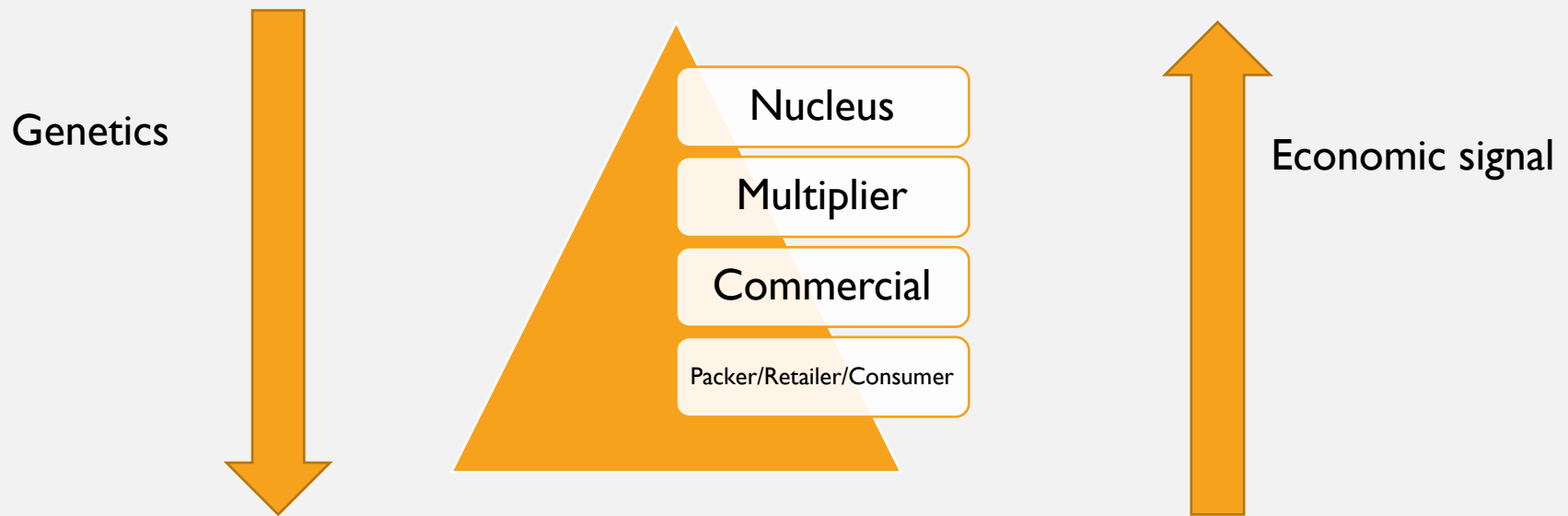


# LEVERAGING COMMERCIAL DATA TO IMPROVE SELECTION AND MANAGEMENT DECISIONS

Matt Spangler

University of Nebraska-Lincoln

# SIMPLIFIED BREEDING PYRAMID



There is substantial lag in both directions

## LONG STANDING DISCUSSION POINT

- A review of those activities, in comparison with the current levels of industry adoption would suggest that the traits not being adequately considered probably includes everything other than growth and calving ease.
- BIF 2018, Focus on Traits not Considered—Dorian Garrick

## DATA HAS A COST/VALUE

- Seedstock sector has paid the cost
- That might not be tenable at scale of commercial data
- Cost could be shared if value is shared and apparent
  - Price for genotyping commercial animals might not need to be exceptionally cheap if cost is shared by all users

# SITUATION

- Many ERT are not captured in seedstock settings.
- This leads to a reliance on indicator traits or simply an omission of EPD for some traits.
- Considerable effort, research and talking, has been placed towards a general suite of traits considered as “novel”
- If a trait is an ERT, it is (likely) not “novel” industry wide.
  - This does not marginalize the importance of closer to biology traits in research settings or emerging trait investigation

## SITUATION

- Phenotypes that are not observed in seedstock settings, but which are profit drivers, are used for management and commerce in commercial animals and thus densely recorded.
- Pedigree is latent
- Phenotypes may be on groups
- Animals change ownership, and management details are lost as they move through the system
- No consistent ID

## CURRENT EFFORTS

- Some partial solutions exist
- Membership/service categories for commercial producers to submit data to breed organizations
- Breeder use of commercial test herds
- Beef x Dairy data
- Commercial genomic tests that leverage genetic evaluations

## (MY) ISSUES WITH CURRENT EFFORTS

- Commercial producers are not seedstock producers
  - Data recording may be different
  - Manage groups, not individuals
  - Notion of sequential culling may be different (model requirements)
- Prediction of commercial animals without leveraging their phenotypes reduces the ROI to the industry of the cost of the prediction
- Returning an EPD is completely insufficient.



