# Analysis of National Livestock Statistics: Assessment for Systematic Reporting Bias

S.M. Howden

*CSIRO Sustainable Ecosystems, GPO Box 284, Canberra, ACT, 2601 (mark.howden@cse.csiro.au)*

**Abstract:** A sound knowledge of the numbers of livestock in Australia is important information for input into industry planning, land management, disease management and for estimating greenhouse gas emissions. However, there have been persistent reports both in the literature and through less formal sources, that there may be a significant but largely unquantified level of under-reporting of livestock in the livestock census and surveys. This has implications for all of the above activities. In this paper an approach of 'inverse modelling' is developed where independent data of industry output (slaughterings, exports) and a simple herd budget approach is used with fitted population parameters (births, deaths) using a least squares metric to test the consistency of the livestock statistics. The analysis indicates there may be a consistent bias towards under-reporting for the beef cattle industry (or some errors in the reporting of calf numbers) but there is no need to assume a similar bias for the sheep industry. Possible approaches for further analyses are discussed briefly.

*Keywords:* Livestock; Statistics; Greenhouse gas emissions; Inventory

## 1. INTRODUCTION

Reliable and robust livestock statistics are needed for a range of purposes in Australia. These purposes include calculation of national greenhouse gas emissions [Howden and Reyenga, 1999], analysis of the risks of natural resource degradation [Brook, 1996], projections of industry productivity and disease surveillance and control. However, there have been suggestions that there may exist some form of systematic bias in the reporting of livestock by graziers [eg Young and Giles, 1982; Mortiss 1995] as happens elsewhere in relation to farmer-based reporting of disease [Goodger, 1978; Ogundipe et al., 1989].

The Australian Bureau of Statistics (ABS) collects livestock statistics through a combination of both a five-yearly farmer-response census (a complete sample of farms above a certain value) and annual surveys (a sub-sample). Sampling error is provided for the surveys and is typically of the order of a few percent, suggesting that the quality control procedures are highly effective. However, estimates of the accuracy of the census are not made, it being difficult to independently check the farmer responses. Anecdotal and other information suggests that farmers may be underestimating the number of animals on their properties [e.g. Mortiss, 1995; White, 1997], but to what degree is unknown. One of the few published studies which enables a comparison of ABS statistics against independent assessment of animals numbers indicates that ABS figures may be under-reporting by up to 60% in one region [Mortiss, 1995]. It is not possible to state whether such a disparity is common or unusual and, if it exists in other regions, whether it is of such magnitude. The aim of this paper is to attempt such an assessment at the national scale for both the cattle and sheep-based industries. The general approach is to use independently-sourced industry production data to parameterise a simple, national herd model. A national model is used as largely undocumented interstate transfers make a more disaggregated approach non-viable.

## 2. CATTLE INDUSTRIES

### 2.1 A Simple Accounting Approach

An initial analysis was applied to the national livestock statistics using an accounting approach. In its simplest form, such an accounting approach takes the form of:

$$\text{Change in stocks}_{2,1} = \text{Inputs} - \text{Outputs} \qquad (1)$$

Where the subscripts 1 and 2 refer to successive periods

To use such an approach, national livestock statistics for the beef and dairy cattle herds were drawn from the ABS livestock statistics and placed in a spreadsheet for the years 1986 to 1999. Data from earlier years were not used for two reasons: 1) 1986 marked a significant change in the size of farm for which the census was carried out, and 2) changes in the livestock industry could introduce significant trends if the time-span is extended too far. The basic livestock numbers were supplemented by output data from published ABS statistics on slaughterings and live cattle exports. These latter data on industry outputs are collected independently from the farmer-response census and surveys and provide the key data with which to assess the census and surveys. The data are associated with levy collection and a range of other reporting requirements and so were considered in the first instance to be robust.

The initial analysis assessed whether the documented outputs from the herd (slaughterings of cattle and calves plus exports) matched documented inputs (implied births calculated as beef and dairy calves less than 1 year old plus slaughterings of calves). Inspection showed that there was a consistent imbalance with herd outputs between 7.3 to 10.1 million (mean of 8.6M) whilst implied herd inputs were 6.2 to 7.8 million (mean of 7.0M). Given that deaths should also be included in the outputs component, there seemed to be a case for further investigation.

