Department of Animal & Dairy Sciences
UNIVERSITY OF WISCONSIN–MADISON

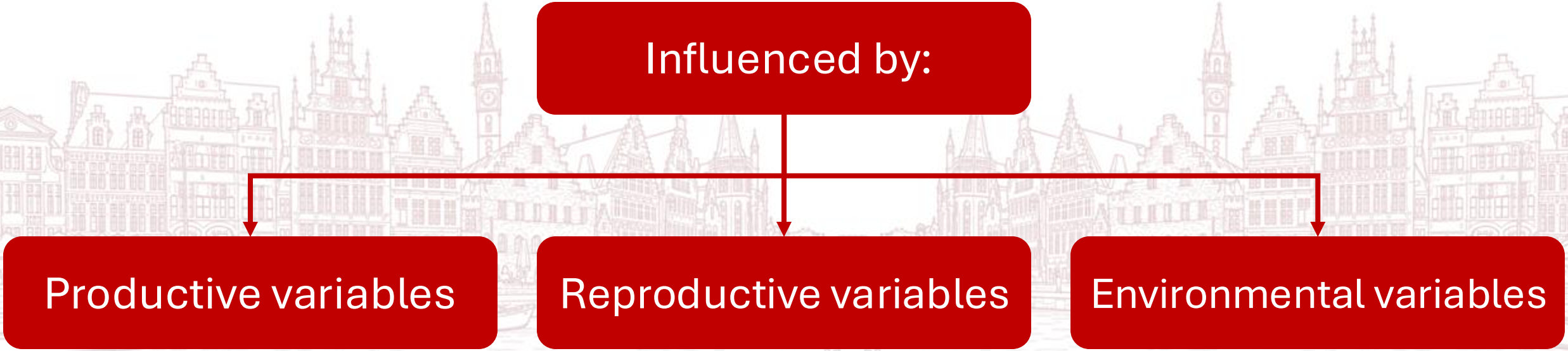
Big Data and Machine Learning to Improve Culling Decision-Making in Dairy Herds

Henrique da Silva, L.; Gong, Y.; **Cabrera, V.E.**

Department of Animal & Dairy Science, University of Wisconsin-Madison

Introduction

- The culling of animals in dairy herds generally results from the interaction between managerial decisions and biological processes inherent to animal physiology and health.



Introduction

- Estimates the probability of survival in future scenarios becomes a strategic tool for decision-making;

- Enables the early identification of critical points that influence culling risks.

Objective

- We aimed to develop two survival curves incorporating a wide range of productive and reproductive variables, using records from farms located in the United States.

Materials and Methods

- 6,729,954 individual records;
- 2680 Dairy herds in the United States;
- 2006 to 2017;
- Council on Dairy Cattle Breeding (CDCB);
- Jupyter using Python language.

Materials and Methods

- XGBoost AFT regression algorithm (GB);
- Kaplan–Meier estimator (KM);
- MAE, RMSE, and R^2 metrics.

Results

Table 1: Comparison of performance metrics for the Kaplan–Meier estimator and XGBoost AFT regression models.

Algorithm	R^2	MAE (years)	RMSE (years)
XGBoost AFT regression	0.74	0.14	0.18
Kaplan–Meier estimator	0.99	0.01	0.01

Results

Table 1: Comparison of performance metrics for the Kaplan–Meier estimator and XGBoost AFT regression models.

Algorithm	R^2	MAE (years)	RMSE (years)
XGBoost AFT regression	0.74	0.14	0.18
Kaplan–Meier estimator	0.99	0.01	0.01