## COMMERCIAL DATA WORTH THE EFFORT

- Traits which are not observed in seedstock (carcass, health, sustained fertility)
- Potential for  $rg < 1$  (fertility  $\rightarrow$  sync vs no)
- Possibility of adding substantially more variants without substantially more phenotypes to train on.
- The “traditional” breeding pyramid could be dynamic
- Improvement of downstream decision making
- Improve efficiency of beef production

## POTENTIAL FOR SYNERGY

- Commercial industry wants predictions for trait complexes like health, etc.
- Genetic evaluations can provide these solutions with data from commercial sector
- Coupling genetic evaluation and predictions to commercial industry enable improvement of next generation, not just management of current population

## STARTING POINT: GENETIC EVALUATION CHANGES

- Systematic effects need to be parsed → See Thallman BIF 2023
- Correlated traits related to sequential culling need to be flexible
- Trait definitions, age windows, need to be flexible

Source	Carcass Wt.	Marb. Score
Diet	2530 ± 769	1717 ± 543
Genetic	1069 ± 142	6025 ± 465
Pen	616 ± 114	377 ± 98
Residual	1548 ± 138	3980 ± 418

## POTENTIAL SOLUTIONS

- Individual animals
  - Confirmed paternity/parentage
  - Chip genotyping
  - LPS
  - Phenotypes can be individual or means
- Groups of animals
  - Pooled samples

## INDIVIDUAL ANIMALS

- Do not need them all
- Random within known breed
  - Random to aid in VCE
- Products of AI sires most helpful
- Could be aided by national animal ID and traceability

## GROUPS OF ANIMALS

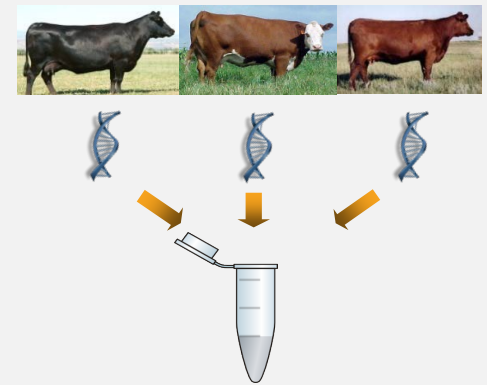
- Commercial animals are managed in groups
- Many phenotypes only recorded at a group level (feed intake)
- Can genotype as a group as well
  - Pooling allele frequency
    - Normalized intensity of red and green signals (McDanel et al., 2012)

# POOLING FOR PREDICTION

- Idea not new
- Kinghorn (2012) → Genomics to manage livestock
- Reverter et al. (2016) → Categorical fertility traits in cattle
- Bell et al. (2017) → GEBV of sires from pooled offspring

# SINGLE-STEP FRAMEWORK

- Recover latent kinship ties
  - Direct ties to genetic evaluation programs
  - Group means may be useful (DNA pooling)
    - Especially when group means are different





