Data quality of the Cattle Tracing System in Great Britain

D. M. Green, R. R. Kao

The Cattle Tracing System (CTS) of Great Britain was examined to investigate the quality of its data in terms of known errors and omissions, and their distribution. The proportions of erroneous or missing data have decreased steadily over time, with a marked improvement in the quality of the data since 2001, when recording became mandatory. There is little variation between regions in the quality of the data, but there are potentially important variations between the types of agricultural premises that cattle move to and from, and in other factors correlated with the type of premises, such as the age of the animals and the number of animal movements associated with the premises.

ANIMAL movements are a considerable risk factor for the transmission of various diseases, including bovine tuberculosis (Gilbert and others 2005) and foot-and-mouth disease (FMD) (Gibbens and others 2001, Kao 2002). Tracing the movements makes it possible to determine an animal's origin, which is important for BSE tracing (Stevenson and others 2005). The Cattle Tracing System (CTS), operated by the British Cattle Movement Service (BCMS) (BCMS 2005), was launched in September 1998, to satisfy European Union legislative requirements and to improve consumer confidence in an industry coping with BSE (Anon 2003). Recording of all cattle has been compulsory since 2001 (Fig 1), and cattle passports have been mandatory for newborn calves since 1996.

More recently, as part of the implementation of the Department for Environment, Food and Rural Affairs (DEFRA) UK Veterinary Surveillance Strategy, a project called RADAR (Rapid Analysis and Detection of Animal-related Risks) (DEFRA 2005a) has been commissioned to improve the collation and use of veterinary surveillance data, including the CTS. RADAR provides CTS data in the form of 'animal stays', a unique lifetime record of the location of each known animal. As with any data source, the CTS data are subject to errors and omissions (Mitchell and others 2005). If the errors are random, any conclusions based upon the data will usually be robust, although the errors will reduce their precision. However, any systematic bias may compromise the utility of the data more seriously. This paper examines the distribution of errors in the CTS animal histories data, to identify where movements are consistently unreported or reported incorrectly.

MATERIALS AND METHODS

Source data

CTS data from the introduction of cattle passports in late 1996 until April 30, 2005, were considered (Fig 1). Unique ID numbers identify each animal and location, and locations are also identified through CPH (county/parish/holding) codes, the location identifiers used in the June Agricultural Census (JAC) for 2003 (DEFRA 2005b), which provides estimates of cattle populations.

Each CTS movement record consists of a movement off one location and a corresponding movement on to another, linked only by the animal's ear-tag number. The CTS requires both sender and receiver to submit a movement report, and either or both may be missing or incorrect. These separate movement records are then paired to provide a record of the locations of each animal through time. When there are multiple intermediate locations within a day, determination of the animal's route through these locations can be difficult (Mitchell and others 2005). When the records are incomplete, that is, either the on- or off-movement record (or potentially both) is missing, or the two reports for a movement contain different information, the discrepancy is recorded in the data as explained below (movement type codes) and the best guess as to the movements of the animal is inferred automatically. When the data are poor for a particular animal, its movements may be impossible to retrieve.

Because the location data record animal locations, not movements, for the analyses below the data were reorganised into a new table describing each movement, by matching the off-movement of each stay with the on-movement of the subsequent stay. This table then contained ID codes for each animal, its source and destination locations, and a movement type code for each off- and on-movement. These type codes are important when considering the quality of the data: they indicate whether a movement was a completely reported movement, or was inferred by either the BCMS or RADAR. Many of the inferred moves were inferred when only one of the two movement reports for an animal movement was received by BCMS. Births and deaths are not considered in this analysis.

The resultant 'moves' table was analysed to provide summary statistics for each location, using the full dataset. The average distance travelled to and from, the number of other premises dealt with (the location 'degree') and the numbers of cattle moved were calculated for each location, for both the on- and off-movements. The average population size and the proportions of movement records that were 'normal' (defined as movement records for which no special inference code was supplied by either the BCMS or RADAR) were also included. Not all the statistics could be inferred for all locations; for example, without coordinates for a location, the distances travelled cannot be calculated. For 18 per cent of the premises active in the movements data, georeferencing (coordinates) data missing in the CTS were taken from the JAC, providing georeferences for all but 1 per cent of the active premises.

