

HIP SCORES OF NEWFOUNDLAND DOGS

An analysis by

John Morgan BSc

September 2010

Version 1.0



HIP SCORES OF NEWFOUNDLAND DOGS

An analysis by

John Morgan BSc

SUMMARY

This paper is a mathematical analysis of the hip scores recorded for Newfoundland dogs.

The following sections form a summary of the essential facts determined in the main part of the paper, using the same section numbers and headings, where further details and explanations can be found.

0.1 Introduction

Canine Hip Dysplasia (CHD) is defined as an abnormal formation of the hip socket that, in its more severe form, can eventually cause crippling lameness, painful arthritis of the joints, or Degenerative Joint Disease (DJD). In order to reduce the incidence of the condition various recommendations have been made on breeding practice. Those recommendations are not based on the appearance of clinical signs (distress, lameness) but on scoring schemes based on the evaluation of radiographs (X-rays) made at a young age when the animal shows no clinical conditions.

In the scheme administered jointly by the British Veterinary Association and the UK Kennel Club dogs are radiographed after their first birthday and the radiographs sent to the BVA for evaluation. This gives a score for nine different aspects of each hip. These scores are added together to give total on a scale 0-53 for each hip. This study is concerned only with the total hip score for two hips, which can be any number in the range from 0-106.

In the two schemes operated in the USA, the Orthopaedic Foundation for Animals uses a similar scheme to that in the UK but lists results in 7 categories, three of which are classed as "normal" and three as "dysplastic", with the intermediate group classed as borderline. In spite of its literal meaning, which essential just means abnormal, the author dislikes the term "dysplastic" as it seems to give a false impression of disease. The Pennhip foundation in the USA uses a single number to describe hip scores, with higher values associated with the term Degenerative Joint Disease, which also seems not to give the correct impression as it appears to be making a clinical judgement. Both of these terms are avoided in this paper, except when making references to other sources.

In this paper only two terms are used to describe hip scores, "Normal" and "CHD". These terms are used to distinguish two groups which seem to have distinguishing characteristics beyond the simple fact of higher hip scores. These characteristics are defined both in terms relevant to genetic considerations and the variation in progeny scores associated with these two groups.

0.2 Newfoundland Hip Scores

Hip Scores for 3,222 Newfoundland dogs scored in the period 1 September 1991 to July 2010 were obtained from the Kennel Club and form the basis for this study.

Previous studies have found that progeny with both parents scored have generally lower hip scores than those with neither parent scored. To ensure data consistency the data were divided into three groups; progeny with both parents scored, those with just one parent scored and those with neither parent scored:

	Number	Mean
Both parents scored	1996	22.9
One parent scored	574	23.5
Neither parent scored	652	28.1

When neither parent has been scored the two parents are representative of the large number of Newfoundlands (about 80%) that have been registered with the KC but not scored. The high score for this group is therefore indicative of the situation in the wider population of Newfoundlands.

By contrast, the group with both parents scored is the result of attempts to reduce hip scores through at least one generation. The reduction from about 28 to about 23 is therefore an indication of the success of the hip scoring scheme.

The hip scores for Newfoundland dogs is relatively high, with an official BMS of 26, based on the scores of 4,000 dogs, This places them as number 140 in a list of 147 breeds on the BVA web site. Similarly, Newfoundland dogs rank as number 136 of 157 breeds in the OFA website in the USA, this time based on the proportion of high hip scores.

It will be shown later that the reason for the high BMS is not that Newfies have generally worse hips than other breeds, as all breeds have a core of animals with normal hips having similar characteristics. The reason for the high BMS is the larger number of animals with high hip scores representative of CHD.

0.3 Predictability of individual scores

It is a common experience, supported by widespread anecdotal evidence, that hip scores can at times appear to be completely random, with low scoring parents having occasional progeny with very high scores, or the reverse. This apparent chaotic behaviour is clearly seen on scatter plots comparing the scores of individual animals with their sire or dam. The plots at first sight appear to be completely random, with examples of progeny having a score of 100 when the dam has a score of 9, or a score of 8 with a dam score of 93. The same chaotic behaviour occurs for the sire-progeny relationship and in other breeds.

In mathematical terms this demonstrates that it is impossible to predict the score of individual progeny from the score of its sire or dam. It is necessary therefore to study the average impact of high or low parent scores in a statistical way, looking for probabilities rather than specific individual forecasts.

04 Hip Score distributions

Charts of the distribution of hip scores show a sharp peak in the number of evaluations when the scores are less than about 25, then a long tail of evaluations for high hip scores, extending over the whole range up to the maximum possible score of 106. There is strong evidence that the distribution of all breeds is similar, demonstrating clearly that this is caused by genetic factors rather than through trauma or other environmental conditions.

As the peak in the distribution is so pronounced it is remarkably similar to a standard statistical distribution known as the Normal Distribution. This and other considerations suggest that hip scores that lie within the curve represented by a standard Normal curve are in fact the hip scores of “normal” dogs. This implies that normal dogs may have scores in the range up to about 21. It might be thought that hip scores higher than this would be considered to be abnormal, or in other words represent CHD. This explanation is too simple. If the scores represented by the “Normal” curve are subtracted from the remaining scores they form another distribution that overlaps the first, with scores starting at 14.