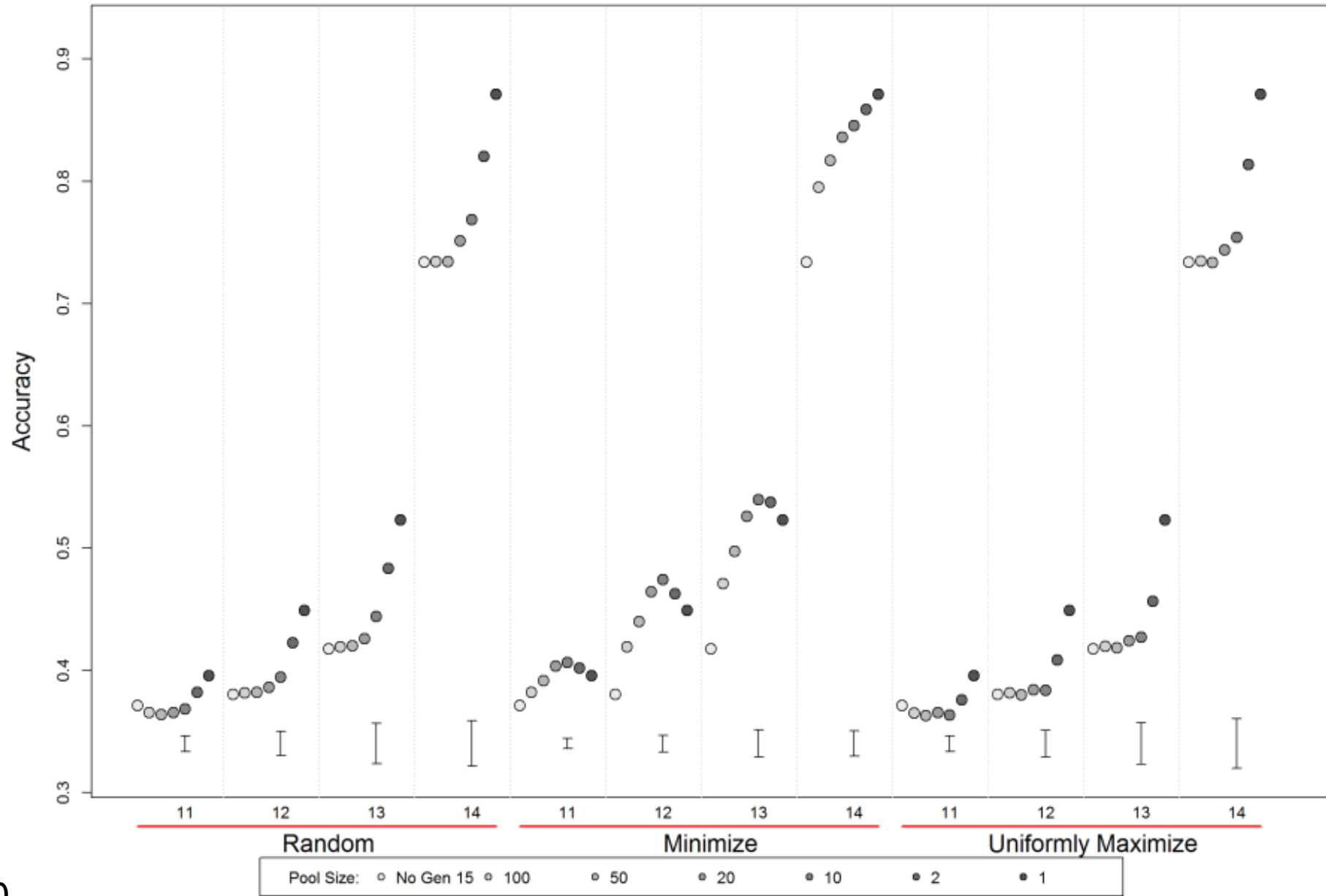
# COMMERCIAL DATA RECOVERY - POOLING

- Improve genetic evaluations and commercial management with genomics
  - DNA Pooling as an interim strategy
    - Cheaper genotyping will open more possibilities
  - Genomically enhanced predictions of groups
- Initial UNL/USMARC investigations
  - Larger pool sizes reduce pool construction error (~20 individuals)
  - Minimizing phenotypic variance within pools leads to greater accuracy

## SSGBLUP MODEL CHANGES

- Individual observations
  - $y = \mathbf{X}b + \mathbf{Z}u + e$
  - $u \sim N(0, \mathbf{H}\sigma_u^2)$
  - $e \sim N(0, \mathbf{I}\sigma_e^2)$
- Individual and pooled observations
  - $y^* = \mathbf{X}^*b + \mathbf{Z}^*u^* + e^*$
  - $u^* \sim N(0, \mathbf{H}^*\sigma_u^2)$
  - $e^* \sim N\left(0, \text{diag}\left(\frac{1}{n_i}\right)\sigma_e^2\right)$