Data analysis

Data were extracted from the database by using C++ and Microsoft Access, and analysed using Minitab. The proportions of off- and on-movement records that were 'normal' (as defined above) were tabulated according to the off- and on-location statistics, including the location degree, the number of animals moved, and the average distance travelled to or from a location. Other movement categories, such as the region and county, type of location, and calendar month, were already present in the data. Movement records were also tabulated according to the animal's age (derived from birthdates provided in the 'livestock data' table), its sex, breed purpose, and the length of stay before or after a movement. There were 'unknown' categories for many of these analyses when categorical data were missing for particular movements, locations or animals.

Veterinary Record (2007) **161,** 439-443

D. M. Green, PhD,
R. R. Kao, PhD,
Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS

The descriptive statistics indicate for which types of movements, animals or locations there is a tendency for movements to be reported incorrectly. However, they do not allow causal factors to be identified, because there will be strong correlations between the explanatory variables. For example, there will be more movements through markets than through farms, and the locations that deal with a larger number of other locations will be largely those through which there are a large number of movements.

A general linear model (GLM) was used to unravel these correlated factors. The response variables used were the proportion of 'normal' movements (on or off) for each location for all the data, transformed into an empirical logit $L=\ln(^1/_2 + v)-\ln(^1/_2 + n - v)$, where v is the number of normal movements and n is the total number of movements for each location (Sokal and Rohlf 1995). This response variable makes it possible to ask what factors are correlated with, or are responsible for, underreporting or misreporting of movements. The continuous explanatory variables used, for example, the average location population and the average movement distance, were all positively skewed, and were transformed before performing the GLM by taking their logarithms. County code was included as an additional factor.

The JAC (DEFRA 2005b) provides population estimates for cattle for each location in the country, identified by CPH code. However, the JAC is not a complete census, and missing records are therefore imputed. The CTS data also provide population estimates for each location, with estimates being available for June 1, 2003, the day before the census. How these two population estimates matched was investigated by cross-referencing them through their CPH codes. The method of principal axes (Sokal and Rohlf 1995) was used to identify the relationship between the two population estimates, and was applied to the square-root-transformed data.

RESULTS

Descriptive statistics

Of all the movement data, 64 per cent were 'normal' movement records as defined above, without a movement type code supplied by the BCMS or RADAR indicating that they had been inferred in some way. Of the rest, 18 per cent had a movement type code supplied by RADAR for one end point, and 12 per cent had a code for both end points; the BCMS supplied a movement type code for one end point of 10 per cent of the records, and for both end points of 1 per cent of them.

Fig 2 shows the number of known cattle movements across time, which had a repeatable seasonal pattern at an approximately constant annual level since 2002 (see also Mitchell and others 2005). The disruption caused by the FMD epidemic in 2001 is clearly visible. The proportion of movement records that are 'normal' reported movement records has been consistent since 2002, whereas the majority of the movements before 1999 were inferred from other sources. Thus, the data from 2002 to 2004 can be taken as representative of the current activity of the CTS system. Overall, off-movements were reported more often than on-movements. The decrease in reported movements in the most recent data was due to delays in recording.

There was little regional variation in the proportion of normal movement records, as shown in Fig 3, although the figure does not include movements to or from slaughterhouses without CPH numbers; this lack accounts for the higher proportion of normal on-movement records than off-movement records (see below for discussion of types of premises). There was more variation at county level (not shown). However, as might be expected from Fig 3, there was no clear pattern, although differences at county level may have been due to differences in the nature of the local industry.

