Journal of Animal Science, 2023, **101**, 1–12
https://doi.org/10.1093/jas/skad099
Advance access publication 2 April 2023
Animal Behavior and Cognition

Effects of dairy cow temperament on milk yield: a systematic review and meta-analysis

Maria G. Marçal-Pedroza[†,‡], Maria Eugênia Andrighetto Canozzi[‖], Mariana M. Campos[$], and Aline C. Sant'Anna[†,1,2,]

[†] Nucleus of Studies and Research in Ethology and Animal Welfare (NEBEA), Department of Zoology, Institute of Biological Sciences, Federal University of Juiz de Fora, 36.036-330, Juiz de Fora, MG, Brazil
[‡]Postgraduation Program in Biodiversity and Nature Conservancy, Federal University of Juiz de Fora, 36.036-330, Juiz de Fora, MG, Brazil
[‖] Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental La Estanzuela, Colonia 70000, Uruguay
[$]Brazilian Agricultural Research Corporation, Embrapa Dairy Cattle, Minas Gerais, 36038-330, Brazil.

[1] CNPQ Researcher.
[2]Corresponding author: aline.santanna@ufjf.br

Abstract

The temperament of dairy cows interferes in milk yield and quality, but there is a lack of consensus throughout the literature. Thus, systematic review (**SR**) and meta-analysis (**MA**) methodologies were used to assess the effects of dairy cow temperament on milk yield. Our literature search included four electronic databases (CABI Abstracts, Web of Science, PubMed, and Scopus) and bibliographies of the publications included on MA. As inclusion criteria, we considered publications about the temperament of lactating cows and its effect on daily milk yield and total milk yield (whole lactation). A random effect-MA was carried out separately for daily milk yield and total milk yield related to each class of cows' temperament, 'low' (low reactivity, calm animals), 'intermediate' (intermediate reactivity), and 'high' (high reactivity, reactive animals). A total of eight publications reporting 75 trials were included in the analyses for daily milk yield, and three publications reporting nine trials for total milk yield. For daily and total milk yield the heterogeneity between publications was high (I^2 = 99.9%). Cows of European breeds with intermediate temperament produced less milk daily than the calm (P = 0.020) and reactive ones (P < 0.001). In the case of primiparous cows, those with intermediate temperament produced less milk daily (P < 0.001) than the reactive ones, while for multiparous, the intermediate produced less than calm (P = 0.032) and reactive cows (P < 0.001). Regarding the stage of lactation, cows evaluated throughout lactation with a calm temperament tended (P = 0.081) to produce more milk than the intermediate ones, but less than the reactive ones (P < 0.001). For total milk yield, reactive cows tended to produce more than the calm (P = 0.082) and intermediate (P = 0.001) ones. Among European and primiparous cows, reactive cows produced more than the intermediate (P = 0.001). According to our results, we cannot confirm what we expected, that calmer cows would be the most productive for both daily and total yield.

Lay Summary

Individual differences in the behavior of dairy cows can affect their productive performance. In an attempt to summarize the scientific information available, we conducted a systematic review and meta-analysis to identify the effects of dairy cows' temperament on milk yield. We hypothesize that calmer cows would produce more milk. We found nine publications with quantitative data available to be included in a meta-analysis. Eight additional publications that addressed the topic of interest but did not present data enough to be included in the meta-analysis (i.e., evaluated the relationships between temperament and milk yield using correlations or regressions) were used to perform a qualitative synthesis. The results of our meta-analysis indicated that the reactive cows were more productive than the calm or intermediate ones, contradicting our initial hypothesis. According to the results of the qualitative synthesis, most of the publications reported a negative association between reactive temperament and milk yield, indicating that calmer cows would produce more milk. We concluded that there are divergences in the information available about the temperament and production of dairy cows. We highlight the need for greater methodological and analytical standardization to allow a broader quantitative synthesis of the temperament effects on milk yield.

Key words: behavior, dairy cattle, performance, personality, reactivity

Abbreviations: SR, systematic review; MA, meta-analysis; NOS, Newcastle-Ottawa Scale; MD, mean difference; 95% CI, confidence interval of 95%

Introduction

Animal temperament is a complex trait that encompasses several behavioral aspects. According to Réale et al. (2007), temperament may be understood as the individual differences in the behavior of animals, in response to their environmental circumstances, given that those differences are relatively consistent over time and in distinct situations. In production animals, this trait may be assessed by observing the behavior of the animals during routine handlings, for example in the milking parlor (milking temperament) (Sawa et al., 2017), or through standardized tests, such as flight speed, reactivity in the handling corral, and flight distance (handling temperament) (Sutherland and Huddart, 2012). For dairy cows, the temperament is usually measured based on the cows' reactivity during milking, considering the intensity of reactions to

Received November 12, 2022 Accepted March 29, 2023.

milking procedure, such as leg movements and kicks (Breuer et al., 2000; Rousing et al., 2004).

In dairy cows, temperament has been associated with productivity (milk yield, quality, and milkability); however, this is still a controversial topic. Contradictory results are reported in the scientific literature. Some studies report that calmer cows produce more milk (Sutherland and Dowling, 2014; Hedlund and L¢vlie, 2015; Cerqueira et al., 2017), with higher fat and protein contents (Kruszyński et al., 2013; Antanaitis et al., 2021). Others show that the reactive ones are more productive, with higher milk yield (Rousing et al., 2004; Sawa et al., 2017), milk fat and protein contents (Cziszter et al., 2016) than the calm ones. In addition, there are still studies that do not find association between temperament and productive parameters (Szentléleki et al., 2008, 2015; Orbán et al, 2011; Sutherland et al., 2012). Furthermore, there is a lack of standardization regarding the measurement used to assess the temperament of the animals throughout the studies, which may hinder the comparison of findings.

The behavior of dairy cows and its relationship with milk yield and quality are topics that interest both consumers and producers, due to their relationships with animal welfare, production efficiency, and sustainability of the livestock industry (Risius and Hamm, 2017; van Dijk et al., 2019; Marçal-Pedroza et al., 2021). Moreover, assessing the effects of temperament on performance may contribute to the improvement of animal welfare, as it aids in the identification of new welfare indicators (Neja et al., 2015).

Thus, in this study, we used systematic review (**SR**) and meta-analysis (**MA**) methodologies to explore the influence of dairy cattle temperament on milk yield and quality. We hypothesize that calmer cows would produce more milk. The aim of this study was to evaluate the scientific evidence available in the literature using SR–MA to identify the effect of the dairy cows' temperament on milk yield.

Materials and Methods

Research question and protocol

This is a theoretical study and therefore did not need to be evaluated by an ethics committee. The systematic review followed the PRISMA guidelines (Page et al., 2021). The search strategy was defined based on PICO terms: population, intervention, comparison, and outcome (Brown et al., 2006). For population, we used the terms "lactating cow" or "dairy cow" or "dairy cattle"; for intervention, "temperament" or "reactivity" or "personality"; and for outcome, "milk production" or "milk yield" or "somatic cell count" or "protein" or "fat".

Dairy cow was the population of interest. The interventions were the different temperament types. As comparison, we considered groups of cows classified as different temperaments in 'Low' (lower reactivity class, also referred to as calm animals), 'Inter' (intermediate reactivity class, also referred to as normal animals), and 'High' (higher reactivity class, also referred to as reactive or nervous or excitable or aggressive animals in the publications reviewed). The outcomes of interest were daily milk yield, total milk yield (whole lactation), and milk quality, but the present study will report only the results regarding yield, despite our database search having included all these measures (Figure 1). To be included in our SR, the publications had to assess at least one of the response variables of interest in association with dairy cows' temperament.

A search protocol was previously developed, and screening tools were adapted from forms used in previous studies (Canozzi et al., 2017, 2019) and tested prior to their application.