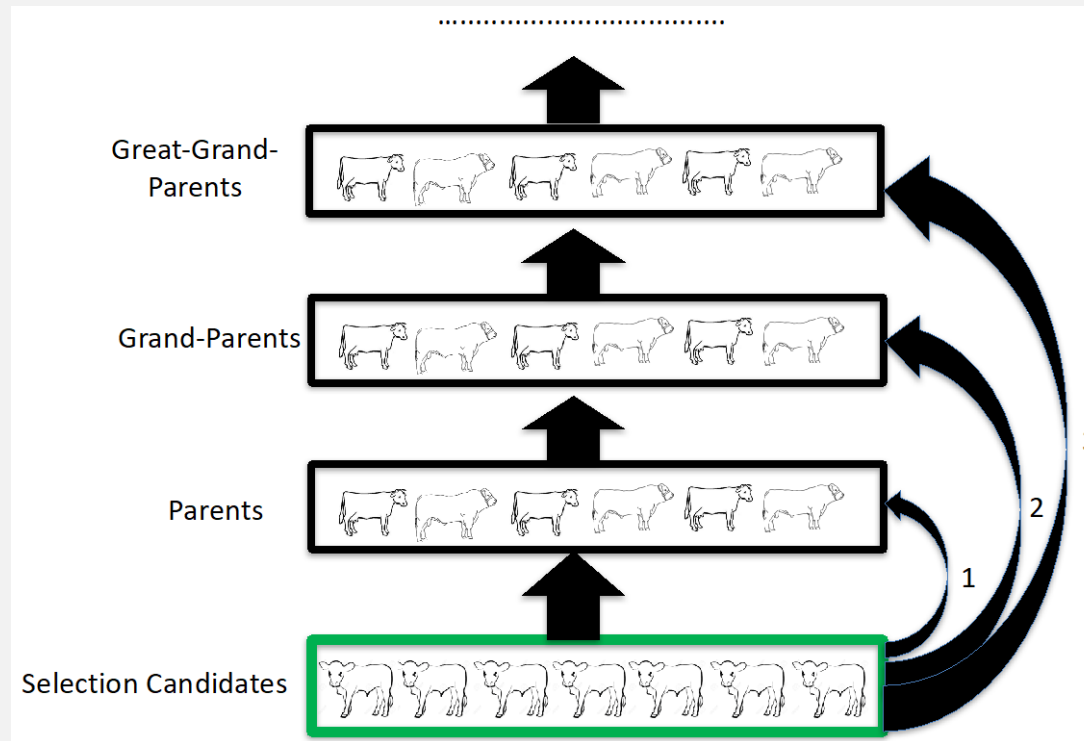
### Sires born in Generation 14 ( $\bar{n}=40$ )