The further investigation used a slightly more sophisticated accounting approach by disaggregating the previous equation further:

$$\text{Total}_2 - \text{Total}_1 = \text{Births} - (\text{Slaughterings} + \text{Deaths} + \text{Exports}) \quad (2)$$

Where:
$\text{Total}_i$ = the combined beef and dairy numbers for all classes and the subscripts refer to successive years

Births can be represented in a couple of ways including through the direct use of the reported calf statistics, but in this case it was preferred to keep these numbers out of the equation so as to 1) provide a check on internal consistency of the analysis, and 2) reduce confounding as these statistics could also exhibit bias. Instead, births were calculated as a function of the numbers of breeding animals reported at the end of the previous reporting period. Given that the beef and dairy cattle industries have different breeding rates, two separate components were used. Hence the equation was expanded to become:

$$\text{Total}_2 - \text{Total}_1 = R_1{*}\text{Beef breeders} + R_2{*}\text{Dairy breeders} - (\text{Slaughterings} + \text{Deaths} + \text{Exports}) \quad (3)$$

Where:
$R_1$ = reproductive rate applying to all beef breeding cows and 75% of heifers from previous year)
$R_2$ = reproductive rate applying to all dairy cows
Total = sum of all beef and dairy cattle classes

This approach tends to be conservative as it assumes that all beef breeders are kept to calve before being slaughtered. Some heifers in the 1-2 year age class will not calve before the next survey. Given the approximately 9 month gestation period and assumed average age of first joining of 15 months this was assumed to be 25%.

Death rates are not reported in the national ABS statistics throughout the period. However, an indication of possible rates can be gained from ABS values for Queensland for the period 1975 to 1994 being 4.2%. This contrasts with documented rates of 1-2% [O'Rourke et al., 1995], 1.3% [Frisch, 1973]; 1.3% [Coates et al., 1987] for studies from research stations in productive northerly regions. However, in other regions with more severe climatic variation, death rates may be much higher in some seasons [e.g. Buxton and Stafford Smith 1996]. Death rates also tend to be higher on farmers properties than on research stations with suggestions of a mean of 6% or more across northern Australia [Mayer et al., 1999] although such rates would be expected to be lower in the southern regions of Australia. We assumed for this analysis a death rate of 4% of the total number of animals at the end of the previous reporting period, acknowledging that this may be conservative. Sensitivity analysis indicates that this assumption does not alter the thrust of the conclusions. Hence the equation is further expanded to:

$$\text{Total}_2 - \text{Total}_1 = R_1{*}\text{Beef breeders} + R_2{*}\text{Dairy breeders} - (\text{Slaughterings} + D{*}\text{Total}_1 + \text{Exports}) \ldots \quad (4)$$

Where:
$D$ = death rate applying to total herd (4%)

Theoretically, the above equation can then be solved (for example using the Excel 'Solver' function) for values of $R_1$, $R_2$ and $D$ using data from all years. However, before this is done, there is a need to include some additional parameters to assess whether there might be bias in reporting the size of beef and dairy herds. This is done by incorporating two new parameters into the equation ($B_1$ and $B_2$) for beef cattle and dairy cattle respectively. These parameters are multiplied against the respective herd numbers. The implementation here assumes that any bias in reporting will be systematic across animal classes within each industry group and thus equally affects both the breeders and total values. When applied to total values and death rates, there is a weighting for $B_1$ and $B_2$ based on the relative proportions of animals reported in the respective industry groups. The equation becomes:

$$(\text{Total}_2 - \text{Total}_1)_{B1,\,B2} = R_1{*}\text{Beef breeders}{*}B_1 + R_2{*}\text{Dairy breeders}{*}B_2 - (\text{Slaughterings} + D{*}\text{Total}_{1\,B1,\,B2} + \text{Exports}) \qquad (5)$$

Where:
$B_1$, $B_2$ = a weighting function based on relative proportions of beef and dairy cattle

The 'solver' approach needs to iterate parameters to achieve a specified target. In this instance it is the 'output' of the herds with the output being recorded slaughterings and live exports. These statistics were used as they are thought to be of low uncertainty arising from relatively strict record keeping due to, amongst other things, industry levies payable on the numbers of animals sold and requirements for export licenses. Hence, these statistics are not changed in the analysis and the other data are fitted to these. The standard method of analysis is to evaluate modelled outputs against measured outputs, this requiring a re-arrangement of the equation so that all unknown or assumed values are on the right-hand side to conform with the basic structure:

$$\text{Outputs} = \text{Inputs} - \text{Change in herd size}_{2,1} - \text{Deaths}$$