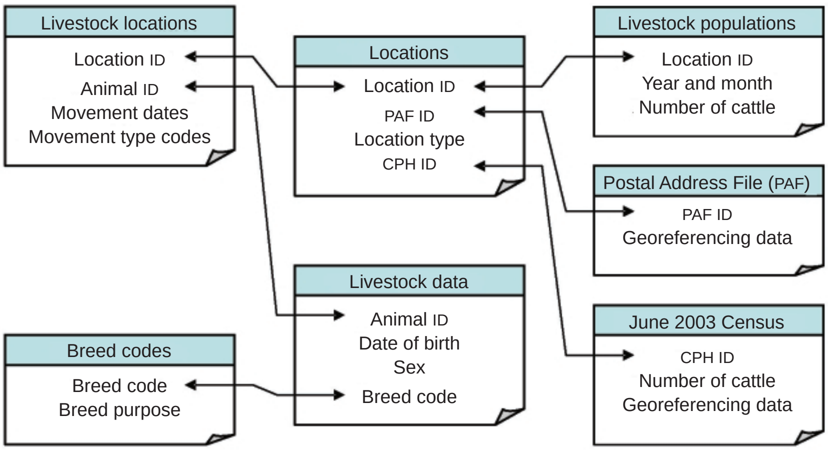

FIG 1: Data tables used for the analysis, showing the contents used and the relationships between them through the Cattle Tracing System Location ID, 'Animal ID', postal address file 'PAF ID', and county/parish/holding 'CPH' ID codes

Although there was no apparent regional bias, individual location- and animal-based predictors did show clear patterns. Table 1 shows the proportions of normal movement records, divided according to the type of premises, both for all the data, and for the data from 2002 to 2004. Market movements were well reported, close to 100 per cent for on- and off-movements, but this was not the case for showgrounds, with only just over 90 per cent of movements being reported. Since 2002, movements to and from animal holdings have been well reported; however, given the sheer volume of movements involving animal holdings, the number of misreported, inferred or unreported movements was large compared with some of the less well reported categories. Movements involving landless keepers had a lower proportion of normal movement records than animal holdings. For those location types that are predominantly sinks (that is, locations with relatively few off-movements), such as slaughterhouses, the proportion of normal movement records was low for all the data, but the proportion increased for the 2002 to 2004 data alone. On-movements to

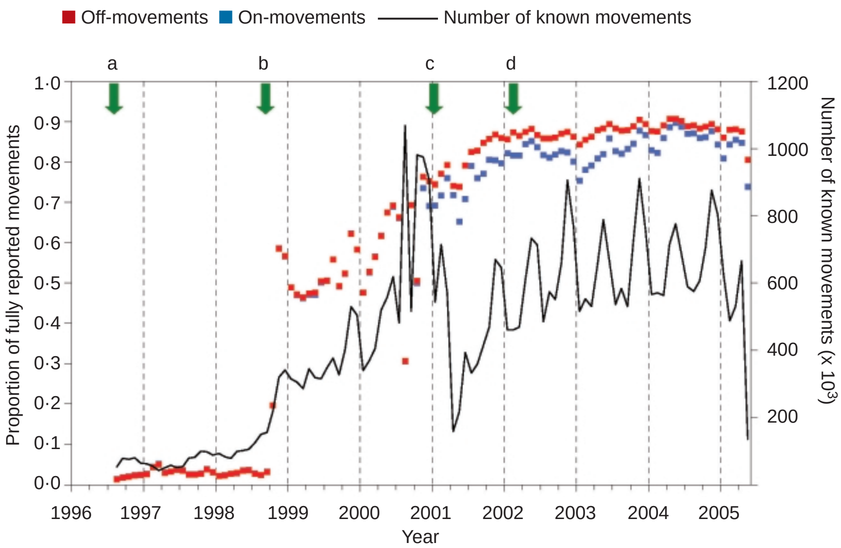

FIG 2: Proportions of off-movement and on-movement Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Scheme or RADAR) and the numbers of known movements between 1996 and 2005. Important dates in cattle tracing: a Newborn cattle require passports, b Recording compulsory for cattle with passports, c Compulsory for all cattle, d Introduction of standstill period