Search methods for the identification of publications

The systematic literature search was conducted from September to December 2020 in four electronic databases—CABI Abstracts (Thomson Reuters, 1910–2020), ISI Web of Science (Thomson Reuters, 1900–2020), PubMed (MEDLINE, 1940–2020), and Scopus (Elsevier, 1960–2020). Additional searches were carried out using the literature cited from the publications included in the MA to include peer-reviewed publications not identified by the literature search as well as abstracts published in conference proceedings that were relevant to the subject. All references were exported to EndNote Web software (Clarivate Analytics, Jersey, England) to organize and manually remove duplicate references.

Publications selection criteria and relevance screening

We applied the screening in all citations identified by the literature search using three stages. Before starting the screening, four reviewers were previously trained using 30 publications.

In the first stage, we aimed to identify possible citations of interest among those selected by the search. Each citation was evaluated by reading only the title and applying five simple questions (Supplementary material – S1). This stage was carried out by two researchers independently. In the next step, the remaining citations were evaluated by the same two reviewers, assessing the title, keywords, and abstract, based on eight questions (Supplementary material – S2). When both evaluators answered "no" to one or more questions, the citation was excluded, and, in case of conflicting answers, both evaluators would consensually make the decision. A citation was considered relevant when it was peer-reviewed or conference proceedings assessing dairy cows' temperament, and its relationships with milk yield. In this last stage, we did not apply any restrictions to language or year of publication. The Microsoft Excel software was used throughout all screening stages.

Methodological assessment and data collection process

The first and last authors were responsible for the extraction of data from the selected publications. The relevance of the previously selected publications was confirmed by reading them in full.

The evaluated publications were restricted to the languages in which the research team was fluent (English, Spanish, and Portuguese). Data extracted from each publication was divided into characteristics related to population, intervention, measures, and outcome data, in addition to journal name, author(s), year of publication, and original language. The data extraction forms were adapted from previous studies (Canozzi et al., 2017, 2019).

We need to highlight the diversity of methods found within the selected publications, with different ways to assess temperament and data analyses, hindering the summarization of results. Furthermore, some of these papers allowed for only a qualitative analysis of data (Breuer et al., 2000; Rousing et al., 2004; Bertenshaw et al., 2008; Szentléleki et al., 2008;

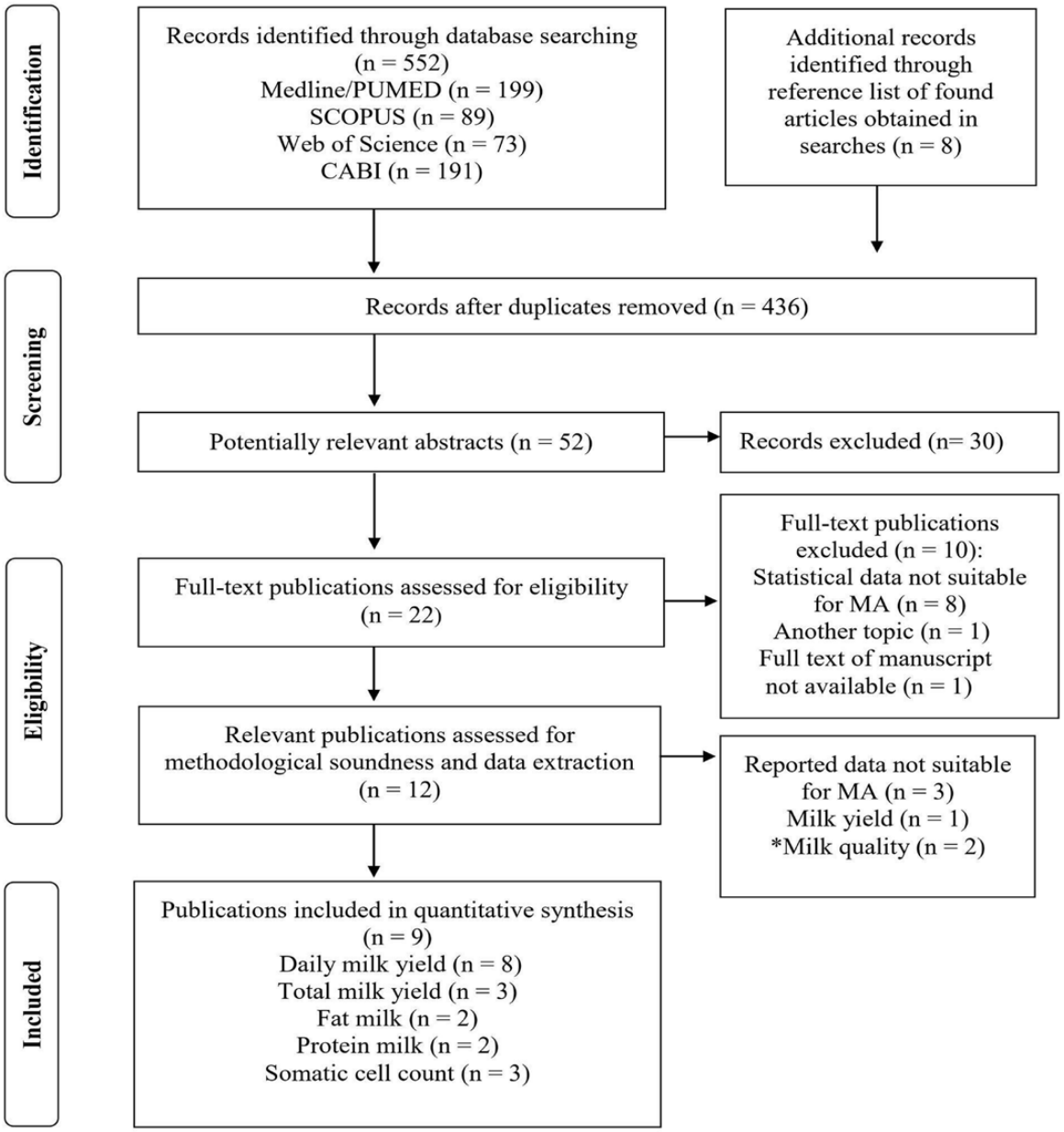

Figure 1. Flow diagram indicating the number of citations and publications included and excluded in each level of the systematic review on temperament of dairy cows and milk yield and milk quality, adapted from PRISMA guidelines (Page et al. 2021). All search results are included in the diagram to allow a better understanding of the total number of records found. *Data from both procedures (milk yield and milk quality) are presented in the flow diagram to allow the researchers to update the same systematic review.

Dodzi and Muchenje, 2011; Sutherland and Dowling, 2014; Hedlund and Lǿvlie, 2015; Cerqueira et al., 2017), as they presented results as correlations and/or regressions, making their inclusion in the MA impossible. Therefore, the included publications were divided into two groups: one for meta-analytical evaluation, and the other for qualitative evaluation.

Considerations for data collection and manipulation

A table with the data were created for each of the results of interest, including mean, standard deviation of mean or another dispersion measure, *P*-value, and the number of evaluated cows in each comparison: (Low vs. Inter), (Low vs. High), and (Inter vs. High), with each comparison for a temperament indicator (measure) being regarded as a 'trial'. For daily yield results, the obtained values refer to the average daily milk yield (in kg/day); and total milk yield (sum of milk yield throughout the whole lactating period, in kg). Some publications presented a greater number of scores and distinct classifications for temperament (Orbán et al., 2011; Gergosvka et al., 2014; Neja et al., 2017), so we standardized them to consider only three temperament types (Low, Inter, High). With these three temperaments, we formed three comparison groups for the analysis of subgroups: group 1 (Low × Inter), group 2 (Low × High), and group 3 (Inter × High).

