure, so that the young genomic bulls have to high GEBVs compared to progeny tested bulls. MS sampling term is unexpectedly above zero for genomic tested females even though almost no selection has occurred before genotyping for youngest birthyear classes. The problem seems to be slightly more pronounced for offspring from progeny tested bulls compared to offspring from genomic bulls. Finally, the too high MS does not seem to be correlated to the genetic level of females without own phenotypes.