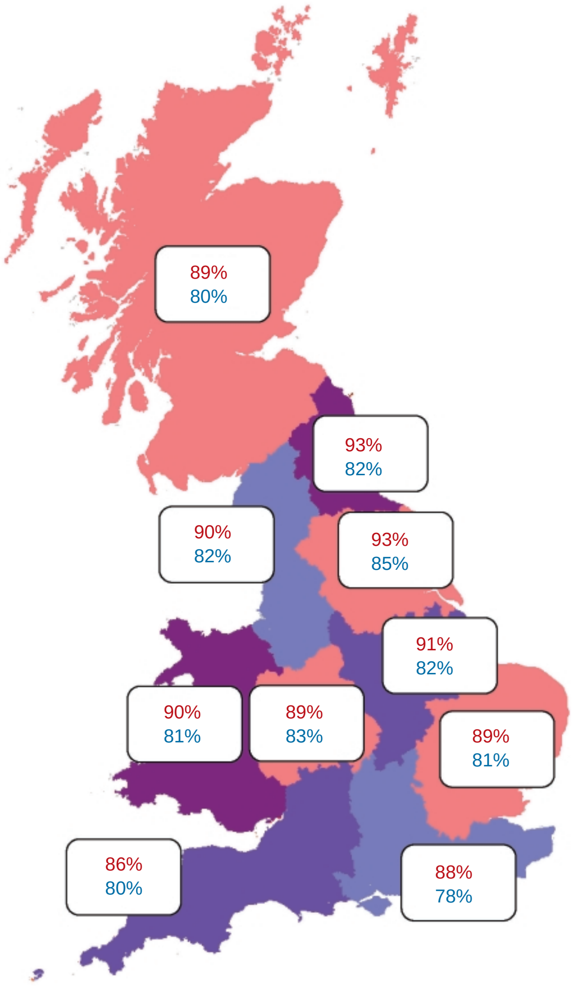

FIG 3: Proportions of off-movement (red) and on-movement (blue) Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Service or RADAR) in different regions of the UK, for all data with an associated county/parish/holding code; off-movements are categorised by the location of the source, on-movements by the location of the destination

slaughterhouses were not well reported, and because there are so many of them, they are partly responsible for the lower proportion of 'normal' on-movement records overall.

Table 2 shows a breakdown of the proportions of normal movement records according to the breed purpose (dairy, beef or dual), inferred from the breed, and sex of the cattle. The proportion was higher for beef cattle than dairy cattle, with dual-purpose cattle intermediate, and the proportion was higher for bulls than for cows.

In Fig 4, the movement records have been categorised according to the distance moved. There was a higher proportion of normal movement records for shorter-distance movements, and movements of unknown distance were less well reported. There were no clear trends when the movement records were categorised according to the average distance moved by cattle to or from the locations (not shown), except that there was a lower proportion of normal movement records from locations with unknown coordinates.

Location degree, that is, the numbers of other premises with which a premises exchanged cattle, and its numbers of on- or off-movements were closely correlated, with a Spearman's rank correlation coefficient between off-degree and off-movements for the 2002 to 2004 data of 0·73, and between on-degree and on-movements of 0·79. Thus, only the distribution of the proportion of normal movement records according to location degree is shown, in Fig 5. There was a slight trend towards a higher proportion of normal movement records for movements to or from locations with a higher degree, except for locations with the largest degree. These trends were due to the differences in degree between types of location; slaughterhouses and markets had a higher degree than agricultural holdings (Table 3), and slaughterhouses had a lower proportion of normal movement records than markets and agricultural holdings.

Fig 6 shows the movements categorised according to the age of the animals. The 'unknown' category consisted largely of older animals whose date of birth predated the system. There was a clear trend towards a higher proportion of normal movement records for younger cattle.

Table 4 shows the relationship with residence time at a location; residence times of less than a week but with an overnight stay had a lower proportion of on- and off-movement records at either end of the stay that were normal movements, but these movements accounted for less than 10 per cent of all the movement records. Single-day stays had a high proportion of normal movements because many were market movements, which were well reported.

Models of error distribution

The sample sizes provided by the CTS data were very large. As a result, the interpretation of statistical P values is uninformative because almost every factor appears to be significant although it may account for a vanishingly small proportion of the error.