This process establishes three zones for the Newfoundland scores:

- Dogs with scores less than 14 are considered to be normal, subject to natural variations within this range
- Dogs with scores above 21 have scores relating to increasing degrees of CHD
- The range 14 to 21 is borderline, with some normal dogs but with an increasing proportion of dogs with CHD as the scores increase in value.

This does not mean that the dogs in these groups have different genetic makeup. It has already been demonstrated that a low scoring parent can have high scoring progeny. It will be demonstrated in a later section that pairs of low scoring dogs can also have high scoring progeny. The relevant genetic factors are present in all dogs and are not isolated in particular groups.

Another factor that emerges from this study is that the peak of the distribution is similar in all breeds. This implies that the hip scores of dogs in the “normal” range are similar in all breeds; a “normal” Newfoundland is the same as a “normal” Labrador for example. The difference between the breeds is mainly a difference between the proportions of animals that can be associated with CHD. A higher proportion of Newfoundlands than Labradors have CHD and this is the factor that makes the BMS higher in this breed.

0.5 Effect of parent scores

The results discussed in the previous section may seem abstract and theoretical, but they are confirmed in the real world when the scores of progeny are compared for parents in the three groups. These progeny scores show significant differences across the range; in the proportion of very high scores, the proportion of low scores and the average or mean value of the progeny.

With both parents having scores less than 14, more than half of their progeny had hip scores lower than 11, while one in five had high scores, greater than 25.

When both parents had scores higher than 21 the situation is almost the opposite, with nearly half of the progeny scoring more than 25 and only one in seven having a score lower than 11.

Of course, the borderline group with scores in the range 14 to 21 are intermediate between these two extremes, with about one in three having high scores (above 25) and one in three having low scores (below 11).

Clearly these figures suggest that breeding from parents that both have high scores is to be discouraged. However, in some cases the availability of suitable breeding stock is limited, or an animal with a high hip score has other characteristics which are very desirable for the breed. In such cases, matching a high scoring parent with a low scoring parent gives the same proportion of high and low scoring progeny as the borderline group.

The mean scores of the progeny follow the same pattern with progeny scores averaging around 18 for parents in the low scoring group, 33 for the high scoring parents and just 25 for the intermediate borderline group. As before, choosing one parent with a high score, one with a low score produces progeny with the same average score as the borderline group.

These real-life statistics prove the point that it is very difficult to reduce the Breed Mean Score, as this is based on the average of more than 4,000 Newfoundland scores dating back to 1980. With only about 150 Newfoundland evaluations each year the impact of low scores from this small group would be totally overwhelmed by the 4,000 that already contribute to the BMS.

0.6 The Triple-Gene Theory

A brief excursion into basic genetic theory confirms that all of the characteristics of Newfoundland scores can be related to a theory that assumes that hip scores are based on the interactions of three gene pairs, where one gene in each pair is inherited from each parent.

This theory explains why there is such a diversity in the observed hip scores and confirms that the dog population can be divided into two main groups; one without obvious signs of CHD and the other certain to show the effects of CHD, with intermediate scores in between. It also confirms the distributions found in Section 4 and the clear separation of parental groups in Section 5. It also correctly predicts that low scoring parents will still produce high scoring progeny and that high scoring parents can also produce a small proportion of low scoring progeny.

More importantly, the genetic theory indicates that successive generations can alternate between being normal or with CHD, confirming that the condition known as CHD is an entirely natural phenomenon, not some strange disease or virus which can be caught!

0.7 Trends in Newfoundland Hip Scores

When the annual mean hip scores are plotted there is an apparent decline year-by-year, with the trend line suggesting that the average score has decreased by 7 points in the period from 1991 to 2010.

However, doubt is cast on this apparent decline. The reason for this becomes obvious when the trend is plotted for those animals whose parents were not scored. This also shows a decline, which is difficult to explain. These animals have not been subject to a breeding policy based on hip scores as nobody knows the hip scores of their parents. This group of progeny are representative of the vast majority of Newfoundland dogs that are registered with the Kennel Club but not scored.

The only obvious explanation for this decline is the under-reporting of high hip scores. There is a great deal of anecdotal evidence for breeders having their dogs radiographed but not scored. It is likely that this occurs when the radiograph shows the potential for a high score. There is no information on the extent of this practice, which must tend to bias the statistics towards low values.

If this theory is correct then it might also be assumed that this practice might grow year-by-year as more and more breeders realise that it helps to reduce the apparent scores of their stock and kennels. Analysis of the trends for scores greater than 25 when parents have not been scored suggests that by 2010 the proportion of lost data may be as high as 15% .

0.8 Breeding Guidelines

It is always up to the breeder to decide which stock to use for breeding purposes, however this paper may help to give guidelines. The most important of these is that even low scoring parents can have progeny with very high scores. This means that hip scores should be only one of many factors to be considered. The main factors must always be the health of parents and close relatives, and of the need to follow the breed standard. If the breed standard is not given priority then there is a clear risk of realising the worst of all worlds: progeny with high hip scores that do not even conform to the breed standard!

