



The Centre For Business Relationships,
Accountability, Sustainability and Society

WORKING PAPER SERIES No. 65

**The relationship between the gender of
veterinary surgeons and the results of
Bovine Tuberculosis tests : Implications
for methods of quality control**

Dr Gareth Enticott



About the BRASS Centre

The ESRC Centre for Business Relationships, Accountability, Sustainability and Society (BRASS) was the first ESRC Centre in Wales. It was established in 2001 with a £3.1 million grant, and in 2006 received its second phase funding of £4.8 million which will take it up to 2011. The Centre is a joint venture between the University's Schools of Business, City & Regional Planning and Law. It brings together the three Schools' existing research expertise on issues of sustainability, business ethics, company law, corporate reporting and business communication.

The Centre started work in October 2001 under the leadership of Professor Ken Peattie of the Business School, Professor Terry Marsden of the Department of City and Regional Planning and Professor Bob Lee of the Law School. The Centre exists to understand and promote the vital issues of sustainability, accountability and social responsibility, through research into key business relationships.

Published by

The Centre for Business Relationships, Accountability, Sustainability & Society
(BRASS)
Cardiff University
55 Park Place
Cardiff CF10 3AT
United Kingdom
<http://www.brass.cf.ac.uk>

© BRASS Centre 2011

ISBNs 978-1-906644-48-2 (print)
978-1-906644-49-9 (web)



The relationship between the gender of veterinary surgeons and the results of Bovine Tuberculosis tests : implications for methods of quality control.

Dr Gareth Enticott

Abstract

This paper analyses differences in results of tests for bovine Tuberculosis (bTB) between male and female vets. Recent concerns about the quality of bTB testing have prompted thinking about the role of quality control mechanisms such as performance management systems and performance indicators. To investigate their suitability, bTB testing data from areas with high incidence of bTB in the United Kingdom was extracted from the Vetnet database. Using the data, the performance indicator 'reactors per 1000 cattle tested' was calculated and compared with veterinarians' gender. Results showed statistically significant differences between gender and vets' reactor detection: male vets were more likely to find bTB reactors than female vets. Differences vary temporally and spatially and are particularly pronounced for short interval tests. Possible explanations for these variations are considered including how the concepts of 'emotional labour' and 'relational distance' may affect vets' behaviour. The presence of these systematic biases raises questions over the use of this performance indicator as a means of quality control, and the need for further research to examine these differences and other quality control methods.

Introduction

Bovine tuberculosis (bTB) is a serious infectious disease affecting cattle. In the United Kingdom (UK), approximately 36000 infected cattle are identified every year (Defra., 2010). The primary means of detecting bTB is by using a single intradermal comparative cervical test, also known as the 'skin test' (for details see: de la Rua-Domenech et al., 2006). Farms in areas where levels of bTB are high are tested at least on an annual basis. Where infected cattle are found, farms are prohibited from moving any livestock onto or off their farm until all cattle pass subsequent tests. The test has a strict protocol detailed in the Animal Health and Veterinary Laboratories Agency's (AHVLA) operations manual (Animal Health, 2010) and in EU directive 64/432/EEC on animal health problems affecting intra-Community trade in bovine animals and swine (Council of the European Economic Community, 1964). On completing a test, vets are required to certify that they have completed the test according to these rules.

A number of high profile cases have publicised concerns over substandard bTB testing practices amongst veterinary surgeons. In 2006, a farmer in Devon complained about the quality of a bTB test after a bull reacted to the test. When it emerged that the vet had not followed the correct procedure, a review into testing practices was commissioned by the UK Government which found that the testing protocol was routinely broken by practicing vets (DNV Consulting, 2006). Unease about the variability of the quality of the application of the skin test by vets was raised by the Independent Scientific Group analysing the effect of badger culling on bTB incidence (Independent Scientific Group, 2003). In 2011, a vet was struck off the register of veterinary surgeons for 10 months after failing to conduct a bTB test according to the test protocol (anon, 2011). Other research has highlighted the role of organisational culture in veterinary practices in creating and legitimising testing practices that diverge from the standard testing protocol (Enticott, 2011).

Similar worries about the variation of medical standards are present in public health research (McDonald et al., 2005, Waring, 2005, Waring et al., 2007). Whilst some of this research has stressed the value of intuitive knowledge held by health professionals (Currey et al., 2006, Currey and Botti, 2006), others have stressed the need to find ways to overcome cultural barriers that prevent health professionals from making errors or following standardised

guidelines and protocols (Dixon-Woods et al., 2009). One approach traditionally used in the public and private sectors has been to use a system of performance management and performance indicators to assess individuals' or organisations' outcomes, effectiveness and/or efficiency. The role of this paper, therefore, is to assess the extent to which performance management and performance indicators provide a suitable way of evaluating the performance of vets conducting bTB tests and means of improving the standards of bTB testing.