Hence:

$$\text{Slaughterings} + \text{Exports} = R_1{*}\text{Beef breeders}{*}\,B_1 + R_2{*}\text{Dairy breeders}{*}\,B_2 - (\text{Total}_2 - \text{Total}_1){*}_{B1,\,B2} - D{*}\text{Total}_{1\,B1,\,B2} \qquad (6)$$

A maximum likelihood statistic was computed from the measured and modelled values. This statistic allows fitting of parameters to minimise the least squares difference between the measured and modelled values (i.e. lower value indicates a smaller difference). The Excel 'Solver' was then applied to seek for parameter values that result in the best fit of modelled vs measured results. Unfortunately, there was a broad range of non-unique solutions possible for the relatively limited data set available.

However, this approach did show the sensitivity of the assumed reproductive rates to the fitting process. Hence, the above procedure was applied to investigate the sensitivity of assumed reproductive rates by fitting these rates when the herd size parameters $B_1$ and $B_2$ were set to 1 (i.e. no bias is assumed in reporting herd sizes). The results shown in Figure 1 indicate that there is a broad combination of parameters which minimise the least squares difference between the modelled and measured outputs. These range from a combination of 0.80 to 0.85 for dairy and 0.70 to 0.85 for beef cattle. Some of these values are unrealistic, for example, beef cattle reproductive rates are unlikely to approach a value of 0.85 and dairy unlikely to be as low as 0.70. When analysed, the ABS data show quite low reproductive rates. These average 0.69 and 0.77 for beef and dairy cattle respectively assuming that 90% of calves slaughtered were sourced from the dairy herd. These calculated reproductive rates are highly consistent across the recorded period with a standard deviation of 0.025 for beef cattle and 0.06 for dairy cattle. Whilst these values fall just below the band of likely values in the above analysis, if these reproductive rates are assumed then there is not only a marked and consistent lack of goodness of fit between the modelled and measured outputs (Figure 2) but also about 0.9 million more calves born than the ABS statistics would indicate.

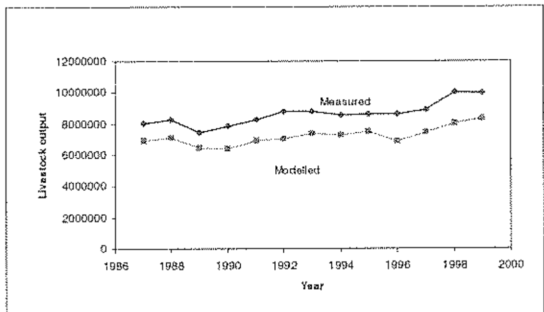


**Figure 1.** The effect of varying beef and dairy reproductive rates on the least squares difference between measured and modelled output. The herd sizes were constrained to recorded ABS numbers.

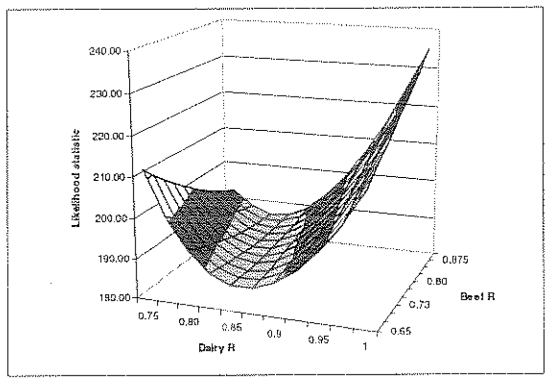


**Figure 2.** Measured and modelled herd outputs for 1987 to 1999 assuming reproductive rates as calculated from the ABS data and with herd sizes constrained to recorded ABS numbers.