A careful distinction should be drawn between high hip scores in a breed (which is certainly the case for Newfoundlands) and the actual occurrence of clinical hip problems such as lameness and arthritic tendencies. If these clinical conditions are prevalent in the available stock then serious consideration should be given to the hip scores of potential parents. If this not the case the hip scores should be treated as just another factor to be taken into account.

It is clearly the case that the breeding from low scoring stock increases the chances of low scoring progeny, but it does not eliminate the chance of high scoring progeny either. Breeding from pairs which both have high scores is clearly to be discouraged as the odds on high scoring progeny is substantially increased.

On some occasions a high scoring animal may have very desirable characteristics for the breed. In these cases it is good to know that the risk of high scoring progeny can be substantially reduced by choosing a mate with a low score of less than 14.

0.9 Conclusions

Newfoundlands dogs have hip scores that are on average higher than those of most other breeds, but there is still a core of normal dogs which have similar scores to the equivalent group in other breeds.

This suggests that all breeds have the same basic genetic influences (Newfies do not have "bad" genes" but rather they have a larger proportion of dogs with high hip scores). The reason for this difference is not clear, it might be that there are specific genetic influences which generate a high proportion of dogs with high scores, or perhaps a specific line of dogs, popular from before hips were scored, still influences present scores.

Whatever the reason it is nevertheless possible to use hip scores to select breeding stock that are more likely to produce low scoring progeny, with guidelines given in the previous section.

A key discovery is that doubt has been cast on the observed decline in annual averages of hip scores over the years. Instead it is suggested that this apparent decline is possibly due to the under-reporting, by as much as 15%, of high hip scores. This suggestion is so important that it merits further comment and study.

More details on all of these topics are given in the main part of this paper.

HIP SCORES OF NEWFOUNDLAND DOGS

An analysis by

John Morgan BSc

1. INTRODUCTION

1.1 Objectives

The primary objective of this study is to examine recorded hip scores for Newfoundlands (“Newfies”) in a systematic way in order to understand the extent to which they can be considered as heritable, that is, the extent to which hip scores in one generation relate to hip scores in the next. The study is primarily concerned with the hip scores as numbers and does not include any veterinary considerations. Nevertheless an attempt will be made to define what could be considered as a normal range of scores for animals of this breed at the usual age for assessment (often shortly after the first birthday).

If it is possible to define a normal range of hip scores then by implication this also defines what is abnormal, or Dysplastic. According to Wikipedia, “**Hip dysplasia** is an abnormal formation of the hip socket that, in its more severe form, can eventually cause crippling lameness and painful arthritis of the joints. It is a polygenic trait, that is affected by environmental factors in the production of the final phenotype. It can be found in many animals and occasionally in humans, but is most commonly associated with dogs, and is common in many dog breeds, particularly the larger breeds. Hip dysplasia is one of the most studied veterinary conditions in dogs, and the most common single cause of arthritis of the hips. “

This definition and many like it hinges on the interpretation of the word “abnormal” and begs the question as to what exactly defines a normal canine hip. Some papers use the words “Canine Hip Dysplasia” or CHD as almost synonymous with Canine Hip Scores, which of course cannot be correct. This paper examines the distribution of hip scores and other factors in an attempt to clarify this issue, always from a mathematical, not clinical, perspective.

1.2 Background

Canine Hip Dysplasia (CHD) is generally believed to be a condition that can herald the onset of crippling arthritis in later life. In order to reduce the incidence of the condition various recommendations have been made on breeding practice. Those recommendations are not based on the appearance of clinical signs (distress, lameness) but on scoring schemes based on the evaluation of radiographs made at a young age when the animal shows no clinical conditions.

While hips are scored in several different countries there are variations in rules and procedures between countries. In the USA two competing breed registries produce scores using radically different approaches. Thus there is no universal agreement on what is meant by a good hip score.

1.3 BVA/KC Hip Scores

Under the scheme organised in the UK by the British Veterinary Association and the Kennel Club, owners arrange for radiographs to be taken of the dog's hips after its first birthday and for them to be sent to the BVA to be scored. Nine separate assessments are made on each hip, with the nine scores added together to give a range for a single hip from 0 (good) to 53 (extremely bad). This paper focuses on the combined score for both hips, which has a maximum value of 106. It is a voluntary scheme used primarily to assess breeding stock which appear to be physically sound with no overt symptoms of any kind.

The Hip score results are kept by the KC in a database that includes details of the dog as well as its parents and their scores if known. The BVA website includes a table showing the Breed Mean Score (BMS), the average score for each breed.

The voluntary nature of the scheme is a significant disadvantage from the statistical point of view as many formal statistical tests apply only when any selection for measurement is made in a representative way. This is not true in this case. Breeders are free to examine the radiographs and choose not to send them for formal scoring. There is plenty of anecdotal evidence for this although the actual extent of the practise is not known. The result may be that the resulting statistics are biased towards lower scores.

The BVA does not indicate which scores it considers to be normal, or which it considers to be evidence of CHD, but advises that "Breeders wishing to reduce the risk of hip dysplasia should choose stock with scores well below the Breed Mean Score (BMS)". It should be noted that this advice only talks about "reducing the risk", not of eliminating risk.

1.4 Hip Scoring in the USA

Two competing hip registries exist in the USA. The Pennhip scheme based on work at the University of Pennsylvania uses radically different radiography, with the dog in a completely different position, and cannot readily be compared with results from the UK system. In this case the hip score is a single number known as the Distraction Index which is generally in the range from 0 to 1, from low to high scores respectively.