A performance management system is a collection of indicators that reflect an individual's or organisation's progress to specific aims and objectives, and which can be compared with other similar individuals or organisations (Audit Commission, 2000). More often than not, performance indicators (PIs) will be statistical, although they may also be qualitative or classified into groups (such as the labels 'excellent', 'good', 'fair' or 'poor'). In the last 20 years, performance management regimes have emerged as a common tool to help judge the effectiveness of public organisations, but they are also an important way of encouraging individuals to change and adapt their behaviour. The emergence of performance management is associated with what has been called the New Public Management (Hood, 1991), which is broadly the application of private sector management and accounting techniques to the public sector. In the UK, these methods have been used extensively within the public sector as part of the Labour Government's public service modernisation agenda (Boyne, 1998). Although this approach began in the 1980s, the use and number of PIs rapidly increased during the 1990s and 2000s as the Labour government sought to ensure minimum standards and demonstrate improved performance in health, education and other public services.

For some, these techniques help an organisation to focus on results rather than process and reward achievement rather than failure (Osborne and Gaebler, 1992). Thus, studies have recorded how systems of metrics, targets and performance management are associated with improved outcomes within the public sector (Boyne and Chen, 2008, Bevan and Hood, 2006). Others however have criticised the effect of PIs for eroding trust within the public sector and creating perverse organisational behaviour by encouraging gaming and fraud (Power, 1997, Mather, 2003, Public Administration Select Committee, 2003).

Whilst PIs may be used to report on organisations as a whole, they may also be used by individuals to improve their own performance. In effect, PIs act as a means of public surveillance: individuals can rate their performance against others and see how much they need to improve. At the same time, the publicness of these PIs creates coercive social pressures to improve performance. By making performance visible, adaptations in behaviour should follow as most people do not want to be seen to be different. However, as some studies have shown, there is no guarantee that individuals will seek to maximise their behaviour. Indeed, regression to the mean has many advantages as over achievement may bring other penalties, such as loss of social status and self-image amongst one's peers (Smith, 1993). Other studies have shown that comparisons of performance indicators between organisations are often false where they fail to take into account the effect of different environmental contexts (Andrews et al., 2006, Andrews, 2004).

When it comes to bTB testing, there are currently no PIs or management systems aimed at providing vets with increased knowledge of their or their colleagues performance in the UK. Given concerns over the quality of testing, a system of performance management may help to provide socially relevant coercive pressures to improve the standard of testing, by for example following the testing protocol more closely. Because of the dynamic nature of disease, PIs would not work as targets to aspire to. Instead, the comparisons they allow may create coercive pressures either through the knowledge that government is monitoring individual performance, or through the publication and sharing of performance indicators between vets in practice. If practice cultures and social relations are important in legitimising alternative testing practices, then PIs may offer a way of reversing this using the same social mechanisms.

Other countries, such as the Republic of Ireland, do use performance management system to monitor the performance of vets undertaking bTB tests. The system in Ireland employs a suite of PIs derived from testing data. Records for bTB testing in England and Wales do not allow for such a detailed analysis. However, using existing bTB testing data, it is possible to calculate some relevant PIs such as reactors per 1000 cattle tested. Whichever way these figures are calculated, it is important that veterinary practitioners recognise them as valid performance measures. This requires an absence of systematic biases between vets with, for example, different social characteristics. If they are, presenting figures for comparison could result in perverse changes in behaviour or wrongly identify 'failing' vets.

Materials and Methods

To test the validity of a performance management regime for veterinary services, the rest of this paper examines the potential of one such PI: Reactors per 1000 cattle tested (henceforth RpT). RpT was chosen as an exemplar indicator as it focuses on the main outcome of a bTB test – the number of animals testing positive or reacting to the test. Traditionally this figure has been used to express incidence of bTB in any given area, but given the subjectivities of the bTB skin test, it may also reflect vets' own testing practices.

The study used data extracted from the Animal Health and Veterinary Laboratories Agency's (AHVLA) Vetnet database. The aim of the analysis was to highlight any systematic differences that could compromise the ability to make comparisons between different vets' performance and render its behavioural change properties problematic. Results of all bTB tests were drawn from Vetnet for the years 2004-2009 in three counties: Devon, Dyfed and Gloucester. These counties were chosen because they have high levels of bTB incidence. In total, data were provided for 89567 bTB tests. Of these, 5667 tests were undertaken by vets directly employed by the Government. The remaining tests were undertaken by 1069 different vets in private practice. This paper only focuses on those tests conducted by private vets (see table 1 for descriptive statistics).