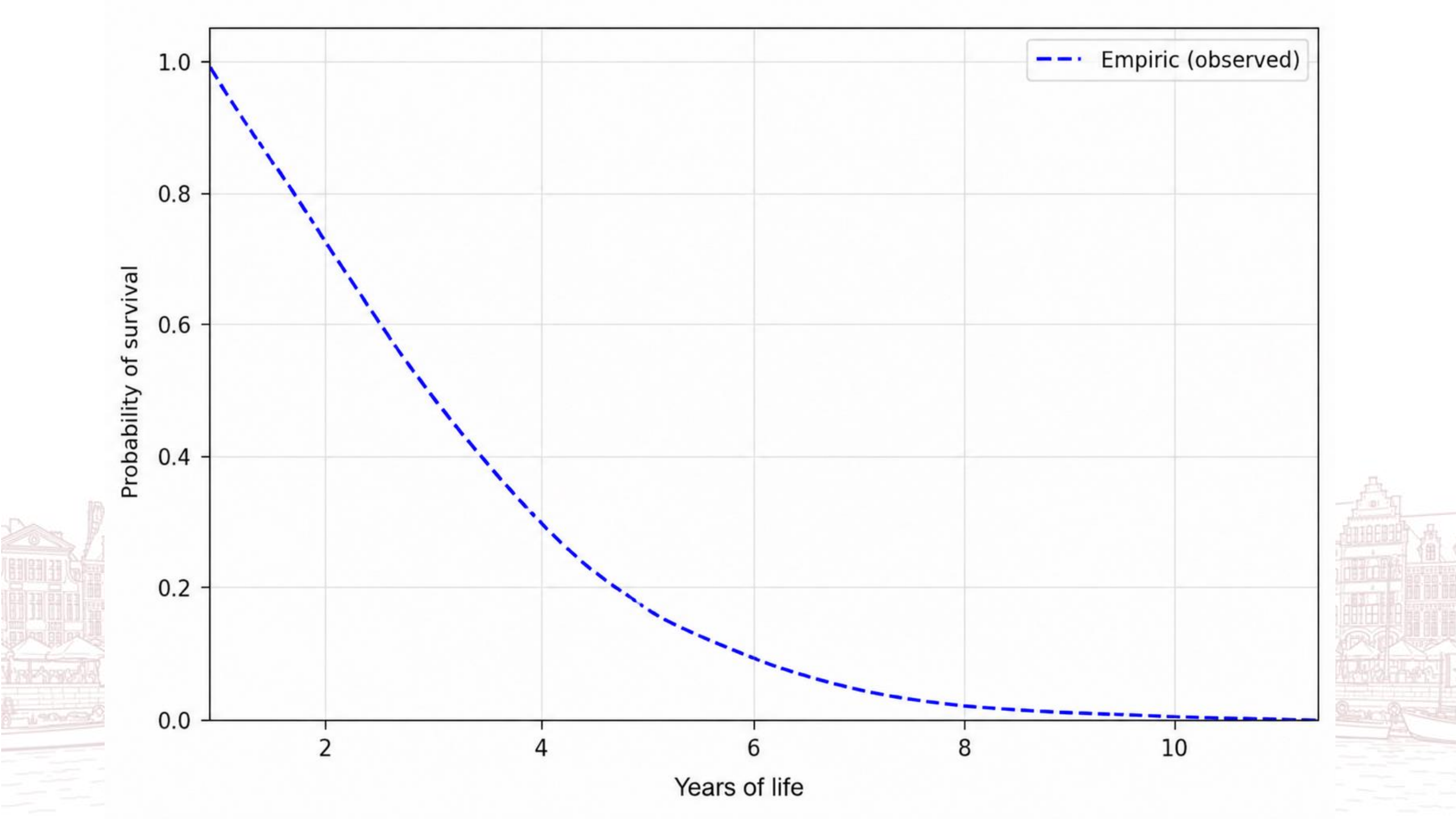
- The KM model showed superior performance, with metrics close to the curve generated with the empirical data.

