

# Data Auditing Tools – Reproduction & Survival

---

## SIL Technical Note

---

Relates to: Characterization of reproduction and survival data held on the SIL database

Written by: Mark Young

Date: 19 June 2007

---

### Summary

- SIL has produced a set of Data Auditing Tools for Reproduction and Survival data
- These can be used by bureaus and breeders to routinely inspect data to monitor flock performance characteristics and to check data for anomalies

### Background

SIL uses sophisticated analyses to estimate genetic merit for key performance traits in sheep. These analyses take account of pedigree and performance information. In addition, they adjust data so known environmental effects do not bias results (e.g. date of birth, birth/rearing rank or mob).

Data integrity is a key factor in getting accurate genetic evaluations. Identifying and correcting data errors is an important part of any system involved in collection, transfer and storage of data. It is also important to assess the amount of data for some traits before deciding on whether to conduct a genetic evaluation for that trait.

Data collection routinely involves people, with varying degrees of skill and experience. Typically, data accuracy is extremely high. However, even the best systems can slip up when it comes to recording performance for individual animals. In addition, errors can occur through no fault of the people collecting the data if equipment fail or information is lost. Many breeders work to minimize human error by collecting and transferring data electronically.

Once data has been entered into the SIL system, it can be used for a genetic evaluation. However, it is useful to verify what data are held by SIL and to check for incomplete or inaccurate data.

SIL has developed Data Auditing Tools to do this. This document considers the **SIL Data Auditing Tools for Reproduction and Survival (DAT-RS)**. These traits are more prone to “errors” since they are derived from information collected over a range of time (from mating through to weaning). Another document will deal with Data Auditing Tools for other traits.

### What are derived traits?

Derived traits are calculated or inferred from other data rather than measured directly. For example weaning weight is measured directly but lamb survival is determined from whether a lamb has a birth fate that indicates it died anytime between birth and weaning. Rearing rank of a lamb is calculated not only from whether that lamb itself survived, but also whether litter mates survived.

Many errors will be picked up at the time data is entered or after the derived traits run. However, some persist because they do not look “wrong” to the SIL database. Usually, the best judgement on whether data looks right is made with input from both the flock owner and their bureau.

### Characterizing the data

It is not realistic to examine the data for every animal on the SIL system in detail. However, summaries of the data are very useful for spotting gross abnormalities, outliers or missing information. These effects can bias results or compromise accuracy of the evaluation.

SIL has a series of tables that characterize data for Reproduction and Survival traits. These focus on three main areas – Mating, Lambing and Survival. A “data audit” can be performed by requesting tables that each focus on particular aspects of the data held by SIL that relate to these events.

The best way to look at the data depends on what the problem is that we want to check for. SIL has produced nineteen tables (3 for Mating, 9 for Lambing and 7 for Survival) that consider the data held from different perspectives. Not all these tables are useful for routine data checking. For this reason, SIL has default sets of tables for different uses e.g. a set from the breeder’s perspective or a set from the bureau perspective (see Appendix 1).

Any set of tables can be requested. The full set is not often needed and it is tedious to look at every table when some may not be informative.

### Specific areas of concern with data for Reproduction & Survival

SIL has used these tools to inspect a large number of flocks to characterize typical issues with data for reproduction and survival as well as with flock structure. These are

- Ensuring ewe fates (DFATE) are fully recorded.
  - Some flocks have few barren ewes. Breeders may not enter barren ewe fates as these ewes are culled and no longer of interest to them. However SIL analyses take account of performance for relatives so this data is needed to assess merit for other animals in the flock. Typically, we see 1-3 % of ewes barren.
  - Many flocks do not record ewes dying between mating and lambing. Typically we expect about 2% of ewes to die in this period.
- Date of birth inconsistencies – outliers may occur (animals born much earlier or much later than the main mob – these may be data entry errors or may require the animals to be set up as a separate lambing mob – your bureau can provide the best advice on what to do about this.
- Incompatible data – such as dead (prior to weaning) lambs that have a weaning weight, or BFATE = dead but STATUS = alive.
- Incomplete data – too few lambs or too few dead lambs coded using the lamb fate code (BFATE). Typical losses from birth to weaning are around 15% but some flocks regularly record less than 5%, suggesting that lamb fate codes are not complete.
- Misidentified sires - birth flock or birth year
- Years not connected through use of a common sire(s).
- Ewe age groups not connected within a year through use of a common sire(s).