Data can be aggregated to county level and for individual vets. However, it is not possible to accurately aggregate data at a practice level. This is because when data are extracted from Vetnet, only the current practice is recorded against each vets test results. As many vets are likely to have changed practice over time, practice level aggregations will be inaccurate except for the most recent cases. Ideally, data would have been extracted for each practice in the three areas. Instead, data was extracted for all tests in each county (based on each farms' unique identifying code known as its CPH code¹). Inevitably this means that where practices conduct tests in neighbouring counties, those data will not be included.

¹ CPH stands for County Parish Holding. Each farm is allocated a CPH number which allows it to be geographically located within its County and Parish.

Table 1: Descriptive Statistics

Test Type	Gender	Number of cattle tested					Number of reactors				
		Mean	No. Tests	Minimum	Maximum	Sum	Mean	No. Reactors	Minimum	Maximum	Sum
Routine	male	95.07	38233	1	1394	3634903	0.43	38233	0	123	16278
	female	86.33	17061	1	1369	1472839	0.31	17061	0	44	5277
Short Interval	male	185.29	16378	1	1211	3034743	1.69	16378	0	215	27731
	female	167.05	6293	1	1309	1051228	1.16	6293	0	50	7274

The aim of the analysis was to look for systematic differences within the data, rather than compare data between counties. In fact, this would have merely reflected differences in disease incidence, rather than differences in practice. Vetnet records relatively few independent characteristics of vets conducting tests. Gender is recorded in the data, but age of vet or length of service is not. These could be calculated by reference to the Royal College of Veterinary Surgeons (RCVS) register of members, but anonymity and confidentiality constraints of the data-release agreement prevented this. The gender of vets working directly for AHVLA is not specified. Because of the limited social characteristic data in Vetnet, the gender of veterinary surgeons forms the basis of the analysis. It is hypothesised that there should be no difference between gender and the results of a bTB test.

Data was converted from a raw text file into a format that could be analysed using the SPSS statistical software package. RpT was calculated for each test by dividing the number of reactors by the number of cattle tested and multiplying by 1000. From these data, mean RpT by gender, test type, county and year was calculated. For there to be no relationship between gender and RpT we should expect there to be similar RpT for male and female vets. That is to say, RpT for male vets should not be consistently higher or lower across the 6 year period when compared to their female counterparts. Analysis was performed in SPSS using a series of independent t-tests and an Ordinary Least Squares (OLS) regression.

A significant limitation with the data was the absence of any indication of the interpretation at which the test was read. Conflating results read at severe interpretation with those at standard interpretation might affect the calculation of RpT. To accommodate these problems, separate RpTs have been calculated for routine tests and short interval tests. Although not all short interval tests will be conducted on severe interpretation, these tests are a different form of epidemiological investigation and deserve separate analysis. Moreover, other individual animal tests such as pre-movement tests and tracing tests have been excluded from the analysis.

Results

When data are pooled for all three counties, analysis reveals that male RpT is higher than female RpT for both routine and short interval tests ($p=0.000$) (see table 2). When these data are analysed by year a similar pattern emerges. For routine tests, analysis shows that the

differences between male and female RpT were statistically significant ($p < 0.05$) in three out of six years. For short interval tests, RpT differences in four out of six years were statistically significant. When data are analysed at a county level, similar patterns emerge. Firstly, it is important to point out that despite being areas of high bTB incidence, RpT values vary considerably between each county for both routine and short interval tests. For example, over the six years, routine RpT in Devon ranges between 4.8-5.8, and 3.5 to 8.9 in Gloucestershire.

Table 2: Differences in RpT for Males and Female Vets

Type of Test	Gender	N Tests	Mean RpT	P value
Routine tests	Male	38237	4.99	0.000
	Female	17061	3.80	
Short Interval tests	Male	16380	10.46	0.000
	Female	6294	7.97	

Table 3: Differences in RpT across county for male and female vets

Type of Test	County	Gender	N Tests	Mean	P value
Routine tests	Devon	male	17091	5.115	0.290
		female	8205	4.635	
	Glos	male	5901	8.316	0.284
		female	1397	6.843	
	Dyfed	male	15245	3.568	0.000
		female	7459	2.329	
Short interval tests	Devon	male	8788	9.695	0.009
		female	3826	7.804	
	Glos	male	3069	10.459	0.597
		female	588	11.260	
	Dyfed	male	4523	11.938	0.000
		female	1880	7.291	

Variations between RpT are also present within counties (see table 3). When data are pooled for all years, male and female RpT is significantly different ($p = 0.000$) for routine tests in Dyfed, but not for the other counties. For short interval tests, RpT differences are also

significantly different in Devon and Dyfed ($p < 0.01$). When data for routine tests are analysed separately by year and county, however, further differences can be found between male and female RpT (table 4). In Dyfed, the largest difference between male and female RpT is 2.68 with the mean 0.9. However, in 5 out of 6 years, male vets have higher RpT than females. These differences are statistically significant ($p < 0.5$) in two cases. Less variation is detected in Devon, with the average difference just 0.5 (maximum 1.8) and males higher on 4 occasions, although none of these differences are statistically significant. Results for Gloucestershire show the highest difference of 4.861 (in 2008) and on average difference of 0.86. Again, male vets have higher RpT in 4 of the 6 years, but only two years are statistically significant.