However, the reproductive rates implied from the ABS statistics could be slightly lower than industry practice. For dairy cattle, limited consultation with farmers would suggest that they consider reproductive rates of 76 calves per 100 cows (i.e. a reproductive rate of 0.76) to be low. There is a trend across the period towards reductions in reproductive rates ($R_2$ = -0.0121*year + 24.819; $R^2$ = 0.67). This may be related to the increasing productivity per head experienced over the period but it is difficult to be definitive about this without more consultation and better data. Very recent trends towards extending the lactation period over the 12 month mark would also reduce the annual reproductive rate. Similarly, best practice in the beef cattle industry is likely to have higher reproductive rates than 0.66 although it is difficult to establish an independent industry average given the disparities between regions, seasons, age and condition classes, and between rates recorded on research stations compared with on-farm values. For northern Australia, a mean annual rate of 0.66 does not seem unreasonable for results from research stations [e.g. Pratchett et al., 1993] but this rate may be lower on-farm [Mayer et al., 1999]. ABS statistics for branding rates in Queensland from 1975 to 1990 give an average of 0.64. For southern Australia, there is no comparable study to that of Mayer et al. [1999] but reproductive rates are likely to be higher than in the north due to the better feeding conditions. Given that the national herd is approximately evenly split between the north and the south, the average industry reproductive rate may be slightly higher than the 0.69 used here but the disparity is likely to be limited.

Consequently, a sensitivity analysis was undertaken of the influence of different reproductive rates on the possible bias in herd sizes needed to minimise the least squares differences between measured and modelled output. Two reproductive rates scenarios were analysed: Low (beef $R$=0.69, dairy $R$=0.78) and High (0.75, 0.90). These combinations should bracket the likely array of rates experienced.

The Low scenario (Figure 3a) approximates the reproductive rates derived from ABS statistics and is consistent with, but perhaps marginally lower than, those which could be expected from national averages based on current industry practice. Under this scenario, the difference between measured and modelled outputs was minimised over a wide range of assumed reporting bias (e.g. 6-15% for beef and0-8% for dairy ) but with central values of 12% for beef and 3% for dairy cattle. However, these scenarios resulted in herd inputs (i.e. births) about 1.9 million higher than implied by the statistics.

The scenario with high reproductive rates (Figure 3b) provided a markedly different view, with the least squares difference minimized by assuming effectively no reporting biases. However, there are about 1.9 million more births calculated than is implied by the ABS statistics.

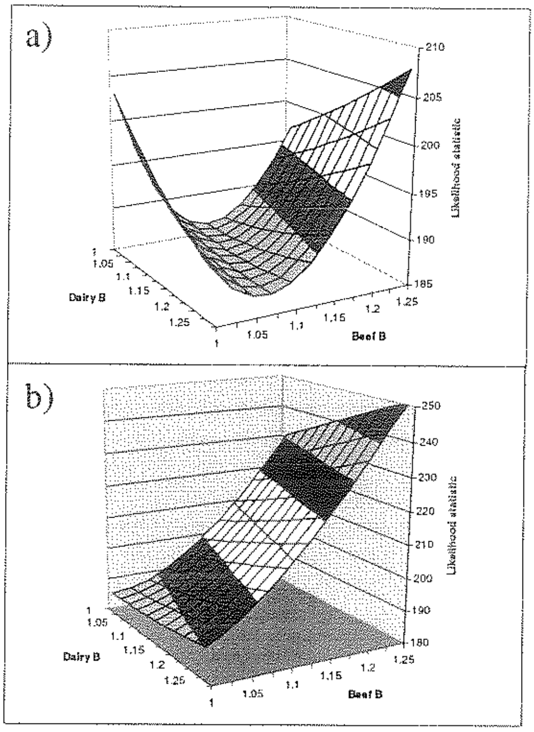


**Figure 3.** The effect of varying beef and dairy herd sizes on the least squares difference between measured and modelled output a) the 'Low' scenario with beef and dairy herd reproductive rates were constrained to values similar to those calculated from the ABS records b) the 'High' scenario with reproductive rates similar to what may be current industry 'best practice'.

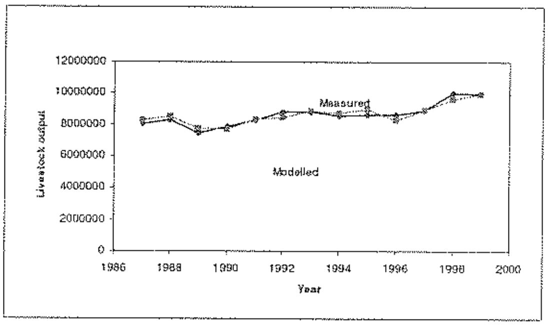


**Figure 4.** Measured and modelled herd outputs for the years 1987 to 1999 for mid-range reproductive rates (0.71, 0.84) and for biases 1.08 and 1.00 for beef and dairy herds respectively.