For two publications that reported only the means values and *P*-values for means comparisons, without a measure of dispersion (Neja et al., 2015; Sawa et al., 2017), an estimate of common standard deviation was calculated using *t*-statistics and assuming the data was normally distributed, based

on the following equation (Ceballos et al., 2009; Mederos et al., 2012):

$$SP = \frac{(x2 - x1)}{t(\alpha dfE)\sqrt{(1/n2) + (1/n1)}}$$

were $\chi_2 - \chi_1$ represents the means difference; $t(\alpha dfE)$ is the percentile of the reference distribution, and n is the sample size of each group.

Quality assessment

The risk of publication bias in the publications was assessed using Newcastle–Ottawa Scale (**NOS**) (Wells et al., 2014). This is an appropriate tool to assess the quality of observational and not experimental randomized trials, based on three main criteria: 'Selection', 'Comparability', and 'Outcome'. The publications receive one 'star' for each quality item included in the criteria of selection and outcome and a maximum of two 'stars' for comparability. In the end, the quality of the publications is expressed on a 9-point scale (Wells et al., 2014).

Meta-analysis

The publications which presented qualitative data that allowed us to estimate the mean difference (**MD**) between the evaluated temperament types and a confidence interval of 95% (**95% CI**) were included in this MA. The statistical analyses were carried out using the Stata V 16.0 software (StataCorp., Texas, EUA).

In subgroup analysis, we carried out an MA separately with datasets consisting of, at least, two individual publications which investigated the same comparative group and the same outcome of interest. The MA results were shown considering MD and 95% CI. Cochran's Q (chi-square test for heterogeneity) and I^2 (percentage of total variation between publications due to heterogeneity and not by chance) were obtained based on the evaluated temperament type (groups 1, 2, and 3) and the outcome variable. The magnitude of I^2 was interpreted in the orders of 25%, 50%, and 75%, and considered as low, moderate, or high heterogeneity, respectively (Higgins et al., 2003).

Publication bias

Publication bias was assessed through a funnel plot and the statistical tests of Begg's correlation and Egger's linear regression. Bias was considered as present based on the visual analysis of the plot and if at least one of the statistical methods was significant ($P < 0.10$). In case there was any indication of the presence of bias, we used the "trim-and-fill" method to estimate its extension (Duval and Tweedie, 2000), which allows us to estimate the number of publications that should be included in the analysis in order for the graph to become symmetrical.

Meta-regression analysis

Univariate meta-regression was performed to identify possible sources of heterogeneity that could influence the results. The variables explored were: year of publication; geographic regions (North America, South America, Europe, Africa, Asia, Oceania); experiment time (days); sample size; racial group (European or Zebu); parity (primiparous or multiparous); lactation stage (beginning = first weeks of lactation or throughout lactation = over the whole lactation); observer effect (unfamiliar person, familiar person or milker); blinding (no, yes, not reported, or not applicable); clustering (no, yes, or not applicable); and identified and controlled confounders (no, yes, or not applicable). The results were reported only for variables that were significant.

Cumulative meta-analysis and influential publications

The cumulative MA was carried out to estimate the effect of the different temperament types on daily and total milk yield each time a new publication was published, to demonstrate the pattern of evidence over time (Borenstein et al., 2009). A sensibility analysis was carried out to check if a certain publication had influenced the effect measurement (MD), by successively removing manually one publication at a time and assessing if MD varied ± 30% after re-inserting the publication and removing the next one.

Results

Publication selection

Our database search identified 552 citations. From that total, 52 were potentially relevant abstracts and 22 were selected for eligibility. Finally, 12 publications were fully read, and among those, nine had their data extracted (Figure 1) and included in this MA, with a total of 84 trials. For daily milk yield, a total of eight publications reporting 75 trials were included, and for total milk yield, it was considered three publications reporting nine trials.

The main characteristics of the included publications are shown in Tables 1 and 2. Three publications were excluded for presenting insufficient data for quantitative analysis (Table 3). We contacted the authors, but no numerical data were obtained, and, since we could not extract them manually, the publication was excluded.

Eight publications evaluated daily milk yield, and three, total yield. The relationship of temperament with daily milk yield was assessed in 26,614 cows, and total milk yield in 23,885 cows.

Risk of bias

The NOS tool was used to analyze the risk of bias, considering the type of publications used in this MA (observational) (Table 4). Of the nine publications included, four (Sutherland and Huddart, 2012; Sutherland et al., 2012; Neja et al., 2015; Sawa et al., 2017) were considered of moderate quality (score between 5 and 7), and the other seven were scored as high quality (scores 8 or 9). This result indicates a moderate to high quality and moderate to low risk of bias in the publications included.

Meta-analysis

In our MA, nine publications were included, six of which evaluated only daily yield and three, daily and total milk yield. The number of publications and types of outcome measures are shown in Table 1. For the analyses, in addition to temperament, the influence of breed, parity, and stage of lactation on milk yield were also evaluated.

Effect of temperament on daily milk yield

The daily yield was the most frequently studied outcome and was shown in eight of the nine publications included in the MA

Table 1. A descriptive summary of each relevant study included in the meta-analysis (n = 9) for daily milk yield and total milk yield.

Reference	Country	Study population (breed/ sample size)	Temperament indicator	*Comparison groups	Outcome parameter
Praxedes et al. (2009)	Brazil	Zebu (Gyr)/ 2.507	Other	Group 1 Group 2 Group 3	Total milk yield
Orbán et al. (2011)	Hungry	Holstein Friesian/ 69 Jersey/ 283	Crush score (reactivity in score in the squeeze chute)	Group 1 Group 2 Group 3	Daily milk yield
Sutherland and Huddart (2012)	New Zealand	Holstein Friesian/ 40	Flight speed (in m/s)	Group 1 Group 2 Group 3	Daily milk yield
Sutherland et al. (2012)	New Zealand	Holstein Friesian/ 30	Flight speed (in m/s)	Group 1 Group 2 Group 3	Daily milk yield
Gergovska et al. (2014)	Bulgaria	Black and White/ 143	Reactivity in scores in the milking parlor	Group 1 Group 2 Group 3	Daily milk yield
Neja et al. (2015)	Poland	Holstein Friesian/ 11.629	Reactivity in scores in the milking parlor	Group 1 Group 2 Group 3	Daily milk yield/ Total milk yield
Neja et al. (2017)	Poland	Holstein Friesian/ 158	Reactivity in scores in the milking parlor	Group 1 Group 2 Group 3	Daily milk yield
Marçal-Pedroza et al. (2020)	Brazil	Zebu-crosses (Girolando)/ 31	Reactivity in scores in the milking parlor/ Steps or kicks/ FSK[1] (or MOV)/ Entrance time/ Crush score/ Flight speed/ Flight distance/ Novel object test	Group 1 Group 2 Group 3	Daily milk yield
Sawa et al. (2017)	Poland	Holstein Friesian/ 12.028	Reactivity in scores in the milking parlor	Group 1 Group 2 Group 3	Daily milk yield/ Total milk yield

* Comparison groups between temperament types, with group 1: low vs. inter; group 2: low vs. high; group 3: inter vs. high. [1]FSK or MOV: Score based on the performance of flinching, stepping, and kicking or sum of the number of kicks and steps during milking.

(I^2 = 99.9%). Mean difference (MD) in daily yield (n = 8 publications, 75 trials) among Group 3 (i.e. Inter vs. High cows) was –0.82 kg of milk/day (95% CI: –1.01, –0.63; P < 0.001), suggesting that Inter cows produced less milk than the High ones, with high heterogeneity among publications (I^2 = 99.4%).