The GLM models were applied to the logits of the proportions of 'normal' off- and on-movements, using county and premises type as discrete factors, and the premises degree, the distance animals moved (off or on as appropriate) and the number of cattle on the premises as covariates. The numbers of on- and off-movements were not included because they are necessarily correlated with the logits. Location type increased the r^2 statistics only slightly, because most locations were agricultural holdings, and was therefore omitted. All the terms were highly significant ($P<0·0005$). The model

TABLE 1: Proportions of off-movement and on-movement Cattle Tracing Scheme records that were fully reported (not inferred by the British Cattle Movement Service or RADAR) for each type of premises, and the total off- and on-movements for each category

	Data for 2002/04			All data		
Type of premises	Off (%)	On (%)	Total (x10⁶)	Off (%)	On (%)	Total (x10⁶)
Unknown	10·0	9·1	4·28	8·5	7·5	8·94
Agricultural holding	95·2	97·5	23·5	76·2	90·2	47·4
Landless keeper	91·5	95·8	0·39	88·4	94·2	0·66
Market	99·8	99·9	10·7	99·5	99·5	17·9
Showground	93·5	92·1	0·15	91·6	90·3	0·16
Slaughterhouse (red meat)	31·0	78·4	7·49	28·4	56·1	16·8
Slaughterhouse and cold store	1·9	99·8	0·03	2·7	13·9	0·17
Slaughterhouse (white meat)			–	69·6	95·6	<0·01
AI/AI subcentre	99·1	98·6	<0·01	97·9	81·9	<0·01
Calf collection centre	99·1	98·7	0·01	98·3	96·7	0·01
BSE collection centre	99·7	99·6	0·11	99·1	99·2	0·25
Hunt kennel	98·7	96·7	<0·01	75·6	4·0	0·03
Knacker's yard	98·4	89·5	<0·01	96·2	42·0	0·05

AI Artificial insemination

TABLE 2: Proportions of off- and on-movement Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Service or RADAR), with the numbers of movements according to breed purpose and sex of the cattle

	Number of movements ($\times10^6$)	Off-movements (%)	On-movements (%)
Breed type			
Beef	33·0	75·6	73·5
Dairy	11·5	70·7	64·5
Dual	0·3	73·1	70·4
Unknown	1·4	76·6	73·2
Sex			
Male	24·4	75·8	74·5
Female	21·9	72·8	67·7

accounted for 19 per cent of the variation in the logit score for off-movements, and for on-movements, 15 per cent.

For both models, county had the smallest *F* ratio and made the smallest contribution to the variance: its removal from the full model reduced the model r^2 to 18 per cent for off-movements and 12 per cent for on-movements. The distance moved had the next smallest *F* ratio in both models: its removal from the full model reduced r^2 to 16 per cent for off-movements and had little effect on the r^2 for on-movements. Of the remaining two covariates, premises degree made the larger contribution to the variance for both models: omitting it from the full model reduced r^2 to 12 per cent for off-movements and 3 per cent for on-movements. To conclude, analysis of variance showed that there was little bias according to geographical area, some according to the distance moved, and most according to the correlated factors of location degree and stocking number.

Population estimates

The agreement between the two estimates of total population size was excellent; the CTS data estimated that the cattle population was distributed over fewer locations, but with a counterbalancing higher mean population at each location. The population estimated by the JAC was 9,268,248 cattle on 92,440 premises, and that estimated by the CTS was 9,305,242 cattle in 87,827 locations, a difference of 0·4 per cent. The difference is more likely to have been underestimate in the imputed data of the JAC than an overestimate in the CTS, although neither would have been without error. For the locations that could be cross-referenced by their CPH codes, the JAC reported 8,367,009 cattle on 84,835 premises, approximately 90 per cent of the census population estimate; for the same locations in the CTS data, there were 9,022,013 cattle.

Where *C* is the square root of the census population estimate, and *R* the square root of the corresponding CTS data, then the equation of the principal axis was

$$R = -0{\cdot}0687 + 1{\cdot}05\ C$$

where the slope of 1·05 is close to, but significantly different from, unity (95 per cent confidence interval 1·042 to 1·054). A slope of 1·0 would be expected if the population estimates from the two data sources were identical.

DISCUSSION

The distribution of missing or incorrect movement records in the CTS data shows some variation according to season but little according to region. It was at the level of the type of enterprise where the largest differences were apparent. That animal movements were well reported for markets is encouraging, because markets are focal points for the spread of disease (Gibbens and others 2001, Green and others 2006, Kao and others 2006). Similarly reassuring is the high proportion of movement records associated with within-day stays that were normal records (Table 3, 2002 to 2004 data). Many of these movements would have involved markets, because for agricultural holdings there are statutory movement standstill periods and few within-day stays would be expected. Animals that are moved frequently clearly pose a bigger risk of transporting disease than animals that are moved infrequently.