Overall, for routine tests, in 13 out of 18 cases (72%) male vets consistently found more reactors than female vets, but only four of these differences were statistically significant (22%). This pattern was even more evident in short interval tests. In Dyfed, for all 6 years, male vets have higher RpT than females with five of these differences being statistically significant. The average difference in RpT is 4.5 and the most it differs by is 6.3. In other words, on average male vets find 4 more reactors for every 1000 cattle tested than their female colleagues. These differences were less pronounced in the other areas. In Devon, the average RpT difference was 1.9 with the largest difference of 3.3. In 5 out of 6 years RpT was higher for male vets (3 years statistically significant). By contrast, in Gloucestershire RpT was highest for female vets for 4 out of 6 years with an average difference of 1.5 and a maximum of 7.9 (no years statistically significant).

It is possible that these associations are influenced by the nature of the herd being tested. For example, were male vets from private practices more likely to only test large herds, this may affect RpT? The data reveal that female vets do undertake fewer tests than male vets, although the relative share of tests has altered during the 6 year study period. In the three counties studied, female vets accounted for 33% of all bTB tests in 2009 compared with 22.8% in 2004. Female vets conduct more routine tests than short interval tests whilst the number of tests conducted by male vets in December and January is almost twice that of female vets. An Ordinary Least Squares (OLS) regression was used to check that the associations between gender and veterinary sector remained present when controlling for herd size and other factors. The dependent variable used was the number of reactors for each test. Independent variables included: number of cattle tested; and number of inconclusive

reactors. Dummy variables were used for gender, test type (routine or short interval), county and year of test. Results indicate that these independent variables explain 13% of the variation in the number of reactors (see table 5). Nine of eleven independent variables are significantly associated with reactor numbers. The strongest associations are between the number of inconclusive reactors and cattle tested. Reactors are more likely to be found in short interval tests. Gender is also positively associated with the number of reactors: the dummy variable indicates that tests by male vets are associated with the discovery of more reactors. Two of the dummy variables for year of the test are not significant, but overall these results suggest that the gender of the veterinary surgeon conducting the bTB test is associated with different levels of reactor diagnosis.

County	Test type	Gender	2004 RpT	2005 RpT	2006 RpT	2007 RpT	2008 RpT	2009 RpT
Dyfed	Routine test	male	3.369	3.835	2.290	3.150	6.364***	3.515*
		female	3.878	2.460	2.045	2.499	3.682	2.452
	Short interval	male	12.406*	10.078***	10.738	11.309**	15.072*	10.789***
		female	6.064	7.169	6.842	6.946	9.280	7.173
Devon	Routine test	male	4.292	5.723	3.320	4.003	5.295	3.888
		female	3.653	3.911	3.664	3.521	4.702	4.090
	Short interval	male	6.914	8.499***	5.857***	7.733***	9.639	8.149
		female	6.962	5.863	4.193	4.386	7.442	6.755
Glos	Routine test	male	4.348	10.216	5.035	6.353	8.235	6.711
		female	5.385	10.720	3.919	5.789	3.374	6.554
	Short interval	male	5.578	7.434	6.745*	7.953*	10.014	8.592
		female	4.240	10.196	8.663	4.600	17.917	9.678

Notes:

*** $p < 0.001$; ** $p < 0.05$; * $p < 0.1$

Table 5: OLS Regression. Dependent Variable: No. Reactors per Test. N - 77,965

	Standardized	t	Sig.	Collinearity Statistics	
	Coefficients Beta			Tolerance	VIF
(Constant)		-4.410	0.000	0.787	1.270
No. Inconclusive Reactors	0.173	45.924	0.000	0.843	1.186
No. Cattle tested	0.225	61.801	0.000	0.980	1.021
Gender (dummy variable)	0.019	5.626	0.000	0.873	1.146
Short interval tests (dummy variable)	0.072	20.227	0.000	0.431	2.318
County: Devon (dummy variable)	-0.030	-5.885	0.000	0.427	2.342
County: Dyfed (dummy variable)	-0.022	-4.261	0.000	0.537	1.862
Year: 2005 (dummy variable)	0.012	2.572	0.010	0.525	1.905
Year: 2006 (dummy variable)	-0.001	-0.325	0.745	0.511	1.956
Year: 2007 (dummy variable)	0.007	1.443	0.149	0.484	2.068
Year: 2008 (dummy variable)	0.033	6.825	0.000	0.465	2.148
Year: 2009 (dummy variable)	0.019	3.820	0.000	0.787	1.270