Effect of temperament on daily milk yield considering breed, parity, and lactation stage

For the effect of breed on temperament, only studies with European breed (n = 7 publications, 35 trials) were evaluated, since only one publication assessed Zebu cows. The comparison among Group 1 (n = 6 publications, 35 trials) resulted in an MD of 0.67 kg/milk (95% CI: 0.10, 1.24; P = 0.020), indicating that daily milk yield was lower for Inter than for Low cows, with high heterogeneity between publications (I^2 = 99.9%). In the comparison among Group 3 (n = 6 publications, 35 trials), MD was –1.18 kg/milk (95% CI: –1.41, –0.95; P < 0.001), with Inter cows producing less milk than High. In summary, for studies with European breeds, cows with intermediate temperament produced less milk than the calm and reactive ones.

Among primiparous animals (n = 4 publications, 50 trials) in Group 3, Inter cows produced less milk (MD = –0.74 kg/milk; 95% CI: –0.93, –0.56; P < 0.001) than High ones, with high heterogeneity among publications (I^2 = 96.4%). Among multiparous (n = 6 publications, 25 trials) in Group 1 (n = 4 publications, 25 trials), Inter cows produced less milk (MD = 0.70 kg/milk; 95% CI: 0.07, 1.35; P = 0.032) than Low ones, with high heterogeneity among publications (I^2 = 99.7%). In Group 3 (n = 5 publications, 25 trials), Inter individuals produced less than High ones (MD = –1.08; 95% CI: –1.54, –0.61, P < 0.001), with a 99.8% heterogeneity. So, intermediate cows produced less than the calm and reactive ones, without difference between the last ones.

When assessing the influence of the lactation stage (n = 3 publications, 50 trials) on daily milk yield, we only found significance for experiments carried out throughout lactation, but not at the beginning of lactation. In Group 1 (n = 3 publications, 13 trials), MD was 0.73 kg/milk (95% CI: –0.09, 1.55; P = 0.081), that is, Low cows tended to have a greater daily milk yield than Inter ones, with high heterogeneity among publications (I^2 = 99.7%). In Group 2 (n = 3 publications, 13 trials), MD was –1.01 kg/milk (95% CI: –1.34, –0.68; P < 0.001), Low cows produced less milk than High, with high heterogeneity among publications (I^2 = 97.5%). In Group 3 (n = 3 publications, 13 trials), Inter cows were less productive (MD = –1.24 kg/milk; 95% CI: –1.99, –0.49; P = 0.001) than the High ones, with high heterogeneity among publications (I^2 = 98.2%). In summary, the daily milk yield was higher for reactive, followed by calm and intermediate cows, which had the lowest milk yield.

Effect of temperament on total milk yield

Results for total milk yield were found in three publications (n = 9 trials), with high heterogeneity among publications (I^2 = 99.9%). In Group 2 (n = 3 publications, 9 trials), we obtained

Table 2. Descriptive characteristics of nine publications included in the meta-analyses (MA).

Variable	Categories	Number of publications
Study design	Observational study	7
	Controlled trial	2
Publication type	Peer-reviewed	8
	Conference proceedings	1
Indicator temperament	Reactivity in scores in the milking parlor	5
	Steps or kicks	1
	FSK (or MOV)[1]	1
	Entrance time (in s)	1
	Crush score	2
	Flight speed (in m/s)	3
	Flight distance (in m)	1
	Novel object test	1
	Other	2
Treatment (type of temperament)	Low	9
	Intermediate	9
	High	9
Year of publication	2009-2014	5
	2014-2020	4
Breed	Not reported	0
	European	2
	Zebu/ Zebu-crosses	7
Calving order	Primiparous	3
	Multiparous	4
	Primiparous and multiparous	1
Lactation stage	Not reported	3
	Beginning of lactation	2
	Throughout lactation	4
Housing system	Not reported	3
	Free-stall or tie stall	3
	Loose housing/ open yard	1
	Pastures/ paddock	2
Milking system	Not reported	6
	Herringbone-milking parlor	2
	Parallel-milking parlor	0
	Tandem_milking parlor	0
	Rotary (Carousel) parlor	1
	Robotic milking parlor	0
Who performed the procedure	Not reported	5
	Unfamiliar person, technician, or researcher (authors)	4
	Familiar person or milker	0
	Other	
Outcome assessed	Daily milk yield	8
	Total milk yield	3
Continent	South America	2
	Oceania	2
	Europe	5
Sample size	$N < 100$	3
	$n \geq 100$ and $n < 1000$	3
	$n \geq 1000$	3

[1]FSK or MOV: Score based on the performance of flinching, stepping, and kicking or sum of the number of kicks and steps during milking.

Table 3. List of relevant publications excluded from the final dataset in the meta-analyses (MA).

Reference	Country	Indicator temperament	Temperament type	Outcome parameter	Reason for exclusion
Szentléleki et al. (2015)	Hungry	Reactivity in scores in the milking parlor	Low/ High	Total milk yield	Insufficient numerical data
Kalińska and Slósarz (2016)	Poland	Reactivity in scores in the milking parlour	Low/ Inter/ High	Fat milk/ Protein milk	Insufficient numerical data
Abdel et al. (2017)	Egypt	Reactivity in scores in the milking parlor	Low/ Inter/ High	Daily milk yield/ Total milk yield/ Fat milk/ Protein milk	Insufficient numerical data

Table 4. Risk of bias assessment in the nine studies included in the final dataset of the meta-analyses (MA).

Reference	Selections				Comparability	Outcome			Total
	Adequate definition of temperament groups	Representativeness of the cows used	Selection of divergent temperament groups	Control for disease or incidents that affected the outcome	Adjustment for confounders	Assessment of outcome	Enough time of outcome recording	Adequacy of outcome recording	
Praxedes et al (2009)	☆	☆	☆		☆☆	☆	☆	☆	8
Orbán et al. (2011)	☆	☆	☆		☆☆	☆	☆	☆	8
Sutherland and Huddart (2012)	☆	☆	☆	☆		☆	☆	☆	7
Sutherland et al. (2012)	☆	☆	☆			☆	☆	☆	6
Gergovska et al. (2014)	☆	☆	☆		☆☆	☆	☆	☆	8
Neja et al. (2015)	☆	☆	☆	☆		☆		☆	6
Neja et al. (2017)	☆	☆	☆		☆☆	☆	☆	☆	8
Sawa et al. (2017)	☆	☆	☆			☆		☆	5
Marçal-Pedroza et al. (2020)	☆	☆	☆	☆	☆☆	☆	☆	☆	9

an MD of –1,217.57 kg/milk (95% CI: –2,589.08, 153.94), indicating that Low cows tended (P = 0.082) to produce less milk than the High ones, with high heterogeneity among publications (I^2 = 99.9%). In Group 3 (n = 3 publications, 9 trials), Inter animals had a yield –1,062.45 kg/milk (95% CI: –1,288.35, –836.54; P < 0.001) lower when compared to High ones, with high heterogeneity among publications (I^2 = 99.9%). It indicates that reactive cows produced more milk than the calm and intermediate ones.

Effect of temperament on total milk yield considering breed, parity, and lactation stage

For breed effect, subgroup analysis was carried out only with European breeds (n = 2 publications, 6 trials), since only one publication evaluated Zebu animals. In Group 3 (n = 2 publications, 6 trials), cows of Inter temperament yielded less milk (MD = –414.97 kg/milk; 95% CI: –656.05, –173.90; P = 0.001) than High ones, with high heterogeneity among publications (I^2 = 99.9%).

For primiparous cows (n = 2 publications, 6 trials), we observed difference only for Group 3 (n = 2 publications, 6 trials). High cows produced 414.97 kg (98% CI: –656.05, 173.90; P = 0.001) more milk than Inter ones, with high heterogeneity between publications (I^2 = 99.9%). Among the three publications included, none assessed total milk yield in multiparous cows.

Regarding the lactation stage, only one of the three publications described it, which made such a comparison impossible.