Other errors can occur but are less common. If it is suspected that flock data are not correct, the SIL bureau can use DAT-RS tables to characterize flock data and work with the breeder to try and identify where a problem is occurring.

### Routine breeder use

It can be helpful to characterize performance of the flock. The default set of tables for breeder provides information about mating performance, lambing performance and lamb survival.

In addition, two tables show how well years are connected and how well ewe age groups are connected within year, through use of common sires. Establishing such connections is critical to removing biases that can occur in a genetic evaluation when these connections are weak.

Finally, these tables can be used to check whether the SIL database contains the right amount of data and whether performance figures are what as expected. Inspection of the data may indicate that some data are missing e.g. there being less lambs for a particular age group of ewes, or survival

figures being too high. Breeders should address the specific concerns listed previously when checking data for their flock.

### Routine bureau use

SIL bureaus provide a skilled service to their breeder clients. They specialise in handling data know how to specify genetic evaluations and design reports to suit their breeder clients.

Experience with a range of clients and the volume of work they get through usually enables bureau staff to more easily spot data inconsistencies than individual breeders. By combining the data inspection skills of both the breeder and the bureau, data integrity is maximized.

Since bureau staff regularly process large volumes of data, quickly, there is value in having simple summary statistics to characterize the data they have added to the database, or to verify what data are present prior to a genetic evaluation being performed. In the case of Reproduction and Survival, the traits that SIL analyses are “derived” from data the bureau enters.

SIL has a default set of tables which aim to characterize data in a standard way. With experience, bureau staff will be able to use these to flag issues in data that need to be followed up. Bureaus can select any combination of tables in order to best characterize data for specific cases.

Bureaus should address the list of specific concerns given previously when checking flock data.

NB: Derived traits routines must be run after Reproduction or Survival data have been added or amended in order for these tables to accurately reflect what is held on the SIL database.

### Other uses

Across Flock Report Managers (AFRMs) can use the tools to check what data are held by flocks in their group when planning analyses or when assessing connectivity across years within a flock or across ewe age groups within a year for one flock.

These tables can be used to survey flock data characteristics to assess whether a particular analysis is appropriate for a breeding group e.g. what proportion of flocks collect particular data (and for how long), whether flocks use different ways to record lamb deaths or to look for reasons for apparent bias in an across flock analysis.

### Responsibilities

Breeders and their bureaus are responsible for data integrity in their flock. SIL DAT-RS tables provide a means to check data held on SIL.

### The DAT-RS tables

A detailed description of the DAT-RS tables and what they can be used for is given in Appendix 1. As well, at the end of this appendix are given suggested lists of tables for routine use by bureaus and by breeders.

Many users will find other ways to use these tables. They are encouraged to tell SIL of the uses they put the tables to so that SIL can revise tables if there is a strong need. Please email silhelp (see below) with your ideas.

### Need more information?

Contact your SIL bureau, send an email to [silhelp@sheepimprovement.co.nz](mailto:silhelp@sheepimprovement.co.nz) or telephone 0800-745-435 (0800-SIL-HELP).