Livestock dealers are also highly connected, may transport animals over long distances and have also been identified as important factors in the 2001 FMD epidemic (Gibbens and others 2001). Identifying the animal movements associated with dealers is not straightforward, because they are not identified as a separate type of premises in the census or CTS data, but are included in the 'agricultural holdings' class; however, they can be identified by their characteristics – they move large numbers of cattle to many locations, typically for long

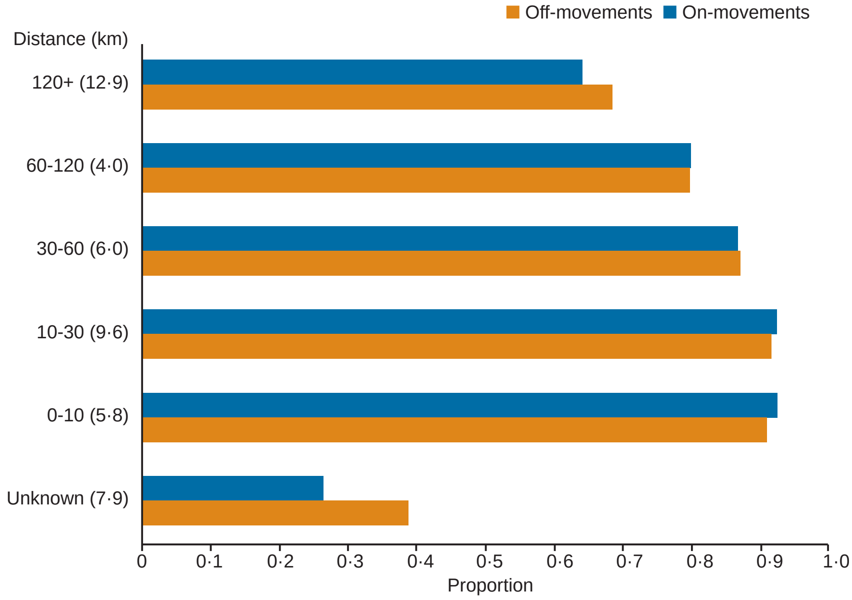

FIG 4: Proportions of off-movement and on-movement Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Service or RADAR) according to the distance moved by the animals (when known). The numbers of movements in each category (x 10^6) are given in parentheses on the y axis

distances, and animals are usually kept on their premises for short periods.

A higher proportion of the movements of beef cattle than dairy cattle had normal movements records. The existence of grants and subsidies for beef production (but not for dairy cattle) may have contributed to the more accurate reporting of the movements of beef cattle (DEFRA 2005c). The reduction in the proportion of normal movement records for older animals may similarly result from a lack of interest in older animals of less value, but it may also result from more practical factors, such as ear tags becoming less readable with time. Furthermore, the probability of movements being to slaugh-