Publication bias

The data included in this MA is quite heterogenous, therefore, results must be interpreted carefully. Both for daily and total milk yield, the asymmetry found in the funnel plot was confirmed by Egger's statistical test (P < 0.001 for both tests), and Begg's test was not significant (P = 0.14; P = 0.75, respectively), with no insertion of new publications by the "trim-and-fill" test.

Meta-regression analysis

Meta-regression results on daily milk yield

Eight publications (n = 75 trials) were inserted in this analysis. Results showed that 99.9% of the variation among

publications was due to chance. None of the eight variables were significantly associated with daily yield, and only three contributed to explaining the variation among publications: sample size (4.6%), lactation stage (4.2%), and identified and controlled confounders (5.5%).

Meta-regression results on total milk yield

Three publications (n = 9 trials) were considered in the meta-regression, and it was evidenced that 99.9% of the variation among publications was due to chance. Meta-regression indicated that with the increase of one year in the year of publication, there was an increase of 233.83 kg in the predicted value (P = 0.050). Publications carried out in Europe showed a 1,905.75 kg (P = 0.019) increase in the predicted value for total milk yield when compared to publications conducted in South America. The number of evaluated animals showed a significant effect, and the increase of one experimental unit rose the predicted value of 0.20 kg of milk (P = 0.022). Publications with animals of Zebu breeds showed a decrease of 1.90 kg in the predicted value (P = 0.019) when compared to those carried out with European cattle. When clustering factors were considered, the predicted value increased by 1.90 kg (P = 0.019) (Table 5).

Cumulative MA and sensitivity analysis

Daily milk yield

In the cumulative MA (2011–2020) for daily yield, there was clear evidence of a change in the estimated yield between temperament groups, going from a positive (MD = 0.16 kg/milk) to a negative value (MD = –0.54 kg/milk). Sensibility analysis showed that removing two publications (Orbán et al., 2011; Sutherland et al., 2012) reduced MD from –0.24 kg to –0.34 and –0.31 kg/milk, respectively. Removing the publication by Neja et al. (2017) increased MD from –0.23 to –0.09 kg/milk.

Total milk yield

In the cumulative MA (2009–2017) for total yield, there was any evidence of changes through the years. Removing the publication by Neja et al. (2015) decreased MD from –796.10 kg to –1,291.86 kg/milk, while removing the publication by Praxedes et al. (2009) increased MD from –796.10 to –171.43 kg/milk.

Qualitative analysis

Some publications assessed the influence of temperament on milk yield using correlation and regression analyses, thus, they were not included in the MA. Due to their relevance, they were considered and analyzed in a qualitative way (Table 6).

All eight publications were carried out with European breeds and evidenced different patterns of relationship between temperament and milk yield. In one of them, milk yield was greater for reactive animals (Rousing et al., 2004), where cows that took more steps in the milking parlor yielded more milk (in kg/day), with Odds Ratios of 1.5 (20-to-30-liter production) and 2.2 (production of over 30 liters). In its turn, Szentleleki et al. (2008) did not find an association between temperament and milk yield using milking reactivity scores as temperament indicators.

Most of the publications (n = 6) reported a negative relationship between temperament and yield, that is, calmer cows produced more milk, as reported by Breuer et al. (2000) (r = –0.38; P < 0.05 for milking reactivity scores); Bertenshaw et al. (2008) (r = –0.25; P = 0.01 for steps); Dodzi and Muchenje (2011) (r = –0.17; P < 0.05 for kicks); Sutherland and Dowling (2014) (r = –0.23; P < 0.05 for milking reactivity scores); Hedlund and L¢vlie (2015) (R^2 = –0.32; P < 0.02 for steps); and Cerqueira et al. (2017) (r = –0.10; P = 0.00 for steps). Bertenshaw et al. (2008) report in the regression analysis, a 7.1% of the variation in productivity occurred due to the number of steps and kicks in the presence of humans (R^2 = 0.07; P < 0.001), which did not occur in the absence of humans (R^2 = 0.002; NS).

Discussion

An SR followed by MA was carried out to quantitatively assess the effects of dairy cows' temperament on milk yield. According to our MA results, calmer cows were not the most

Table 5. Univariate meta-regression results showing significant (P < 0.05) and marginally significant (0.05 ≤ P < 0.10) covariates investigated as potential sources of study heterogeneity for total milk yield. The explained results for each of the covariates included in the meta-analysis are presented for daily production.

No. of studies [1] (trials) [2]	Covariate (trials)	Estimate [3]	*95% CI* [4]	P-value	I^2 (%)	Adj-R^2 (%)
Total milk yield 3 (9)	Null model	–796.10	–1,765.62, –173.41	0.095	99.9	NA
	Publicarion year (9)	233.83	–0.52, –468.18	0.050	99.9	0
	Continent	–	–	0.019	99.9	0
	South America (9)	Reference				
	Europe (9)	1,905.75	413.93, 3,397.57	0.019		
	Sample size (9)	–2,563.72	0.04, 0.36	0.022	99.9	0
	Cattle group (9)	-	-	0.019	99.9	0
	Zebu (9)	Reference				
	Europe (9)	1,905.75	–3,397.57, –413.93	0.019	–	–
	Clustering (9)	-	-	0.019	99.9	0
	No (9)	Reference				
	Yes (9)	1,905.75	413.93, 3,397.57	0.019		

I^2 between-study residual variation; Adj-R2 percentage of the residual variation.
[1] Number of studies included in the meta-regression.
[2] Number of trials included in the meta-regression.
[3] Standard mean difference of the effect size.
[4] These values represent 95% confidence intervals (CI) for the effect size.

Table 6. A descriptive summary of each relevant study (n = 8) that was included in the qualitative synthesis (could not be included in the MA) for daily and total milk yield.

Reference	Country	Study population (breed/ sample size)	Temperament indicator	Outcome parameter
Breuer et al. (2000)	Australia	Holstein Friesian/ 100-200	Reactivity in scores in the milking parlor, steps, and other	Total milk yield
Rousing et al. (2004)	Denmark	Holstein Friesian/ 1.196	Steps, kicks, and other	Daily milk yield/
Bertenshaw et al. (2008)	United Kingdom	Holstein Friesian/ 148	Steps and kicks	Daily milk yield
Szentléleki et al. (2008)	Hungary	Holstein Friesian/ 17	Reactivity in scores in the milking parlour	Daily milk yield
Dodzi and Muchenje (2011)	South Africa	Holstein Friesian/ 7, Jersey/ 7, and crossbred/ 7	Steps, kicks, FD, and FS[1]	Total milk yield
Sutherland and Dowling (2014)	New Zealand	Holstein Friesian/ 150	FSK, FD[1]	Total milk yield
Hedlund and Lǿvlie (2015)	Sweden	Holstein Friesian/ 29, and Swedish Red and White cattle/ 27	Steps, kicks, and NOT	Daily milk yield
Cerqueira et al. (2017)	Portugal	Holstein Friesian/ 2.903	Steps, and kicks	Total milk yield

[1]FD: Flight distance; FS: Flight speed; FSK: Score based on the performance of flinching, stepping, or kicking during milking; NOT: Novel object test.

productive for both daily and total milk yield, against our initial hypothesis. Despite the significant number of publications, only nine had enough information to be included in the quantitative synthesis (MA).