The registry operated by the Orthopedic Foundation for Animals (OFA) is particularly relevant to this study as it is very similar to that operated by the BVA/KC. The dogs are radiographed in the same position and a similar list of criteria are used to establish a grade on a seven point scale. The three grades at the lower end of the range are classified as normal, those at the high end are called dysplastic and the one in the middle of the range is called borderline. Dogs classified as borderline may be re-scored at a later date to determine if their status has improved or deteriorated, in which case they are re-classified.

As with the BVA/KC scheme, the OFA registry is voluntary. Furthermore, OFA offers the possibility to pre-assess dogs at an age as early as 4 months to determine if it would be useful to keep the animal for breeding purposes, in which case it would have a full evaluation at a later date. This option makes it even more likely than in the case of the UK scheme the statistical results are biased towards low values.

Results from this scheme are summarised on the OFA website, which includes details of the percentage of evaluations in each breed that are classed as dysplastic. This is a useful statistic as it offers a benchmark for the proportion of dogs with high hip scores.

1.5 Comparison of hip scoring schemes

OFA publishes an approximate comparison of hip scoring grades and scores, shown in Table 1. This is useful for a variety of reasons, not least because it includes definitions for “Normal” and “Dysplastic” scores, something that is unfortunately missing from the BVA/KC web pages. These definitions form a useful background to the work in this paper, although an attempt will be made to define breed specific boundaries rather than the one-size-fits-all version in Table 1.

Other Hip Dysplasia Registries—An Approximation					
OFA Grades			FCI (European)	BVA (UK/Australia)	SV (Germany)
NORMAL	Excellent	E	A-1	0-4 (no > 3/hip)	Normal
	Good	G	A-2	5-10 (no > 6/hip)	Normal
	Fair	F	B-1	11-18	Normal
Borderline		B	B-2	19-25	Fast Normal
DYSPLASTIC	Mild	M	C	26-35	Noch Zugelassen
	Moderate	Mod	D	36-50	Mittlere
	Severe	S	E	51-106	Schwere

Table 1. Comparison of Hip Score Categories published by the Orthopedic Foundation for Animals (OFA) in the USA.

1.6 A question of perspective

The BVA/KC system has hip scores ranging from 0 to 106, while OFA uses just seven grades to describe the hip score, based on similar criteria. The level of detail in the BVA score is undoubtedly of great value when making a statistical analysis such as presented in this paper but gives a possibly misleading impression of unjustified precision. It is based on 18 different assessments of features that are not independent of each other. If one of the nine assessments of a single hip has a high value then it is likely that many of the other eight assessments would also have high values. If the condition is genetic then it is likely that the second hip will have a similar score. In the extreme case the 18 assessments may have little more meaning than the same measurement repeated 18 times. It is sometimes helpful to divide the BVA score by 18. A score of 9 is then “only” an average score of 0.5 on a seven point scale. This sometimes helps to put scores into perspective.

2. NEWFOUNDLAND HIP SCORES

2.1 Available data

Hip Scores for over 3,000 Newfoundlands were obtained from the Kennel Club with test dates ranging from 1983 to July 2010. Until 1 September 1991 the KC published scores only when they were 8 or less, so all results before that date were discarded. This left 3222 scores for analysis. This forms a data set of useful size. Only seven other breeds had a higher number of evaluations in the summary published on the BVA web site in November 2009. These include Labradors, with more than 57,000 evaluations (by far the highest), and nearly 22,000 Golden Retrievers. Previous studies of those two breeds has provided important insights which help in the analysis of the much smaller number of Newfoundland hip scores in this study.

Previous studies have found that progeny with both parents scored have generally lower hip scores than those with neither parent scored. To ensure data consistency the data were divided into three groups; progeny with both parents scored, those with just one parent scored and those with neither parent scored:

	Number	Mean
Both parents scored	1996	22.9
One parent scored	574	23.5
Neither parent scored	652	28.1

When neither parent has been scored the two parents are representative of the large number of Newfoundlands (about 80%) that have been registered with the KC but not scored. The high score for this group is therefore indicative of the situation in the wider population of Newfoundlands.

By contrast, the group with both parents scored is the result of attempts to reduce hip scores through at least one generation. The reduction from about 28 to about 23 is therefore an indication of the success of the hip scoring scheme.

In most cases the group with both parents scored has been used in the analysis, as heritability is a main focus of this paper. In some cases the use of all the data is appropriate, and in others the results when neither parent is scored form a useful "natural" baseline, as these represent cases free of any change due to a breeding policy based on hip scores. This will be made clear in the text.

2.2 Relative rankings

The official Breed Mean Score recorded by the BMA had a value of 26 in November 2009. With this score Newfoundlands rank as number 140 in a list of 147 breeds. Only seven other breeds have a higher BMS.

In the USA, the OFA statistics list the percentage of dysplastic dogs, roughly equivalent to a score higher than 25 in the UK system. Using this percentage, Newfoundlands rank as number 136 in the list of 157 breeds.

These statistics clearly indicate that this breed must have a very high incidence of CHD, but this is not the whole story, as will be discussed in later sections.

