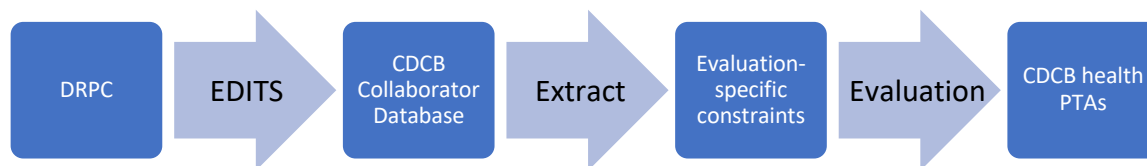# CDCB Guidelines for Health Data Management

By Kristen Gaddis, PhD

Health evaluations are a relatively new offering from the CDCB, at least compared to traits such as yield that have been evaluated for over more than 50 years. Evaluations for six common health events were introduced by CDCB for Holsteins in 2018 and Jerseys in 2020 and planned for Brown Swiss in 2022. These traits are not only relatively new, but they also rely on a separate stream of data consisting of health events reported on farm. The following description includes the flow of data to CDCB, how these data are processed, and how the data used in genetic evaluations are selected.

DRPC → EDITS → CDCB Collaborator Database → Extract → Evaluation-specific constraints → Evaluation → CDCB health PTAs

Data flow to CDCB

Health data are received by CDCB as Format 6 records (Format 6 documentation). Although only six events currently receive a genetic evaluation, we can accept any of the 20 health event codes and 4 management codes as described in the documentation. Health and management events are 2 to 4 characters in length and should also be associated with an event date (YYYYMMDD) and optional details, along with information similar to that provided in Format 4 records (e.g., ID, birthdate, sire and dam information, herd, calving date, etc.). The Format 6 documentation was most recently updated in 2017, which primarily incorporated updated and expanded detail options for health events. Use of the detail section is not required for the record to be deemed usable at this time. Up to 20 health events can be reported in a single Format 6 record for a specific cow-lactation. Up to 50 health events can be stored in the CDCB Collaborator database for any cow-lactation.

Preliminary editing of health events

Prior to storage in the CDCB Collaborator database, all submitted phenotypic data must pass a series of preliminary checks, Format 6 records included. These checks are handled by the CDCB "EDITS" system which provides a first level of quality control for phenotypic data. As a result of this process, records may be either accepted, rejected, or modified. A sample of example

reasons for error are provided in Table 1 below for health events. In addition to confirming that health segments are acceptable, extensive checking is performed on the remaining information provided in a Format 6 record. The length of the record is checked. Animal identification is verified, as well as sire and dam identification. Birthdate information is confirmed, and cross-reference data is checked. Data provided in the submitted Format 6 record is compared to what is currently available in the Collaborator database for consistency. These checks are completed for each Format 6 file submitted to CDCB. Upon completion of this process, output files are produced that include rejected records along with the respective error code. Data providers can then use that information to correct the record and resubmit. Error code documentation is available through the CDCB website (https://queries.uscdcb.com/formats/geterr.cfm). Accepted records are uploaded to the CDCB collaborator database and stored with the appropriate lactation record.

Table 1. Example reasons for modification or rejection – health events

Error	Action
Number of health event segments does not agree with length of record	Modify
Cow already has 50 health events	Reject
Type of health code is invalid	Reject
Date of health event is non-numeric	Reject
Date of health event is after current date or processing date	Reject
Date of health event is before calving date	Reject
Date of health event is after the next calving date	Reject
Date of health event is not the same as calving date for dystocia	Modify
Date of health event cannot be in herd designated and is not with certainty another herd	Reject
Health event source code is invalid	Reject
Health event source code is blank	Modify
Prevent non-DRPC source from updating DRPC source health event	Modify
Detail is invalid. Detail information ignored.	Modify

Data selection for evaluations

Much like other trait evaluations calculated by CDCB, additional constraints are placed upon the stored health data in order to select those to be used for the evaluation. General constraints in the initial extract of data include dairy cows with lactations since 1990, parities 1 to 5, pedigree information for both sire and dam, a calving date at least 1.5 years after her birthdate, and an overall lactation days in milk (DIM) of less than 700 days. A constraint of 700 DIM is used to eliminate lactation records that are likely two separate lactations.

Constraints are applied individually for each of the six health events currently evaluated by CDCB which include milk fever, displaced abomasum, ketosis, mastitis, metritis, and retained placenta. All records of any of these six events are extracted from the collaborator database for

further analysis. Healthy contemporary cows for each health event are identified as cows with a lactation record in the herd-year without the health event of interest reported. For each event, the first incidence reported most immediately following calving is retained. Thus, only one record per event per cow-lactation are retained as a possible phenotype for the evaluation, resulting in a 0 or 1 binary trait. Records are required to occur within an event-specific time frame, as shown in Table 2. For a cow's record to be included, either with a health event or as a healthy contemporary, her lactation DIM must exceed the individual event constraint. This ensures that each included cow has equal opportunity to express the health event of interest.

Table 2. Time (days in milk) constraint for each health event and overall incidence

Health event	Days in milk	Overall Incidence (%)
Milk fever	30	1
Displaced abomasum	60	2
Ketosis	60	4
Mastitis	210	10
Metritis	90	6
Retained placenta	10	3

Constraints are also implemented on a herd-year basis. Each herd-year must have at least 5 cows with lactation records. For each herd-year and event, there must be at least one reported incidence. This ensures that we include only herd-years that are recording the health event. Incidence by herd-year and event is also calculated and compared to the overall population average. Herd-years with a reported health event incidence greater than 3 standard deviations beyond the population incidence is deemed to be over-reporting the event and excluded from further evaluation for that health event. These herds may be using the health event acronym as a management indication as opposed to reporting true incidences of the event. Under-reporting is also considered on a herd-year and event basis. Records from herd-years with a reported incidence of a health event less than 10% of the overall population incidence are excluded from further analyses for that health event. Current incidence rates for each evaluated health event are provided in Table 2.

A summary of these steps is shown as a flow chart in Figure 1 to exemplify how a record may flow through the health event-specific editing constraints. Records that were available in the collaborator database and meet the additional constraints in place for data extraction are used as phenotypes in CDCB's official evaluations of health traits.

Figure 1. Flow chart of data editing to select phenotypic records for genetic evaluation.

Collaborator database

Lactations since 1990

Extract data

✓ Parities 1 to 5
✓ Pedigree data
✓ DIM < 700
✓ Calving date – Birth date > 1.5 years

Milk fever | Displaced abomasum | Ketosis | Mastitis | Metritis | Retained placenta

Lactation DIM ≥ 30 | Lactation DIM ≥ 60 | Lactation DIM ≥ 60 | Lactation DIM ≥ 210 | Lactation DIM ≥ 90 | Lactation DIM ≥ 10

Event = incidence reported within 30 d of calving | Event = incidence reported within 60 d of calving | Event = incidence reported within 60 d of calving | Event = incidence reported within 210 d of calving | Event = incidence reported within 90 d of calving | Event = incidence reported within 10 d of calving

Healthy = no incidence reported within 30 d of calving | Healthy = no incidence reported within 60 d of calving | Healthy = no incidence reported within 60 d of calving | Healthy = no incidence reported within 210 d of calving | Healthy = no incidence reported within 90 d of calving | Healthy = no incidence reported within 10 d of calving

Calculate population incidence for each health event $(\bar{X})$

Compare herd-year incidence
✓ HY incidence < $\bar{X}$ + 3 SD
✓ HY incidence > $0.10\bar{X}$

Repeat for each health event

Milk fever phenotype