Effect of temperament on daily milk yield

In general, our MA results for daily milk yield evinced that cows classified as reactive (High) produced more than intermediates, and even more than the calm ones (Low), which differed from what we expected. According to Abdel-Hamid et al. (2017), reactive cows, possibly, spend more energy on motor activities, such as walking and standing. Additionally, reactive cows in the milking parlor drop teat cups more often and direct less liquid energy to lactating, which leads to a lower yield (Marçal-Pedroza et al., 2021). However, there are authors who argue that reactive cows are more aggressive during feeding and ingest greater amounts of food, resulting in greater productivity (Sawa et al., 2017). Despite our sensibility analysis not identifying it, the study by Marçal-Pedroza et al. (2020) could be influencing these results, since rumination frequency during milking was used as a temperament measurement. In this particular study, a significant relationship between temperament and milk yield was reported for the behavioral indicator of rumination in the milking parlor. In this specific case, cows classified as High ruminated more during milking, therefore being calmer and more relaxed, and reaching greater milk yield than the Low ones who spent less time ruminating. This classification was different from the other publications included in this MA, in which the High category animals were the most reactive.

The high variability found for the eight analyzed publications may be due to the different methods used to measure reactivity as an indicator of the cows' temperament. This makes it difficult to compare the data in published literature, since some methods may be more sensitive to recording the intensity of the behavioral responses of the animals than others (Sutherland and Huddart, 2012).

The effect of temperament on daily milk yield was assessed considering the subgroups of breed, parity, and lactation stage. Among the evaluated publications, only Marçal-Pedroza et al. (2020) studied Zebu cows. In the European cows, Inter animals produced less than the Low and High ones. For two (Orbán et al., 2011; Sutherland and Huddart, 2012) of the seven publications evaluated in the MA for European breeds, there was no evidence of any effect of temperament on daily milk yield, with only five publications leading to these results. Thus, it is evident that we need to be careful when interpreting results, mainly due to the low number of publications available.

Regarding the effect of parity, primiparous cows of Inter temperament yielded less than those of High temperament. Again, we highlight the work of Marçal-Pedroza et al. (2020), which, by using the frequency of rumination as temperament measurement, primiparous in the High category were the ones with the most rumination and higher milk yield. According to Sawa et al. (2017), the selection of animals to increase productivity may also increase the risk of selecting animals with undesirable temperaments, which might remain in the herd due to their greater milk yield (Praxedes et al., 2009).

Regarding multiparous cows, productivity was lower for Inter than for Low and High cows. In general, multiparous individuals are more used to the milking process, and their reaction to handling may be smaller, which possibly results in better productive performance for the calmer and for reactive ones compared to the intermediates (Sutherland and Huddart, 2012).

When considering lactation stage, the temperament classes differed only throughout the lactation, with a higher daily milk yield for reactive, followed by calm and intermediate cows that had the lowest milk yield. Among the four publications analyzed, two failed to find an influence of temperament on productivity (Orbán et al., 2011; Sutherland and Huddart, 2012), while the other two (Gergovska et al., 2014; Sawa et al., 2017) found greater productivity in High cows,

in a total of 12,068 evaluated cows, and argued that High animals could have yielded more due to greater consumption. Whereas Gergovska et al. (2014) reported that High cows, despite their greater production, had an irregular lactation curve, which does not occur for Low cows.

Effect of temperament on total milk yield

Only three of the publications included in the MA evaluated the effect of animal temperament on total milk yield (over the whole lactation), which may compromise the interpretation of these findings. In general, High cows were more productive than Low and Inter ones. Regarding the breed effect, only two publications with European breeds were considered. In that case, High cows had greater productivity than Inter ones. Moreover, among the primiparous animals, also the High yielded more than Inter ones, possibly due to the previously mentioned relationship between greater feed intake and high milk yield in reactive animals.

Frequently used reactivity indicators for dairy cows have been the number of steps and kicks in the milking parlor (Rousing et al., 2004; Cerqueira et al. 2017; Marçal-Pedroza et al., 2020), but there is no consensus among authors regarding the real interpretation of these movements. Steps may represent a stress indicator, mainly for animals classified as aggressive (Wenzel et al., 2003), or have another meaning, e.g., younger animals with a high parasitic rate (ticks) may take more steps than those with a lower rate, signaling discomfort rather than a more excitable (or reactive) temperament (Rousing et al., 2004). This divergence of interpretation of the animals' temperament may lead to an incorrect association between temperament type and productivity variables. As highlighted by Sawa et al. (2017), the relationship between temperament and milk yield depends on several factors, such as the temperament indicator used, studied breed, age of the animals, and parity.

Meta-regression analysis

Of the eight covariables analyzed (year of publication, geographic region - continent, experiment duration, sample size, breed, parity, lactation stage, and controlled confounders), only three contributed to explaining the variation between publications: sample size, lactation stage, and controlled confounders have shown a direct correlation with the daily milk yield of cows. As for the total milk production, some variables showed an association with milk production, but none of them contributed to explaining the variability found between the publications.

Meta-regression indicated that with every one-year increase in the year of publication, there was an increase in MD, which is possibly related to the period of publication of the selected papers since all nine publications were published starting from the 2000s, a period of growing interest in issues related to behavior, productive performance, and welfare of farm animals (Hemsworth et al., 2000; Rousing et al., 2004; Broom, 2010; van Dijk et al., 2019). Another element we need to highlight is that most studies carried out in Europe showed an increase in MD for total milk yield when compared to studies conducted in South America (Praxedes et al., 2009; Marçal-Pedroza et al., 2020). It could be attributed to the longer period of selection for high productivity in the European breeds, resulting in higher productivity for these animals compared to the Zebu breeds and local crossbreeds used in Latin America (Brito et al., 2021). In spite of the lower milk production, the use of Zebu breeds and their crosses (such as Girolando), more adaptable to warm climates, would result in higher sustainability of dairy production in tropical regions (Canaza-Caio et al., 2016; Brito et al., 2021). The number of evaluated animals had a significant effect, which is probably because the publications had a great variation in sample size (from 30 to 12,028 animals).

For daily milk yield, there was clear evidence of change in the estimated MD, going from a positive value to a negative one, indicating that milk yield increases for the higher temperament classes (Inter and High). The exclusion of the publications by Orbán et al. (2011) and Sutherland et al. (2012) lead to a reduction in MD, but the daily yield of the reactive animals continues to be higher than that of calm and intermediate cows. Both publications together evaluated only 382 dairy cows, all of European breeds. In turn, the exclusion of Neja et al. (2017) resulted in increased MD, also maintaining greater production for reactive cows, and in their study, only 158 animals of European breed were evaluated.

Differently from daily yield, no tendencies were evidenced for total milk yield. The removal of Neja et al. (2015) decreased MD, and it was conducted with 11.629 cows of European breeds, but the total yield of the reactive cows remained higher than the intermediate and calm ones. The opposite happened when we excluded Praxedes et al. (2009), leading to an increase in MD, but the milk yield of reactive cows remained higher. Praxedes et al. (2009) investigated the production of 2,507 animals of Zebu breed, with a lower sample size when compared with the publications by Neja et al. (2015). The last one, published by Sawa et al. (2017), evaluated 12,028 cows. Neja et al. (2015) and Sawa et al. (2017) used European animals, which has possibly led to this variation alongside the fact that Zebu cows, in general, have lower milk yield than European breeds.

Qualitative analysis

The publication of Rousing et al. (2004), which evaluated the cows' temperament based on the number of steps in the milking pen, was the only one to find that High cows yielded more milk, in agreement with our results from MA. For these authors, the occurrence of steps is an indication of discomfort during the milking process, mainly in younger animals, and does not necessarily indicate reactive temperament, which could explain why High cows were more productive. In turn, Bertenshaw et al. (2008) and Dodzi and Muchenje (2011) reported that primiparous individuals which took more steps and kicks while milking were less productive. Hedlund and Lǿvlie (2015) found the same pattern of association with nervous cows producing less milk, which was seen only in the first lactations. Cerqueira et al. (2017), who evaluated multiparous and primiparous cows, observed that the relationship between reactivity and production is associated with parity: cows with a greater number of calvings, i.e., the oldest of the herd, which took more steps, had a lower yield.