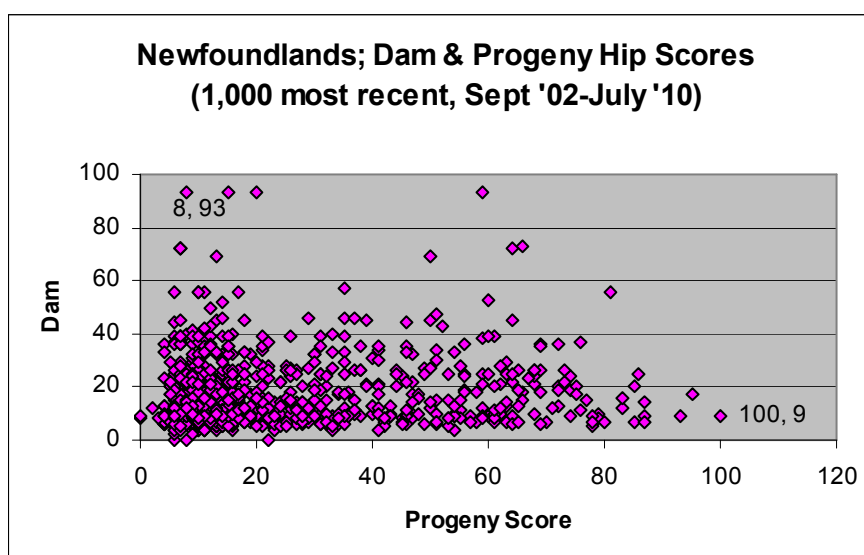
3. PREDICTABILITY OF INDIVIDUAL SCORES

3.1 Dam-Progeny Scatter Chart

It is a common experience, supported by widespread anecdotal evidence, that hip scores can at times appear to be completely random, with low scoring parents having occasional progeny with very high scores, or the reverse. Figure 1 is a graphic demonstration of this effect. Each point on the chart represents one animal, with its score shown along the horizontal axis and the score of its dam along the vertical axis. The chart is restricted to the most recent 1,000 evaluations to make sure the results are as current as possible and also to ensure that the chart is not totally obliterated in ink. A few of the data points are labelled with the actual scores.

The occasional random nature of the distribution is clear to see. Most scores, for both the Dam and Progeny, are concentrated in the range from 5 to 25 but there are many cases outside this zone. The highest progeny score is the data point on the extreme right, with a score of 100 when the score of its Dam was 9. At the other extreme a Dam with score of 93 has one progeny with a score of 8.

Figure 1
Scatter plot of progeny scores (horizontal axis) against Dam scores (vertical axis). Each point represents one animal. Actual values are shown for a few points; these show the progeny score followed by the score of its dam.



3.2 Other Scatter Charts

The scatter chart showing the relationship between Sire & Progeny is very similar to that shown for the Dam in Figure 2 and need not be shown here. The same pattern is also shown in other breeds, where for instance a Labrador with a score of 94 is found to have parents with scores 9 and 7 for the sire and dam respectively.

This is exactly the feature of hip scores that this paper attempts to address. Is this a random effect caused by external environmental factors such as early trauma? Or can it be explained by genetic considerations?

This confirms what we already know; that high hip scores can result, apparently at random, from low scoring parents. But is there a genetic cause?

3.3 The Mathematical verdict

Mathematical correlation defines the extent to which one set of numbers (progeny scores in this case) depend on an associated set of numbers (scores of dam or sire). This is a number on a scale of 0 to 1, where a value of 1 would mean that the progeny score could be predicted accurately without error. The progeny-sire and progeny-dam correlations for Newfoundlands are just 0.05 and 0.20 respectively, both being far too low to be of any use in helping to predict an individual hip score.

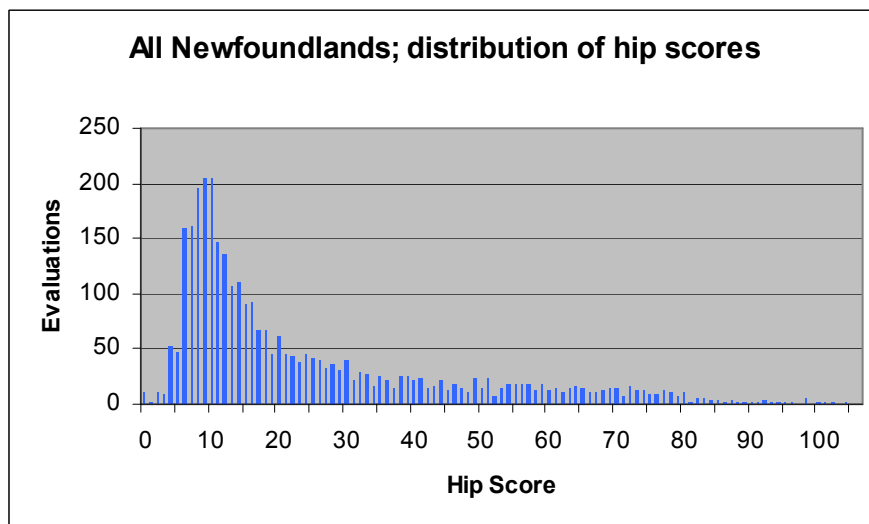
The charts and the mathematical correlation both make it clear that it is impossible to predict individual hip scores from the scores of the parents. The remaining possibility is to use the available information in an attempt to predict the probabilities of scores within certain ranges. This will be addressed in a later section.