## Appendix 1 – List of all DAT-RS tables

- 
- M1. Mating & Lambing Summary – ewe count by ewe age group
    - Are more ewes mated than lambbed?
      - If the difference is small, it suggests barren ewes and dead ewes were not recorded
      - High % of ewes not lambing – perhaps should be fated as dead, barren or “unknown”?
    - Are ewe hoggets regularly mated
    - Check of ewe age groups present and numbers in each group
    - Zero counts occur when there are ewes in an age group but they are coded as “Not mated”
    - Zero age rows occur when a ewe is identified as lambing in her own birth year – most likely an animal ID error
  - M2. Mating & Lambing Traits – ewe count by ewe age group
    - Shows which variable mating and lambing information may come from
    - Check of numbers for different types of data
  - M3. Dam Fates – ewe count by ewe age group
    - Which ewe fate codes (DFATE) are being used and how frequently
    - Look for too few ewes dead or barren
    - How much fostering? This can affect SIL’s ability to assess merit for Survival
- 
- L1. Birth Rank – lamb count by ewe age group
    - Check of lamb numbers by ewe age group and birth rank
  - L2. Sires Used
    - Check of sires used – ID and breed
    - Check of ewes mated per sire & sex ratio in progeny – NB: male, female and unknown sex
  - L3. Sires Used Across Years – lamb count
    - Sires connecting years – sires that connect years based on numbers of progeny born
    - Quick pointer as to whether link sires are used regularly to connect years to remove bias
    - NB: This is **not actual connectedness** since it is based on number of progeny from the pedigree and NOT the number of progeny with measurements for any trait
  - L4. Sires Used – lamb count by ewe age group
    - All sires used – showing which were used across ewe age groups within a year
  - L5. Sires Used Across Ewe Age Groups – lamb count
    - Only sires used across age groups within a year
    - Quick pointer that sires are used to connect ewe age groups to remove bias from the analysis
  - L6. Litter Size – ewe count by ewe age group
    - Check of litter sizes
    - Check for unexpected data – too few or too little in different groups
    - Same as format as L7 but counts here
  - L7. Litter Size – ewe percent by ewe age group
    - Check of litter sizes – compare fecundity of different ewe age groups
    - Check for unexpected data – too few or too little in different groups
    - Same as format as L6 but percentages here
  - L8. Birth Date - by ewe age group
    - Check of outliers – very early or very late birth dates (may give high “Std dev”)
    - If Std Dev = zero, all lambs born on same day
    - Other fields indicate whether problem is early or late
  - L9. Lambs Missing Parentage – lamb count
    - Check of pedigree data – should this many lambs be missing one or both parents?
-

- 
- S1. Birth Fates - lamb count by ewe age group
    - Which lamb fate codes (BFATE) are being used and how frequently
    - Are they being used consistently from year to year?
    - High survival % suggests lamb fates not complete
  - S2. Lamb Survival to Weaning - by ewe age group
    - Check of survival rates
    - Flags inconsistent fate data
      - Dead lambs with a fate code of died before weaning
      - Why would lambs alive at weaning not have a weaning weight?
  - S3. Lamb Survival to Weaning – birth rank by ewe age group
    - Shows average survival for different birth rank – ewe age combinations
    - Check to see if lamb fates have been consistently recorded for all groups
  - S4. Lamb Survival to Weaning - sex by ewe age group
    - Shows average survival for different sex – ewe age combinations
    - Check to see if lamb fates have been consistently recorded for both sexes
    - Some breeders record all dead lambs as Unknown sex. This will show as high survival for males & females
  - S5. Lamb Survival to Weaning - sire by birth year
    - List of sires used with raw measure of survival % in progeny
    - Useful for checking sire ID
    - Useful check of survival figures
  - S6. Birth & Rearing Rank Interaction – lamb count
    - Check of cross fostering
      - Are any lambs born AND raised as triplets?
      - Are any lambs reared in a larger litter than they were born into?
    - Proportion of lambs in different litter sizes for birth-rearing rank
    - Lambs without a RR are not counted
    - Table S7 is the same but it is based on ewe counts
  - S7. Birth & Rearing Rank Interaction – ewe count
    - Check of cross fostering
      - Are any lambs born AND raised as triplets?
      - Are any lambs reared in a larger litter than they were born into?
    - Proportion of ewes with different litter sizes for birth-rearing rank
    - Lambs without a RR are not counted
    - Table S6 is the same but it is based on lamb counts

---

Suggested set of DAT-RS tables for routine use by breeders - M1, M3, L3, L5, L7, S1, S3, S5

- To describe data held on SIL for the last three birth year cohorts of animals
- To check connectedness between years and between ewe age groups within year
- To check that the right amount of data is held on the SIL database
- To check for apparent abnormalities
- Default year range is from this year minus 5 to this year (e.g. 2002-2007)

---

Suggested set of DAT-RS tables for routine use by bureaus - M1, M2, M3, L2, L3, L5, (L7), L8, L9, S1, S2, (S3), (S5)

- To compare data added to SIL with that for the last few years
- To check connectedness between years and between ewe age groups within year
- To check that the right amount of data is held on the SIL database
- To check for apparent abnormalities
- Default year range is from this year minus 3 to this year (e.g. 2004-2007)
- *Bracketed tables are included so that the bureau default set of tables includes all tables in the breeder set*

<p>NB: Derived traits routines must be run after Reproduction or Survival data have been added to or amended on the SIL database in order that these tables accurately depict what should be used in genetic evaluations</p>
--