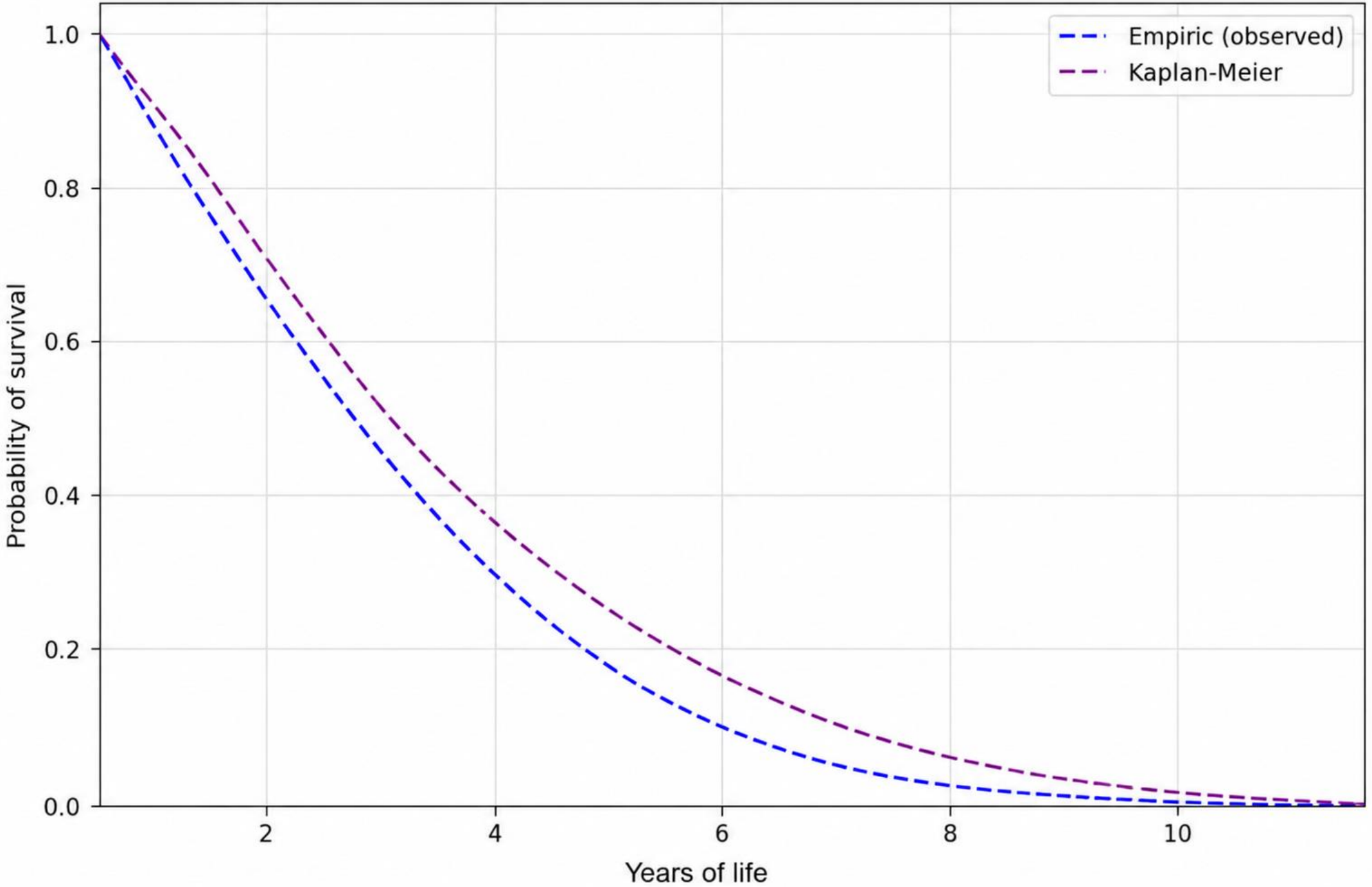
Results

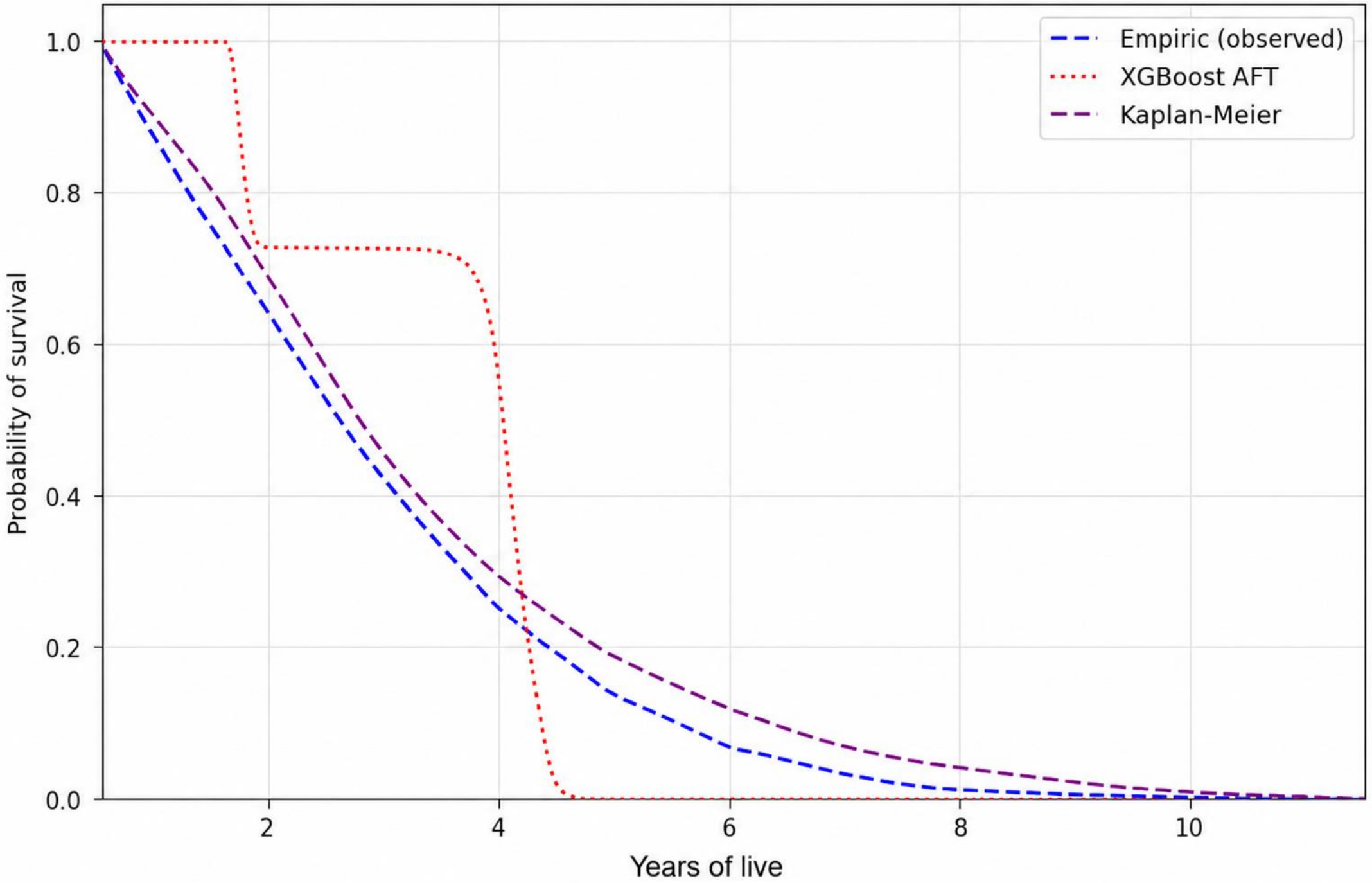
Table 1: Comparison of performance metrics for the Kaplan–Meier estimator and XGBoost AFT regression models.