4. HIP SCORE DISTRIBUTIONS

4.1 Newfoundland Hip Score Distributions

Figure 2 shows the distribution of hip scores for all Newfoundlands. The distribution has a very distinctive shape, with a high peak when hip scores are low and a long tail containing relatively few evaluations extending to high hip scores.

Figure 2.
Distribution of Hip Score for all Newfoundlands, for the period Sept 1991 to July 2010, showing the number of evaluations for each possible score.

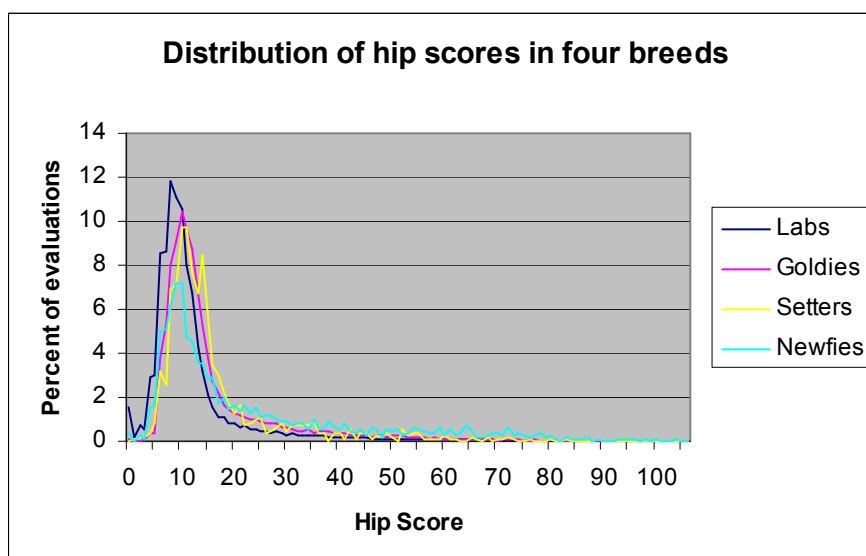


4.2 Hip Score Distributions in other breeds

The distinctive shape of the distribution of scores shown in Figure 2 is not confined to this breed. Figure 3 shows the percentage of hip scores in four breeds, all showing the same distinctive shape.

There is strong evidence from the summary statistics published by the OFA in the USA that all breeds have similar distributions, implying that this is a characteristic imposed by the underlying genetics, rather than through environmental effects.

Figure 3.
Distribution, in percentages of evaluations, of hip scores in four breeds when both parents have been scored. The breeds are Labrador, Golden Retrievers, English Setters and Newfoundlands. All four show similar distributions.



4.2 Normal or what?

The peak in the distributions shown in Figures 2 and 3 triggers an immediate reaction in the mind of any statistician, as this distinctive shape arises when you take independent measurements of almost anything. If you measure the exact length of a kilo 2” nails and plot the resulting distribution it will have a bell-shaped curve similar to these peaks. The heights of 12 year-old boys will look the same, as will the lengths of the backs of any particular breed of dogs. There is nothing new in this, the chest sizes of 5,000 Scottish militiamen in the early part of the 19th century show a similar distribution.

Statisticians have found that this type of distribution can be represented by an equation that forms the basis of mathematical statistical analysis. This is so prevalent that it is usually called the Normal distribution, but if you want to be posh you can call it the Gaussian, named after the prolific mathematician Carl Friedrich Gauss (1777-1855).

Figure 4
Distribution of Newfoundland hip scores expressed as percentages of all evaluations, together with a Normal distribution chosen as the best fit of the distinctive peak