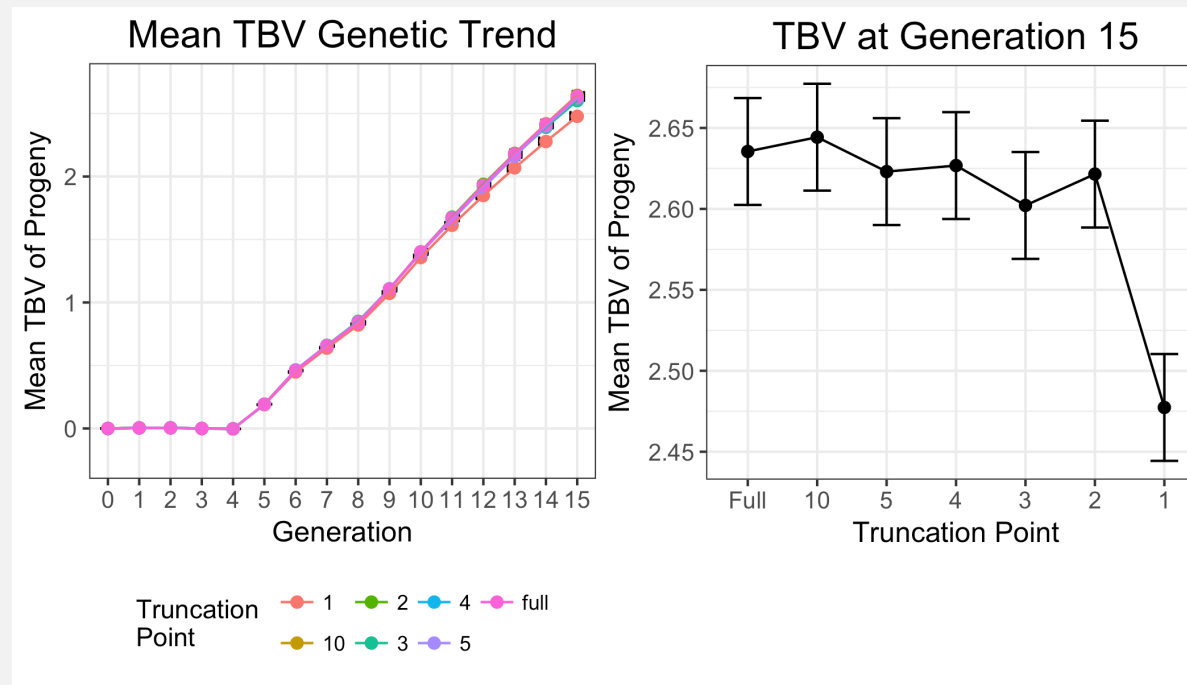
## PRACTICAL POOLING ISSUES

- How to group
- When to group
- How many in a pool
- Resolving heterosis
- Commercialization of pooled genotyping

# NOT ALL DATA EQUALLY VALUABLE



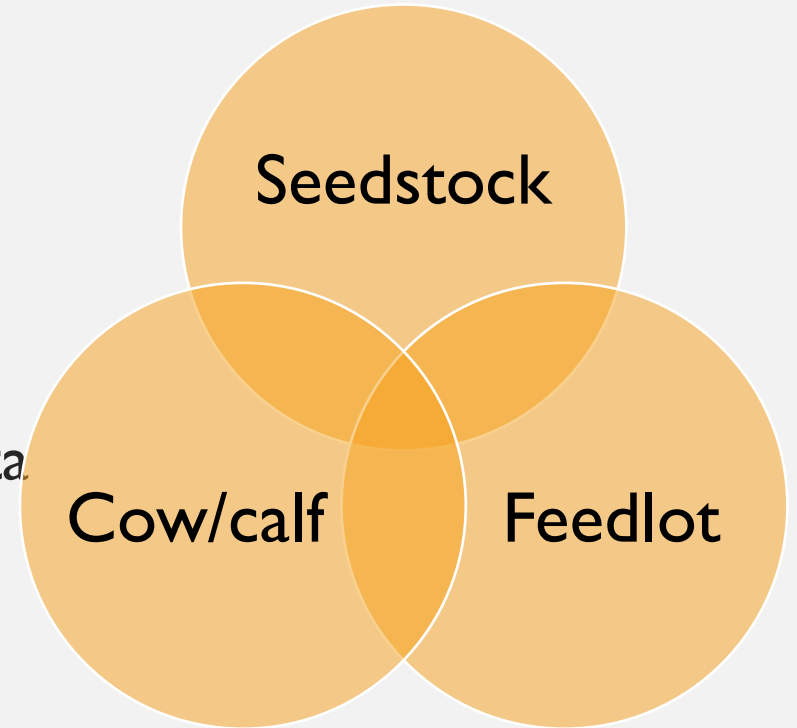
- Truncating data from 2 or more generations back did not impact TBV at generation 15.



Howard et al., 2018 J.Anim. Breed. Genetic.

# DATABASES

- Data could flow directly to genetic evaluation
- Similar to how seedstock producers interact now
- Data could be stored in segment/entity database
- Accessed via API
- Caution in using external EPD if informed by shared data.



# CHANGES TO GENETIC EVALUATION

- Input data formats and flow
  - Volume, group data, ID convention
- Output data formats and flow
  - Predictions more than EPD
- Modelling
  - Contemporary group now broken into components
  - Pooled data
  - Many indicator traits not available on commercial animals (no sequential culling bias)



## DELIVERABLE TO COMMERCIAL SECTOR

- Given  $P=G+E$  the predictions returned to the commercial sector should contemplate all known sources of variation
- $P= \text{EBV} + \text{Heterosis} + \text{breed} + \text{Sex} + \text{Age} + \text{Feedlot} + \text{Season} + \text{Diet} \dots$

## FUNDAMENTAL QUESTIONS

- What is the appetite to make garnering commercial data work?
- How cheap can genotyping/sequencing get?
- Is collective bargaining of interest?
- How flexible can genetic evaluation become?

## GENOMICS IS...

- The tool to allow us to gain phenotypes