There is strong congruence between modelled and measured values of herd outputs (e.g. Figure 4; $R^2$ = 82%) for a large range of assumed values of reproductive rates and reporting bias.

## 3. SHEEP INDUSTRY

A similar approach to that used for cattle was adopted for the sheep industry although the final equation is more simple as there was only one industry to deal with:

$$(\text{Total}_2 - \text{Total}_1)*B = R*\text{Breeding ewes}*B - (\text{Slaughterings} + D*\text{Total}_1*B + \text{Exports}) \qquad (7)$$

Where:

$\text{Total}_i$ = total number of sheep and lambs and the subscript refers to the period
$R$ = reproductive rate applying to breeding ewes
$D$ = death rate applying to the whole flock
$B$ = factor to change the size of the flock
Breeding ewes = breeding ewes plus the maiden ewes intended for breeding reported at the end of the previous period.

This equation was re-organised as for cattle (Equation 6) to have the measured outputs (slaughterings and live exports) on the left-hand side and the modelled values on the right. A maximum likelihood statistic was calculated as for cattle, so as to enable evaluation of the sets of parameters that minimise the least squares difference between measured and modelled outputs.

In the first instance, an optimising solution was sought for Equation 7. As with beef cattle, there was a range of non-unique solutions although these were found over a much more limited parameter space and so may be used as a guide to appropriate settings. A general set of parameters found were 0.71 for reproductive rate applying to breeding ewes ($R$), 0.07 for the death rate applying to the whole flock ($D$) and 0.99 ($B$ factor to change the size of the flock).

The result of 0.99 for $B$ implies that no consistent change in flock size is needed to explain the observed outputs of the system, unlike the beef and dairy industries. As a consistency check, the values for reproductive rates (71%) are slightly higher than the average of 67.6% reported directly in the ABS statistics but are consistent with the calculated average value of 71.5 using the ABS data for lambs, lambs slaughtered and breeding ewes and ewes intended for breeding. The implied death rate of 7% is higher than that used for cattle but whilst there are no compiled national figures available it is not inconsistent with the high death rates that can occur with inclement weather which are well documented from various sources.

A further sensitivity analysis was made of reproductive rates and flock sizes following evaluation of the death rate parameter (a value of 0.07 was found to be suitable under a range of conditions). This analysis (not shown) suggested that within the range of reproductive rates both reported (67.6%) and implied (71.5%) by the ABS statistics, the least squares difference between measured and modelled outputs was minimised by a flock size very similar to that reported by the ABS.

## 4. DISCUSSION

The results of this analysis do not provide unique solutions to fitting the model, largely due to the limited data set used to fit the model parameters and thus definitive statements cannot be made. However, this analysis suggests that there is some discrepancy in the reports used to compile the ABS cattle statistics. If ABS statistics are used to calculate reproductive rates consistent with reported herd size and composition (and these seem to be only marginally below what appears to be happening with industry), then there is significant and consistent difference between measured and modelled outputs. Further, if these reproductive rates are applied in a simple herd model, then the measured outputs can be reconciled by assuming herd sizes greater than those reported (viz. the dairy herd 3% bigger than reported and beef herd 12% greater but with a range of possible combinations). However, such an analysis results in a calculation of about 1.9 million more calves than the statistics would indicate.

Alternatively, if greater birth rates are assumed then there is no need to assume any systematic errors in reporting total numbers. However, the reproductive rates are then consistently lower than those implied by the statistics again giving rise to more births calculated than the statistics would indicate.

Such results for cattle are not inconsistent with the suggestions raised by Young and Giles [1982], Mortiss [1995] and White [1997] for extensive livestock and informally for other livestock industries (e.g. piggeries: C. Brewster, NSW Agriculture). These levels of under-reporting could partly arise from the ABS census and surveys not recording animals on very small farms – but the effects of this should be limited as the ABS has assessed this factor. An alternative cause of the above differences could stem from the reliability of the slaughtering and export numbers. If these are inflated, then this will inflate the size of the herds needed to supply these activities (ie $B_1$ and

$B_2$ will be greater than unity). However, there does not seem to be an *a priori* reason for such inflation of figures given the levy arrangements noted previously.