Adjusted R Square: 0.132
F – 1075.916
sig. 0.000

Discussion

Overall these data suggest that male vets find more reactors than female vets. Differences are present on an annual basis, and at a county level. There are fewer significant differences when data is analysed both spatially and temporally, but the evidence still points to a relationship between the outcome of a bTB test and the gender of the vet. In particular, results for short interval tests consistently reveal differences in RpT between male and female vets. These results suggest that the actual number of reactors may vary considerably simply as a result of the gender of the vet conducting the test. For example, if the male RpT had been applied to all tests in Dyfed, an additional 2260 cattle would have been classified as reactors between the years 2004 to 2009. On the other hand, if the female RpT had been applied to all cattle tests, 5575 fewer cattle would have been classified as reactors between the years 2004-09.

These results have a number of implications. Firstly, the presence of systematic biases in the data questions the use of RpT as a performance indicator and means of behavioural change. A key aim of any performance monitoring system is to allow comparisons between individual scores and the overall distribution of scores (e.g. the mean, or percentiles). However, this analysis suggests that unless a range of different versions of RpT is calculated, or a correction factor is factored into RpT, then the performance indicator will not be able to provide meaningful comparisons between vets. At least, separate RpT indicators should be calculated for vets for different types of test and county. However, the grouping of statistics into spatial units provides no guarantee of avoiding ecological fallacies. Vets' own knowledge of local epidemiological variations within counties may also contribute to the rejection of RpT as an accurate measure of performance by vets. Moreover, the need for a range of different versions of RpT will add to the cost of production and dissemination, as well as making comparisons less than straightforward and potentially out of date.

Secondly, these results raise the question of why such variations exist. *A priori*, there is no logical reason for variations in the data to exist, nor can male or female RpT be said to be any more or less "right". It is nevertheless important to explain why these differences exist. Practice level factors may provide one explanation. In response to these findings, vets have suggested that the disparity between male and female RpT is because practice partners are more likely to send young female vets to 'easier' tests. These may be tests with good handling facilities, friendly farmers and stock that are easier to manage. The classification of work and customers into easy/difficult groupings is an established coping mechanism amongst service workers (Mennerick, 1974) and also in veterinary practices (Sanders, 1994b, Sanders, 1994a). Studies of policing (Martin, 1999) also show how gender is associated with certain skills thereby affecting the kinds of police work allocated to female police officers. There is some evidence of gender differences in the distribution of bTB tests: the average herd size for routine bTB tests for female vets' is 86 cattle compared to 95 for male vets; and 167 compared to 185 for short interval tests. However, there should be no logical association between the perceived difficulty of tests, herd size and levels of bTB. Moreover, for this theory to hold, it must occur constantly in all practices, and relies on tests being allocated by people with the same foresight and knowledge of what constitutes 'easy' test. Whilst it is likely that there are some biases in the way tests are distributed in a practice, overall it seems

unlikely to have a significant impact upon RpT. Indeed, results from the OLS regression show that gender differences remain when controlling for herd size.

Another explanation put forward by vets in response to these findings is that female vets are likely to be younger and more likely to be influenced by farmers to not record reactors in borderline cases. Whilst this may be possible, such an explanation is also highly reductionist, treating the characteristics of gender as somehow pre-given. By contrast organisational research shows that difference results from the pressures and cultural norms of membership of social groups (Vaughan, 2005, Janis, 1972). Nevertheless, other research does find gender differences in the practices of medical doctors. In a study of breast cancer patients, Silliman et al. (1999) found that female surgeons provide better standard of care than their male counterparts; Glaesmer and Deter (2002) found that male doctors write prescriptions more frequently; Boulis and Long (2004) found that male and female surgeons have different referral rates for different illnesses; whilst Huston et al. (2001) revealed that male doctors of similar age and training are more likely to recommend hormone replacement therapy than female doctors see also (Beran et al., 2007, but see Coulter et al., 1995, Rathore et al., 2001).

In medical research, gender differences are often explained by different styles of communication between male and female doctors. In patient consultations, female doctors appear to pay more attention to the history of disease presentation than male doctors (Adams et al., 2008). A meta-analysis of gender effects in medical communication (Roter et al., 2002) found that female doctors communicate in a different way, using more positive and emotional talk and spend more time with their patients (see also Roter et al., 1991). For example, Roter et al. (1999) found that female doctors were more likely to engage emotionally with patients such as through laughter. These styles of behaviour have been famously referred to as 'emotional labour' (Hochschild, 1983) and whilst it has been associated with female jobs such nursing (Smith, 1992), its commercialisation throughout the service sector means its association with gender is disputed. Sanders (1994b) for example, shows how emotional labour is practiced by all members of a veterinary practice, regardless of gender. Similarly, research into different styles of environmental regulation (Marsden et al., 2000, Hutter, 1988, Lowe et al., 1997) also fails to find an association between gender and persuasive and emotional styles of regulation as opposed to more formal legalistic approaches. Indeed, in relation to bTB testing, emotional labour may manifest itself more in other activities after the bTB test – such as helping farmers complete paperwork and taking

time to explain how the farmer can manage the ir herd – rather than affect the way the disease is diagnosed.