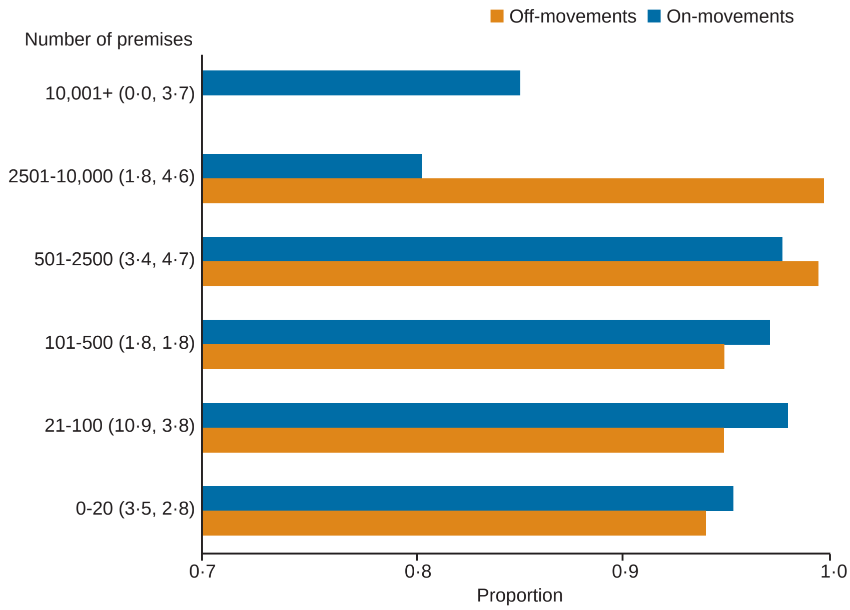

FIG 5: Proportions of off-movement and on-movement Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Service or RADAR) according to the number of other premises with which a premises exchanged livestock. The total numbers of off- and on-movements (x 10^6), respectively, are given in parentheses on the y axis

TABLE 3: Median numbers (interquartile ranges) of cattle movements, distances travelled to or from premises (km), and the numbers of other premises with which cattle were exchanged (degree) for selected classes of premises for the Cattle Tracing System

	Animal holdings	Markets	Slaughterhouses
Off-movements			
Number	78 (16-200)	6700 (28-33,000)	3 (1-50)
Distance	34 (22-53)	56 (39-82)	45 (16-78)
Degree	18 (6-37)	590 (120-1300)	3 (1-19)
On-movements			
Number	13 (2-63)	6700 (31-33,000)	570 (14-8600)
Distance	26 (14-49)	20 (15-28)	43 (26-67)
Degree	6 (2-13)	560 (150-1200)	520 (84-3000)

ter (movements that are less likely to be reported) increases with age.

The proportion of movement records that were normal decreased with the distance moved. A bias towards shorter distances in the estimate of the distance travelled by animals would lead to an underestimate of the potential long-distance transmission of disease to susceptible areas.

Throughout the data, the proportions of normal movement records were lower for the locations that act predominantly as sinks: the slaughterhouses. From the point of view of the spread of disease this poorer recording should be unimportant, because sinks do not contribute to the spread of disease through animal movements. However, when considering the transfer of cattle into the human food chain, it may be of importance; the risk of BSE is one concern (Anon 2003, Stevenson and others 2005). Similarly, any bias towards the movements of older animals being less well reported may be significant, because BSE infectivity is more widely distributed in older animals in more advanced stages of disease (Wells and others 1998).

Analysis carried out by RADAR has uncovered many cases in the CTS data where an animal moves off one location, after moving onto another location, with no evidence of movements in between. Consideration of solely the locations and the recorded movements between them will therefore miss potentially infectious links between these types of pairs of locations. There will be other omissions from the dataset than those inferred from other moves of the same cow. Identification of further omissions is a current subject of study, but this paper is concerned only with the errors and omissions apparent from the data presented. However, the omissions identified above may be surrogates for generally less well recorded data.

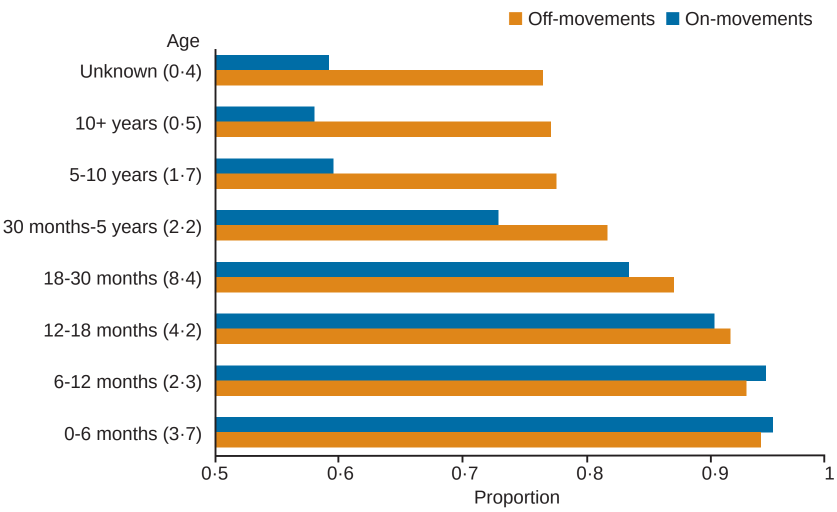

FIG 6: Proportions of off-movement and on-movement Cattle Tracing System records that were fully reported (not inferred by the British Cattle Movement Service or RADAR) according to the age of the cattle for the data for 2002 to 2004. The total numbers of movements (x 10^6) are given in parentheses on the y axis