There is also the possibility that this form of analysis does not account for differences in the data sources used. For example, the dates differ between the livestock census/survey (31 March) and the date at which slaughtering and export numbers are accumulated (30 June). This will add some discrepancies to the annual results. Similarly, a potential overlap between the census/survey date and the time of the end-of-wet-season muster in northern Australia may introduce some year to year variation in numbers. However, a herd budgeting approach such as this is an inherently conservative method of analysis as animals that miss one year tend to be accounted for in the following year or vice versa. Another source of error could be the assumed death rates. However, if death rates are higher, consistent with those suggested by the work of Mayer et al. [1999], the discrepancy increases; even if they are decreased to zero, there remains a significant imbalance. Given the current data sets available, it is not possible to be definitive about the reasons for the above discrepancies or bias. This is clearly an area of work that needs further attention given the importance of robust and reliable livestock statistics for a range of applications.

Interestingly, the statistics for sheep were internally consistent such that no reporting bias or variation in reproductive rates was needed to explain the flock outputs.

Independent checking of the livestock statistics is possible via aerial survey and dung sampling [Landsberg and Stol 1996]. However, these approaches are likely to be expensive. Alternatively, a more disaggregated analysis of the kind undertaken here may provide further insight as to sources of the possible discrepancies.

## 5. ACKNOWLEDGEMENTS

This report has been improved by the contributions and critical assessments of Drs David Mayer (QDPI), Greg Hood and Graham Turner (CSIRO Sustainable Ecosystems), Nigel Gibson (ABS), Beverley Henry, Greg McKeon, Wayne Hall and Grant Stone (QDNR), Roger Hegarty (UNE) and Penny Reyenga (Australian Greenhouse Office).

## 6. REFERENCES

Brook, K.D. Development of a national strategic drought alert system: Research Summary, Vol 1, Final Report on QPI20 to the Land and Water Resources Research and Development Corporation, Queensland Department of Primary Industries, Brisbane, 11pp., 1996.

Buxton, R. and D.M. Stafford Smith, Managing drought in Australia's rangelands: four weddings and a funeral, *Rangeland Journal*, 18, 292-308, 1996.

Coates, D.B., L. Mannetje and G.W. Seifert, Reproductive performance and calf growth to weaning of Hereford and Belmont Red cattle in subtropical, subcoastal Queensland, *Australian J. Experimental Agriculture.*, 27, 1-10, 1987.

Frisch, J.E., Comparative mortality rates of Bos indicus and Bos taurus cattle in central Queensland, *Australian J. Exp. Agriculture and Anim. Husbandry*, 13, 127-133, 1973.

Goodger, W.J., Economic losses associated with anaplasmosis in California beef cattle utilizing three methods of economic models, *California Veterinarian*, 32, 12-14, 1978.

Howden, S.M. and P.J. Reyenga, Methane emissions from Australian livestock: implications of the Kyoto Protocol, *Australian J. Agricultural Research*, 50, 1285-91, 1999.

Landsberg, J. and J. Stol, Spatial distribution of sheep, goats and kangaroos in woody rangeland paddocks, *Rangelands J.*, 18, 270-291, 1996.

Mayer, D.G., P.M. Pepper, G.M. McKeon and A.D. Moore, Modelling mortality and reproductive rates for management of beef herds in northern Australia. *Proceedings of the IASTED International Conference on Applied Modelling and Simulation*, September 1999, Cairns, Australia, 142-147, 1999.

Mortiss, P.D., The environmental issues of the Upper Burdekin catchment. Project report Q095017 to the Land and Water Resources Research and Development Corporation, Queensland Department of Primary Industries, Brisbane, 1995.

Ogundipe, G.A.T, S.B. Oluokun, and Esuruosmo, G.O., The development and efficiency of the animal health information system in Nigeria, *Preventative Veterinary Medicine*, 7, 121-135, 1989.

O'Rourke, P.K., G. Fordyce, R.G. Holroyd and I.D. Loxton, Mortality, wastage and lifetime productivity of Bos indicus cows under extensive grazing in northern Australia. I. Seasonal mating in the speargrass region, *Australian J. Experimental Agriculture*, 35, 285-295, 1995.

Pratchett, D., J. Wallace and S. Young, An evaluation of cattle types for north-western Australia, *Aust. J. Exp. Agric.*, 33, 677-683, 1993.

White, M.E., *Listen, our land is crying: Australia's environment: problems and solutions*, Kangaroo Press, Kenthurst, NSW, Australia, 1997.

Young, M.D. and G.E. Giles, The efficiency and equity of livestock rating systems and the accuracy of data obtained from them, *Aust. Rangeland Journal*, 4, 21-28, 1982.