Algorithm	R^2	MAE (years)	RMSE (years)
XGBoost AFT regression	0.74	0.14	0.18
Kaplan–Meier estimator	0.99	0.01	0.01

- The GB model yielded inferior metrics. However, it remains technically viable for survival predictions.


Empiric (observed)
Probability of survival
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
Years of life


Empiric (observed)
Kaplan-Meier
Probability of survival
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
Years of life


Empiric (observed)
XGBoost AFT
Kaplan-Meier
Probability of survival
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
Years of live

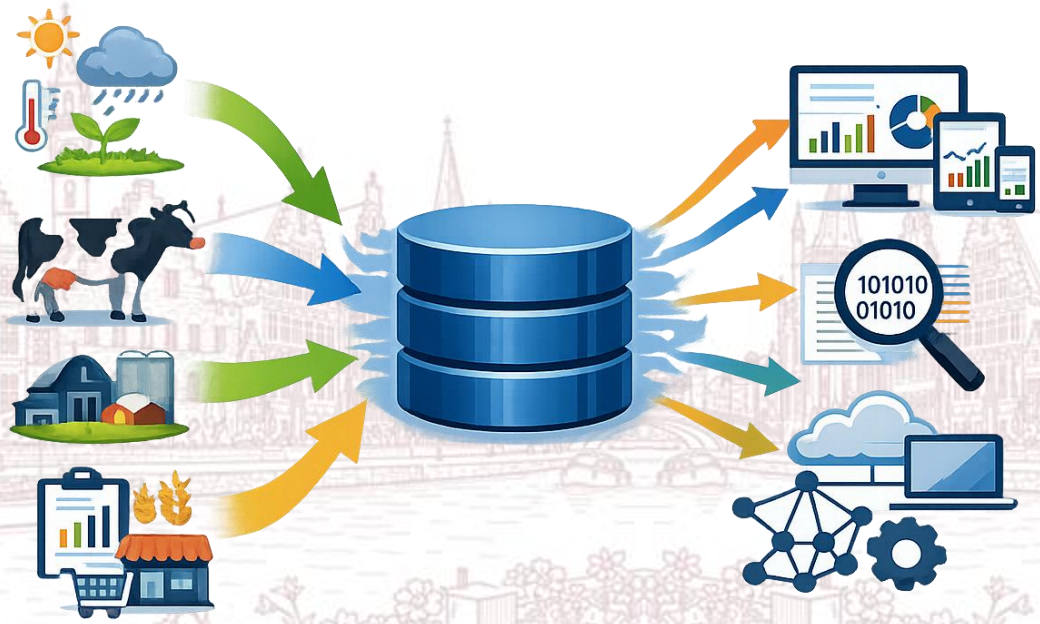
Results

- Although GB presented lower metrics than KM, its performance was satisfactory;

- In some circumstances, its use could be justified because it allows:
  - Integration of a larger number of productive, reproductive, and managerial variables, which is not possible with the KM model
  - Better representation of specific production units (herds/farms)

Results

- Thus, the tested machine learning–based model (GB) offered a promising path for more realistic and individualized survival estimates, incorporating heterogeneity among cows and management practices.



Conclusions

- Both models demonstrated strong predictive capacity;
- The GB model may allow more accurate predictions tailored to the unique reality of each producer.
- The GB model appears promising when we consider a range of different data sources.

Department of
Animal & Dairy Sciences
UNIVERSITY OF WISCONSIN-MADISON

Thank You

vcabrera@wisc.edu
Phone: +1 608-2658506