An alternative explanation is that variations in RpT are due to the ‘relational distance’ (Black, 1976) between vets and farmers. Studies of regulation show that strict interpretation of laws and rules occurs when regulators are physically and culturally distant from the people and organisations they are regulating. Where there are similarities or previous working relationships, approaches to regulation are less formal and more negotiated (Hood et al., 1999, Grabosky and Braithwaite, 1986). As vets are allowed to test their own clients, relational distance is low meaning there is potential for negotiated decision making, particularly where vets have known farmers for many years. Accordingly, the theory of relational distance would suggest that high levels of RpT may be the result of longstanding close relationships with their clients which leads to classifying more cattle as reactors when they are only inconclusive reactors. For the management of the farm and the disease this practice may make sense to both vet and farmer, but as it constitutes false certification is likely to occur only when a relationship of trust exists between the vet and the farmer (see Enticott, 2011). The most recent analysis of the structure of the veterinary profession (Institute of Employment Studies, 2010) shows that the average age of male vets is 53 and 38 for female vets, potentially indicating lower relational distance amongst male vets and may help to explain to explain the higher levels of RpT amongst male vets. For the theory of relational distance to be confirmed further analysis on vets’ age and contacts with farmers is required.

More generally it is worth noting that low levels of relational distance have been encouraged by an increasing proportion of bTB tests being handed to the private sector in the three counties studied. In 2004, 9.9% of all herd tests were undertaken by vets working directly for the government. By 2009, this figure had almost halved, standing at 5.2% of all herd tests. Moreover, the share of different types of test has also changed. The majority of tests undertaken by government vets are short interval (88% in 2004, 78% in 2009) but their share of short interval tests has dramatically declined. In 2004 31% of all short interval tests were undertaken by government vets, but this had slumped to 6.6% in 2008 before rising again in 2009 to 11.1%. These data suggest a decreasing relational distance at a time when concerns about the quality of bTB testing by private vets have been increasing. If relational distance accounts for variations in RpT, then quality control efforts may work better by focussing on

the organisational structures of bTB testing in order to reduce the possibilities of conflict of interest.

Thirdly, the analysis raises questions about the way data from bTB tests are held by AHVLA. If concerns about the standard of bTB testing are real, then a data management system capable of easily producing analysis of the testing results for counties and individual vets is required. Limitations of this analysis include the inability to include the interpretation applied to test results. The database should also accurately collect information on vets' practice to allow analysis of variations in RpT between practices, as well as hold some limited personal data such as age or date of qualification. Clearly, ongoing analysis of these data is essential for a performance management regime designed to assess veterinary standards and assist their improvement.

Finally, even with up to date and accurate information, the meaning of these indicators needs to be checked with those vets who regularly test for bTB. What would they make of them? How frequently would they want to see them? How would they affect their behaviour? The best performance indicators are usually drawn up in conjunction with those people they are assessing. There can be little point in producing a set of performance indicators if those people believe they can do little about their performance, do not trust their meaning, or think they are powerless to do anything about their level of performance. For example, if vets do not connect good testing practices to a range of RpT values deemed to be normal, or if vets believe the number of reactors they find is down to chance, then the publication of RpT is unlikely to bring about any significant changes to bTB testing practices.

Conclusion

This paper has examined the potential role of performance indicators to help assess and improve the quality of tests for bTB by veterinary surgeons. The results show that female vets consistently find fewer reactors than their male counterparts. These systematic differences may vary temporally and spatially, but they do not appear to be related to factors associated with the test, such as herd size. There is no clear reason to explain these differences, but one possibility is the close relational distance between male vets and their clients may lead to more cattle being passed as reactors than inconclusive reactors. Although these results only relate to one possible PI, they suggest that using RpT as a performance indicator has considerable problems. In reality, a good performance management system

should use a range of different indicators to produce reliable judgements on performance. Nevertheless, this evidence suggests that considerable care needs to be taken if performance indicators are to be used to judge the standard of vets' bTB testing. Whilst these findings mirror similar research within human medicine, they also suggest that addressing some of the conflicts of interest that are perpetuated by the governance of bTB testing may prove a better strategy than developing a performance management system. Further research is required to unpack the meaning of the differences in RpT between male and female vets but the implications these findings have for the bTB testing regime are such that it should be a priority.