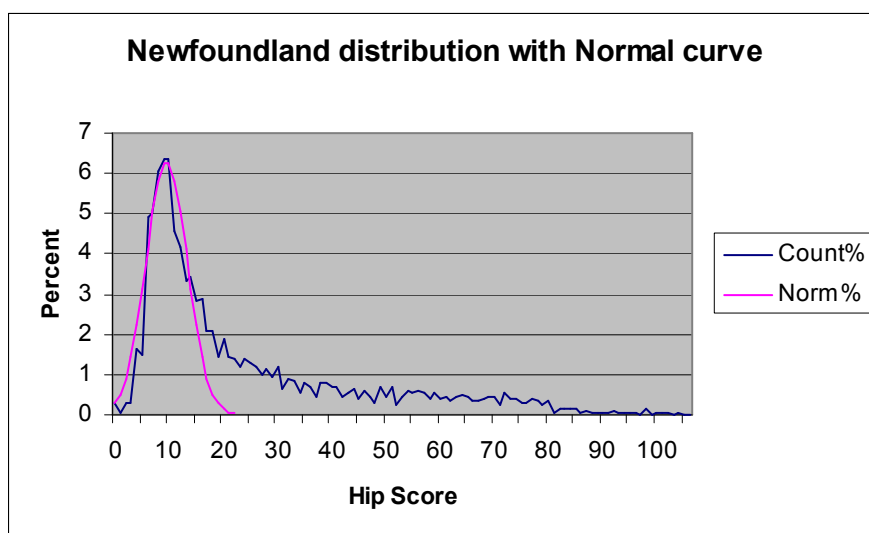


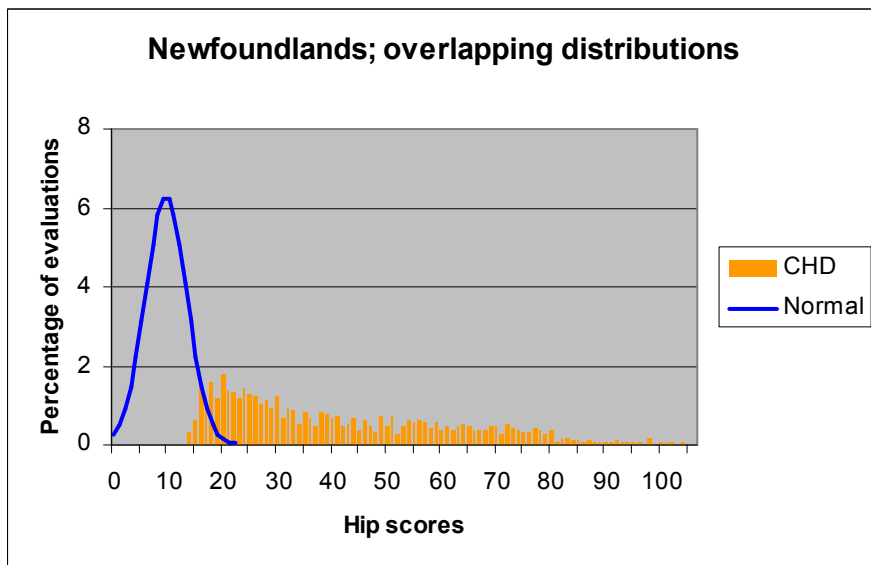
Figure 4 shows the extent to which a mathematical “Normal” curve fits the distribution of Newfoundland scores.

It is not an exact fit but close enough to indicate that something interesting is happening.

It is suggested that the scores represented by the Normal curve are indeed the normal range of scores associated with dogs free of overt symptoms of CHD. Those “outside” this curve are not explained by this Normal distribution and the obvious conclusion (supported by other studies) is that these represent dogs with varying amounts of CHD.

This theory is perhaps clarified by Figure 5, where the evaluations “under the curve” have been subtracted from the total number of evaluations, leaving residual values representative of CHD.

Figure 5.
As Figure 4 but with “Normal” evaluations replaced by the mathematical curve, the remaining data shown as CHD.



4.3 Comparison with Labradors

Figure 6 shows the distribution of hip scores in Labradors, a breed with a BMS of only 15, arranged in the same way as for Figure 5.

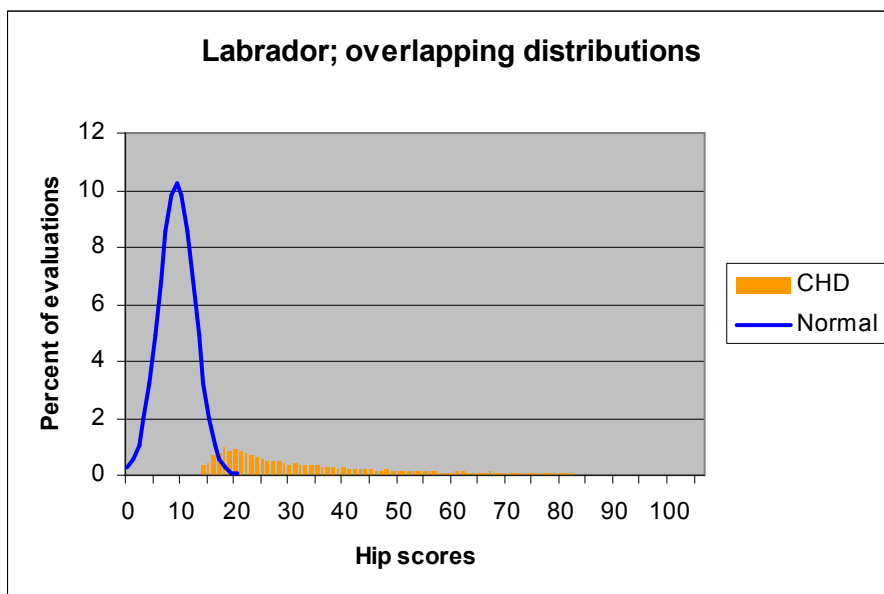
The striking thing about this comparison is the similarity of the two regions classed as “normal”. In both breeds the peak of the distribution has a value close to 9, so it could be said that the hips of a “Normal” Newfoundland are almost identical to the hips of a “Normal” Labrador.

The main difference between the figures is that the Newfoundlands have a much larger proportion of dogs with CHD (as shown in orange in Figures 5 and 6). The difference between the two breeds is not the breed as a whole having “worse” hips, but is due to the higher proportion of Newfoundlands with high scores representative of CHD.

Figure 6.
Labradors, showing overlapping results for normal dogs and for dogs with increasing degrees of CHD.

The blue line indicating normal hip scores is centred at the same hip score as for Newfoundlands.