The quality of the human-animal relationship during the milking routine is possibly mediating the relationships between temperament and milk yield, as reported by Breuer et al. (2000) and Hemsworth (2003). Therefore, with high-quality handling, based on application of good practices, even the cows with the reactive temperament (more susceptible to stress) might express their best productive potential under adequate environmental conditions (Praxedes et al., 2009; Marçal-Pedroza et al., 2020).

Our SR/MA has some limitations that must be considered. Firstly, the low number of publications found on the subject.

Secondly, some publications which could have been included did not present the data in a format that allowed it to be extracted for a MA. Even after trying to contact the authors to obtain details, as suggested by Lean et al. (2009), we were not successful to reach the numerical data. Additionally, some publications were analyzed separately from the MA in a qualitative manner, due to the relevance of their results. Also, the lack of standardization of the methods of temperament assessment in dairy cows associated with the large variation in productive performance of the animals made the analysis and interpretation of the results a challenging task. Putting it all together, the results obtained in this MA, reporting the greater production by High cows, may be due to how the behavior is interpreted in these studies (reactivity considering the leg movement levels). It is important to highlight the fact that the animals being less agitated, or even still, during the milking procedures does not necessarily mean a calmer temperament, but a fear state (Munksgaard et al., 2001). Understanding animal reactivity as an indicator of temperament type requires, aside from objective measurements, an interpretation of the intrinsic traits of animals, what could be achieved based on the inclusion of physiological measures.

Conclusion

This is the first SR-MA that assessed results published in the scientific literature on the effect of dairy cows' temperament on productivity. Our results of the MA did not support the original hypothesis, as we obtained that reactive cows generally produce greater milk yield than those of calm and intermediate temperament. On the other hand, correlation and regression data support our hypothesis of calm cows being more productive. This contrast leads us to further questions: which indicators should we use to classify animal temperament? And when should this classification be applied? In addition to the need for standardization of protocols for behavioral assessments, in order to allow for a better understanding of the results, and the need for more studies reporting this type of assessment for cows of Zebu breeds.

Acknowledgments

We would like to express our gratitude to Jéssica Tatiana Morales-Pineyrúa (INIA, Uruguay) and Rogério Ribeiro Vicentini (UFJF, Brazil) for contributing to the selection of the publications covered in this review. This work was supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Brasil (Fapemig) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq # 311794/2020-3). The study was part of the doctoral thesis of the lead author of the Graduate Program in Biodiversity and Nature Conservation at the Federal University of Juiz de Fora (UFJF), Juiz de Fora, Campus, Brazil.

Conflict of Interest Statement

The authors have declared no conflicts of interests.

Literatures Cited

Abdel-Hamid, S. E., D. M. A. Fattah, H. M. Ghanem, and E. A. A. Manaa. 2017. Temperament during milking process and its effect on behavioral, productive traits and biochemical, parameters in Friesian dairy cows. *Adv. Anim. Vet. Sci.* 5:508–513. doi:10.17582/journal.aavs/2017/5.12.508.513.

Antanaitis, R., V. Juozaitienė, V. Jonike, V. Čukauskas, D. Urbšienė, A. Urbšys, W. Baumgartner, and A. Paulauskas. 2021. Relationship between temperament and stage of lactation, productivity and milk composition of dairy cows. *Animals*. 11:1840. doi:10.3390/ani11071840.

Bertenshaw, C., P. Rowlinson, H. Edge, S. Douglas, and R. Shiel. 2008. The effect of diferent degrees of 'positive' human–animal interaction during rearing on the welfare and subsequent production of commercial dairy heifers. *Appl. Anim. Behav. Sci.* 114:65–75. doi:10.1016/j.applanim.2007.12.002.

Borenstein, M., L.V. Hedges, J. P. T., Higgins, and H. R. Rothstein. 2009. *Introduction to Meta-analysis*. John Wiley and Sons, Ltd, The Atrium: Chichester, UK.

Breuer, K., P. H. Hemsworth, J. L. Barnett, L. R. MatthewsG. J. Coleman. 2000. Behavioural response to humans and the productivity of commercial dairy cows. *Appl. Anim. Behav. Sci.* 66:273–288. doi:10.1016/S0168-1591(99)00097-0.

Brito, L. F., N. Bédère, F. Douhard, H. R. Oliveira, M. Arnal, F. Peñagaricano, A. P. Schinckel, C. F. Baes, F. Miglior. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal* 15:100292. doi:10.1016/j.animal.2021.100292.

Broom, D. M. 2010. Animal welfare: an aspect of care, sustainability, and food quality required by the public. *J. Vet. Med. Educ.* 37:83–88. doi:10.3138/jvme.37.1.83.

Brown, P., K. Brunnhuber, K. Chalkidou, I. Chalmers, M. Clarke, M. Fenton, C. Forbes, J. Glanville, N. J. Hicks, J. Moody, et al. 2006. How to formulate research recommendations. *BMJ*. 333:804–806. doi:10.1136/bmj.38987.492014.94.

Canaza-Cayo, A. W., J. A. Cobuci, P. S. Lopes, R. A. Torres, M. F. Martins, D. S. Daltro, and M. V. G. B. Silva. 2016. Genetic trend estimates for milk yield production and fertility traits of the Girolando cattle in Brazil. *Livest. Sci.* 190:113–122. doi:10.1016/j.livsci.2016.06.009.

Canozzi, M. E. A., A. Mederos, S. Turner, X. Manteca, C. McManus, S. R. O. Menegassi, S. J. O. J., Barcellos. 2019. Dehorning and welfare indicators in beef cattle: a meta-analysis. *Anim. Prod. Sci.* 59:801–814. doi:10.1016/j.rvsc.2017.07.014.

Canozzi, M. E. A., A. Mederos, X. Manteca, S. Turner, C. McManus, D. Zago, and Barcellos, J. O. J. 2017. A meta-analysis of cortisol concentration, vocalization, and average daily gain associated with castration in beef cattle. *Res. Vet. Sci.* 114:430–443. doi:10.1016/j.rvsc.2017.07.014.

Ceballos, A., J. Sánchez, H. Stryhn, J. B. Montgomery, H. W. Barkema, and J. J. Wichtel. 2009. Meta-analysis of the effect of oral selenium supplementation on milk selenium concentration in cattle. *J. Dairy Sci.* 92:324–342. doi:10.3168/jds.2008-1545.

Cerqueira, J. O. L., J. P. P. Araújo, I. Blanco-Penedo, J. Cantalapiedra, J. T. Sørensen, and J. J. R. Niza-Ribeiro. 2017. Relationship between stepping and kicking behavior and milking management in dairy cattle herds. *J. Vet. Behav.* 19:72–77. doi:10.1016/j.jveb.2017.02.002.

Cziszter, L. T., D. Gavojdian, R. Neamt, F. Neciu, S. Kusza, and D. E. Ilie. 2016. Effects of temperament on production and reproductive performances in Simmental dual-purpose cows. *J. Vet. Behav. Clin. Appl. Res.* 15:50–55. doi:10.1016/j.jveb.2016.08.070.

Dodzi, M. S., and V. Muchenje. 2011. Avoidance-related behavioral variables and their relationship to milk yield in pasture-based dairy cows. *Appl. Anim. Behav. Sci.* 133:11–17. doi:10.1016/j.applanim.2011.04.014.

Duval, S., and R. Tweedie. 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56:455–463. doi:10.1111/j.0006-341x.2000.00455.x.

Gergovska, Z., I. Marinov, T. Penev, and T. Angelova. 2014. Effect of milking temperament on productive traits and SCC in Black-and-White cows. *Int. J. Curr. Microbiol. Appl. Sci.* 3:1–11.

Hedlund, L., and H. Lǿvlie. 2015. Personality and production: nervous cows produce less milk. *J. Dairy Sci.* 98:5819–5828. doi:10.3168/jds.2014-8667.

Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. Measuring inconsistency in meta-analysis. *BMJ*. 327:557–560. doi:10.1136/bmj.327.7414.557.

Hemsworth, P. H. 2003. Human-animal interactions in livestock production. *Appl. Anim. Behav. Sci.* 81:185–198. doi:10.1016/s0168-1591(02)00280-0.

Hemsworth, P. H., G. J. Coleman, J. L. Barnett, and S. Borg. 2000. Relationships between human-animal interactions and productivity of commercial dairy cows. *J. Anim. Sci.* 78:2821–2831. doi:10.2527/2000.78112821x.

Kalinska, C., J. Slósarz. 2016. Influence of cow temperament and milking speed on herd life, lifetime milk yield and reasons of cow culling. *Annals of Warsaw University of Life Sciences-SGGW. Anim. Sci.* 55(2):177–186.

Kruszyński, W., E. Pawlina, and M. Szewczuk. 2013. Genetic analysis of values, trends and relations between conformation and milk traits in Polish Holstein-Friesian cows. *Arch. Anim. Breed.* 56:536–546. doi:10.7482/0003-9438-56-052.

Lean, I. J., A. R. Rabiee, T. F. Duffield, and I. R. Dohoo. 2009. Invited review: use of meta-analysis in animal health and reproduction: methods and applications. *J. Dairy Sci.* 92:3545–3565. doi:10.3168/jds.2009-2140.

Marçal-Pedroza, M. G., M. M. Campos, J. P. Sacramento, L. G. R. Pereira, F. S. Machado, T. R. Tomich, and A. C. Sant'Anna. 2021. Are dairy cows with a more reactive temperament less efficient in energetic metabolism and do they produce more enteric methane? *Animal*. 15:100224. doi:10.1016/j.animal.2021.100224.

Marçal-Pedroza, M. G., M. M. Campos, L. G. R. Pereira, F. S. Machado, T. R. Tomich, M. J. R. Paranhos da Costa, and A. C. Sant'Anna. 2020. Consistency of temperament traits and their relationships with milk yield in lactating primiparous F1 Holstein - Gyr cows. *Appl. Anim. Behav. Sci.* 222:104881. doi:10.1016/J.APPLANIM.2019.104881.

Mederos, A., L. Waddell, J. Sánchez, D. Kelton, A. S. Peregrine, P. Menzies, J. VanLeeuwen, and A. Rajic. 2012. A systematic reviewmeta-analysis of primary research investigating the effect of selected alternative treatments on gastrointestinal nematodes in sheep under field conditions. *Prev. Vet. Med.* 104:1–14. doi:10.1016/j.prevetmed.2011.10.012.

Munksgaard, L., A. M. Depassilé, J. Rushen, M. S. Herskin, and A. M. Kristensen. 2001. Dairy cow's fear of people: social learning, milk yield and behavior at milking. *Appl. Anim. Behav. Sci.* 73:15–26. doi:10.1016/S0168-1591(01)00119-8.

Neja, W., A. Sawa, M. Jankowska, M. Bogucki, and S. Krężel-Czopek. 2015. Effect of the temperament of dairy cows on lifetime production efficiency. *Arch. Anim. Breed.* 58:193–197. doi:10.5194/aab-58-193-2015.

Neja, W., D. Piwczyński, S. KrężeL-Czopek, A. Sawa, and S. Ozkaya. 2017. The use of data mining techniques for analysing factors affecting cow reactivity during milking. *J. Cent. Eur. Agric.* 18:342–357. doi:10.5513/JCEA01/18.2.1907.

Orbán, M., K. K. Gaál, F. Pajor, A. Szentléleki, P. Póti, J. Tőzsér, and L. Gulyás. 2011. Effect of temperament of Jersey and Holstein Friesian cows on milk production traits and somatic cell count. *Arch. Anim. Breed.* 54:594–599. doi:10.5194/aab-54-594-2011.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 372:1–9. doi:10.1136/bmj.n71.

Praxedes, V., R. Verneque, M. Pereira, M. Pires, M. Machado, and M. Peixoto. 2009. Evaluation of factors influencing milk ejection in the Brazilian Gyr dairy cattle. *Braz. Agric. Res. Corp.* 40:142–145.

Réale, D., S. M. Reader, D. Sol, P. T. McDougall, and N. J. Dingemanse. 2007. Integrating animal temperament within ecology and evolution. *Biol. Rev.* 82:291–318. doi:10.1111/j.1469-185x.2007.00010.x.

Risius, A., and U. Hamm. 2017. The effect of information on beef husbandry systems on consumers' preferences and willingness to pay. *Meat Sci.* 124:9–14. doi:10.1016/j.meatsci.2016.10.008.

Rousing, T., M. Bonde, J. H. Badsberg, and J. T. Sorensen. 2004. Stepping and kicking behavior during milking in relation to response in human–animal interaction test and clinical health in loose housed dairy cows. *Livest. Prod. Sci.* 88:1–8. doi:10.1016/j.livprodsci.2003.12.001.

Sawa, A., M. Bogucki, W. Neja, and S. Krężel-Czopek. 2017. Effect of temperament on performance of primiparous dairy cows. Ann. *Anim. Sci.* 17:863–872. doi:10.1515/aoas-2016-0085.

Sutherland, M. A., and F. J. Huddart. 2012. The effect of training first-lactation heifers to the milking parlor on the behavioral reactivity to humans and the physiological and behavioral responses to milking and productivity. *J. Dairy Sci.* 95:6983–6993. doi:10.3168/jds.2011-5211.

Sutherland, M. A., and S. K. Dowling. 2014. The relationship between responsiveness of first-lactation heifers to humans and the behavioral response to milking and milk production measures. *J. Vet. Behav. Clin. Appl. Res.* 9:30–33. doi:10.1016/j.jveb.2013.09.001.

Sutherland, M. A., A. R. A. R. A. R. Rogers, and G. A. G. A. Verkerk. 2012. The effect of temperament and responsiveness towards humans on the behavior, physiology and milk production of multiparous dairy cows in a familiar and novel milking environment. *Physiol. Behav.* 107:329–337. doi:10.1016/j.physbeh.2012.07.013.

Szentléleki, A., J. Herve, F. Pajor, D. Falta, and J. Tozser. 2008. Temperament of Holstein Friesian cows in milking parlour and its relation to milk production. *Acta Univ. Agric. Silv. Mendel. Brun.* 56:201–208. doi:10.11118/actaun200856010201.

Szentléleki, A., K. Nagy, K. Szeplaki, K. Kekesi, and J. Tozser. 2015. Behavioural re sponses of primiparous and multiparous dairy cows to the milking process over an entire lactation. *Ann. Anim. Sci.* 15:185–195. doi:10.2478/aoas-2014-0064.

van Dijk, L., H. J. Buller, H. J. Blokhuis, T. van Niekerk, E. Voslarova, X. Manteca, C. A. Weeks, and D. C. J. Main. 2019. HENNOVATION: Learnings from promoting practice-led multi-actor innovation networks to address complex animal welfare challenges within the laying hen industry. *Animals*. 9:24–38. doi:10.3390/ani9010024.

Wells, G., B. Shea, D O'Connell, J. Peterson, V. Welch, M. Losos, P. Tugwell, P., 2014. *The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of Non-Randomized Studies in Meta-analyses*. Ottawa (ON): oxford.asp&. Ottawa Health Research Institute. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed September 27, 2022

Wenzel, C., S. Schönreiter-Fischer, and J. Unshelm. 2003. Studies on step-kick behaviour and stress of cows during milking in an automatic milking system. *Livest. Prod. Sci.* 83:237–246. doi:10.1016/s0301-6226(03)00109-x.