TABLE 4: Proportions of movements that were fully reported (not inferred by the British Cattle Movement Service or RADAR), according to the time cattle spent on a holding, for the Cattle Tracing System data for 2002 to 2004

	Off-movements (%)	On-movements (%)
Single-day	97·1	85·3
One day to one week	26·3	52·8
Longer term	94·0	80·2

Further work could attempt to identify specific correlations between the different variables related to bias; for example, whether seasonal variation in the proportion of normal movement records reflects seasonal variation in the source and destination location types of cattle movements. Likewise, if animals move longer distances to slaughter than for other reasons, this might explain the fall in the proportion of normal movement records with distance.

ACKNOWLEDGEMENTS

The authors thank Steve Holdship, Andy Mitchell, Simon Roberts, Lisa Smith, Giles Paiba and John Wilesmith for valuable input. Fig 1 was inspired by Angela Wright. This work was funded by DEFRA and the Wellcome Trust. The CTS data extract used in the study was provided by the DEFRA RADAR system (www.defra.gov.uk/animalh/diseases/vetsurveillance/radar/index.htm).

References

ANON (2003) Identifying and tracing livestock in England. London, National Audit Office
BCMS (2005) CTS Online. www.bcms.gov.uk. Accessed February 16, 2006
DEFRA (2005a) Disease surveillance and control. www.defra.gov.uk/animalh/diseases/vetsurveillance/radar. Accessed February 16, 2006
DEFRA (2005b) Farming Statistics. www.defra.gov.uk/esg/work_htm/publications/cs/farmstats_web/default/htm. Accessed February 16, 2006
DEFRA (2005c) Farming – Cattle & Sheep Schemes. www.defra.gov.uk/farm/schemes/livstck.htm. Accessed February 16, 2006
GIBBENS, J. C., SHARPE, C. E., WILESMITH, J. W., MANSLEY, L. M., MICHALOPOULOU, E., RYAN, J. B. M. & HUDSON, M. (2001) Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Veterinary Record* **149,** 729-743
GILBERT, M., MITCHELL, A., BOURN, D., MAWDSLEY, J., CLIFTON-HADLEY, R. & WINT, W. (2005) Cattle movements and bovine tuberculosis in Great Britain. *Nature* **435,** 491-496
GREEN, D. M., KISS, I. Z. & KAO, R. R. (2006) Modelling the initial spread and foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B* **273,** 2729-2735
KAO, R. R. (2002) The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends in Microbiology* **10,** 279-286
KAO, R. R., DANON, L., GREEN, D. M. & KISS, I. Z. (2006) Pathogen dynamics on a well characterised network. *Proceedings of the Royal Society B* **273,** 1999-2007
MITCHELL, A., BOURN, D., MAWDSLEY, J., WINT, W., CLIFTON-HADLEY, R. & GILBERT, M. (2005) Characteristics of cattle movements in Britain – an analysis of records from the Cattle Tracing System. *Animal Science* **80,** 265-273
SOKAL, R. R. & ROHLF, F. J. (1995) Analysis of frequencies. In Biometry. 3rd edn. New York, Freeman. pp 685-793
STEVENSON, M. A., MORRIS, R. S., LAWSON, A. B., WILESMITH, J. W., RYAN, J. B. & JACKSON, R. (2005) Area-level risks for BSE in British cattle before and after the July 1988 meat and bone meal feed ban. *Preventive Veterinary Medicine* **69,** 129-144
WELLS, G. A., HAWKINS, S. A., GREEN, R. B., AUSTIN, A. R., DEXTER, I., SPENCER, Y. I., CHAPLIN, M. J., STACK, M. J. & DAWSON, M. (1998) Preliminary observations on the pathogenesis of experimental bovine spongiform encephalopathy (BSE): an update. *Veterinary Record* **142,** 103-106