Acknowledgements: This research was funded by the Economic and Social Research Council (RES-000-22-2578 and RES-173-27-0213). The author is grateful to the AHVLA for providing the data and to officials in the Welsh Assembly Government and AHVLA (Carmarthen) for their assistance and comments on a previous version of this paper.

References

- ADAMS, A., BUCKINGHAM, C. D., LINDENMEYER, A., MCKINLAY, J. B., LINK, C., MARCEAU, L. & ARBER, S. 2008. The influence of patient and doctor gender on diagnosing coronary heart disease. *Sociology of Health and Illness*, 30, 1-18.
- ANDREWS, R. 2004. Analysing Deprivation and Local Authority Performance: The Implications for CPA. *Public Money and Management*, 24, 19-26.
- ANDREWS, R., BOYNE, G. A. & ENTICOTT, G. 2006. Performance failure in the public sector: Misfortune or mismanagement? *Public Management Review*, 8, 273-296.
- ANIMAL HEALTH 2010. *Operations Manual*, Worcester, Animal Health.
- ANON 2011. Vet to be suspended for false TB certification. *Veterinary Record*, 168, 315-316.
- AUDIT COMMISSION 2000. *On Target: The Practice of Performance Indicators*, London, Audit Commission.
- BERAN, M. S., CUNNINGHAM, W., LANDON, B. E., WILSON, I. B. & WONG, M. D. 2007. Clinician gender is more important than gender concordance in quality of HIV care. *Gender Medicine*, 4, 72-84.
- BEVAN, G. & HOOD, C. 2006. Have targets improved performance in the English NHS? *BMJ*, 332, 419-422.
- BLACK, D. 1976. *The Behaviour of Law*, New York, Academic Press.
- BOULIS, A. K. & LONG, J. A. 2004. Gender differences in the practice of adult primary care physicians. *Journal of Women's Health*, 13, 703-712.
- BOYNE, G. A. 1998. Public services under new labour: Back to bureaucracy? *Public Money and Management*, 18, 43-50.
- BOYNE, G. A. & CHEN, A. A. 2008. Performance targets and public service improvement. *Journal of Public Administration Research and Theory*, 17, 455-477.
- COULTER, A., PETO, V. & DOLL, H. 1995. Influence of sex of general practitioner on management of menorrhagia. *British Journal of General Practice*, 45, 471-475.
- COUNCIL OF THE EUROPEAN ECONOMIC COMMUNITY. 1964. *Council Directive 64/432/EEC of 26 June 1964 on animal health problems affecting intra-Community trade in bovine animals and swine (Consolidated version)* [Online].

Available:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1964L0432:20071113:EN:PDF>

[Accessed 27th April 2010.]

- CURREY, J. & BOTTI, M. 2006. The influence of patient complexity and nurses' experience on haemodynamic decision-making following cardiac surgery. *Intensive and Critical Care Nursing*, 22, 194-205.
- CURREY, J., BROWNE, J. & BOTTI, M. 2006. Haemodynamic instability after cardiac surgery: Nurses' perceptions of clinical decision-making. *Journal of Clinical Nursing*, 15, 1081-1090.
- DE LA RUA-DOMENECH, R., GOODCHILD, A. T., VORDERMEIER, H. M., HEWINSON, R. G., CHRISTIANSEN, K. H. & CLIFTON-HADLEY, R. S. 2006. Ante mortem diagnosis of tuberculosis in cattle: A review of the tuberculin tests, [gamma]-interferon assay and other ancillary diagnostic techniques. *Research in Veterinary Science*, 81, 190-210.
- DEFRA. 2010. *Bovine Tuberculosis: The Government's approach to tackling the disease and consultation on a badger control policy*, London, Defra.
- DIXON-WOODS, M., SUOKAS, A., PITCHFORTH, E. & TARRANT, C. 2009. An ethnographic study of classifying and accounting for risk at the sharp end of medical wards. *Social Science and Medicine*, 69, 362-369.
- DNV CONSULTING 2006. *Review of TB testing procedures*, London, Defra and the Welsh Assembly Government.
- ENTICOTT, G. 2011. The Local Universality of Veterinary Expertise and the Geography of Animal Disease. *Transactions of the Institute of British Geographers*.
- GLAESMER, H. & DETER, H. C. 2002. Gender-specific aspects of physicians work: Is there a difference in drug prescribing behaviour of male and female physicians. *Geschlechtsspezifische aspekte der ärztlichen tätigkeit: Verschreiben ärzte häufiger medikamente als ärztinnen?*, 52, 386-391.
- GRABOSKY, P. & BRAITHWAITE, J. 1986. *Of Manners Gentle. Enforcement Strategies of Australian Business Regulatory Agencies*, Oxford, Oxford University Press.
- HOCHSCHILD, A. R. 1983. *The Managed Heart*, London, University of California Press.

- HOOD, C. 1991. A New Public Management for all seasons. *Public Administration*, 69, 3-19.
- HOOD, C., SCOTT, C., JAMES, O., JONES, G. & TRAVERS, T. 1999. *Regulation inside government waste watchers, quality police, and sleaze-busters*, New York, Oxford University Press.
- HUSTON, S., SLEATH, B. & RUBIN, R. H. 2001. Physician gender and hormone replacement therapy discussion. *Journal of Women's Health and Gender-Based Medicine*, 10, 279-287.
- HUTTER, B. M. 1988. *The reasonable arm of the law? the law enforcement procedures of environmental health officers*, Oxford, Clarendon.
- INDEPENDENT SCIENTIFIC GROUP 2003. *Minutes*, London, Defra.
- INSTITUTE OF EMPLOYMENT STUDIES 2010. *The 2010 RCVS Survey of the UK Veterinary and Veterinary Nursing Professions*, Brighton, IES.
- JANIS, I. 1972. *Victims of Groupthink*, Boston, Houghton Mifflin.
- LOWE, P., CLARK, J., SEYMOUR, S. & WARD, N. 1997. *Moralizing the environment countryside change, farming and pollution*, London, UCL Press.
- MARSDEN, T., FLYNN, A. & HARRISON, M. 2000. *Consuming interests the social provision of food choice*, London, UCL.
- MARTIN, S. E. 1999. Police Force or Police Service? Gender and Emotional Labor. *The ANNALS of the American Academy of Political and Social Science*, 561, 111-126.
- MATHER, G. 2003. Beyond targets, towards choice. *Political Quarterly*, 74, 481-492.
- MCDONALD, R., WARING, J., HARRISON, S., WALSHE, K. & BOADEN, R. 2005. Rules and guidelines in clinical practice: A qualitative study in operating theatres of doctors' and nurses' views. *Quality and Safety in Health Care*, 14, 290-294.
- MENNERICK, L. 1974. Client Typologies: A Method of Coping with Conflict in the Service Worker-Client Relationship. *Sociology of Work and Occupations*, 396-418.
- OSBORNE, D. & GAEBLER, T. 1992. *Reinventing Government*, Reading, MA., Addison-Wesley Publ. Co.
- POWER, M. 1997. *The Audit Society*, Oxford, Oxford University Press.
- PUBLIC ADMINISTRATION SELECT COMMITTEE 2003. *On Target? Government by Measurement*, London, The Stationery Office.

- RATHORE, S. S., CHEN, J., WANG, Y., RADFORD, M. J., VACCARINO, V. & KRUMHOLZ, H. M. 2001. Sex differences in cardiac catheterization: The role of physician gender. *Journal of the American Medical Association*, 286, 2849-2856.
- ROTER, D., LIPKIN JR, M. & KORSGAARD, A. 1991. Sex differences in patients' and physicians' communication during primary care medical visits. *Medical Care*, 29, 1083-1093.
- ROTER, D. L., GELLER, G., BERNHARDT, B. A., LARSON, S. M. & DOKSUM, T. 1999. Effects of obstetrician gender on communication and patient satisfaction. *Obstetrics and Gynecology*, 93, 635-641.
- ROTER, D. L., HALL, J. A. & AOKI, Y. 2002. Physician gender effects in medical communication: A meta-analytic review. *Journal of the American Medical Association*, 288, 756-764.
- SANDERS, C. 1994a. Biting the hand that heals you: encounters with problematic patients in a general veterinary practice. *Society and Animals*, 2, 47-66.
- SANDERS, C. R. 1994b. Annoying owners: Routine interactions with problematic clients in a general veterinary practice. *Qualitative Sociology*, 17, 159-170.
- SILLIMAN, R. A., DEMISSIE, S. & TROYAN, S. L. 1999. The care of older women with early-stage breast cancer: What is the role of surgeon gender? *Medical Care*, 37, 1057-1067.
- SMITH, P. 1992. *The Emotional Labour of Nursing*, Basingstoke, Macmillan.
- SMITH, P. 1993. Outcome-related performance indicators and organisational control in the public sector. *British Journal of Management*, 4, 135-151.
- VAUGHAN, D. 2005. The normalization of deviance: signals of danger, situated action, and risk. In: MONTGOMERY, H., LIPSHITZ, L. & BREHMER, B. (eds.) *How professionals make decisions*. London: Lawrence Erlbaum.
- WARING, J., HARRISON, S. & MCDONALD, R. 2007. A culture of safety or coping? Ritualistic behaviours in the operating theatre. *Journal of Health Services Research and Policy*, 12.
- WARING, J. J. 2005. Beyond blame: Cultural barriers to medical incident reporting. *Social Science and Medicine*, 60, 1927-1935.