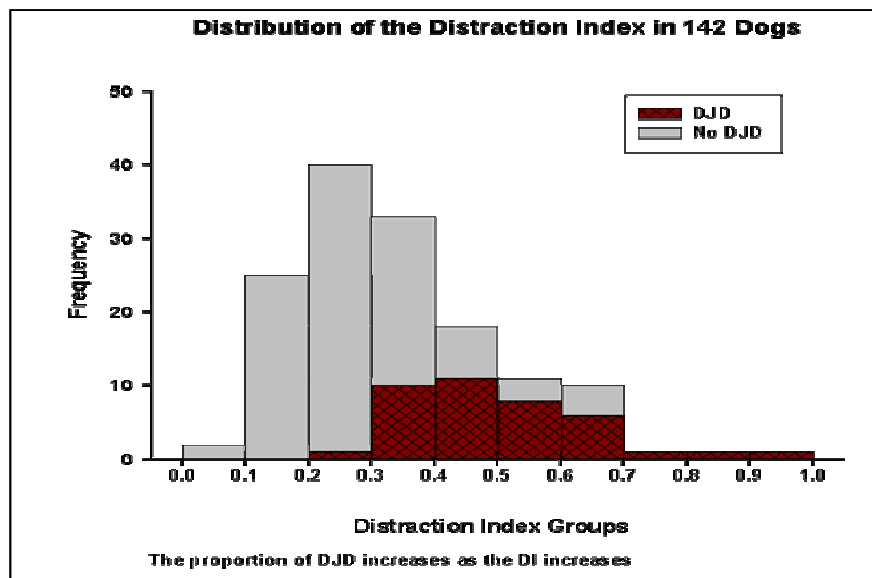
The key difference from Newfoundlands is that the proportion of evaluations classified as CHD is much lower.



4.4 Normal and CHD ranges

It is worth noting that the website of the Pennhip group in the USA contains a similar diagram, showing overlapping zones of dogs with and without what they call Degenerative Joint Disease. This is reproduced in Figure 7, effectively showing a low resolution pixellated version of the distributions shown in Figures 5 and 6. This diagram provides support to the theory developed in this paper.

Figure 7.
Distribution of Distraction Index (a hip score) by the PennHip group in the USA



Figures 4 and 5 suggest that “Normal” Newfoundland dogs free of the overt signs of CHD have scores in the range up to 21. Dogs with CHD occur with scores from 14 and higher. This establishes three zones:

- Dogs with scores less than 14 are considered to be normal, subject to natural variations within this range
- Dogs with scores above 21 have scores relating to increasing degrees of CHD
- The range 14 to 21 is borderline, with some normal dogs but with an increasing proportion of dogs with CHD as the scores increase in value.

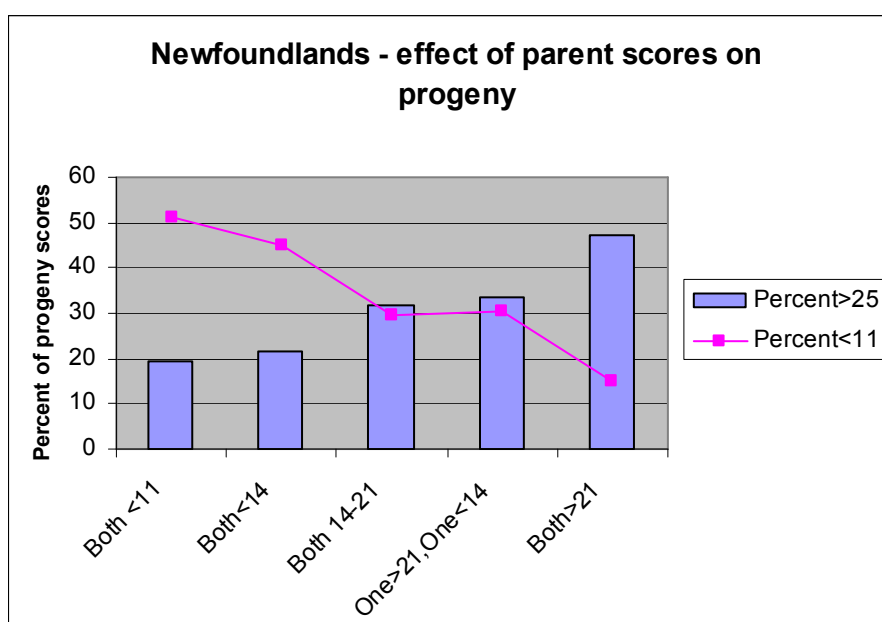
This does not mean that the dogs in these groups have different genetic makeup. It has already been demonstrated (Figure 1) that a low scoring parent can have high scoring progeny. It will be demonstrated in a later section that pairs of low scoring dogs can also have high scoring progeny. The relevant genetic factors are present in all dogs and are not isolated in particular groups.

5. EFFECT OF PARENT SCORES

5.1 Effect of parent scores on progeny

The previous section was long and rather complicated, but important to the understanding of the impact of parental scores on their progeny. Figure 8 shows how the proportion of good and bad hip scores of Newfoundland progeny varies according to the scores of their parents. It was established in Section 3 that it is impossible to predict the score of an individual puppy or even the average score for a particular litter, but this diagram shows how progeny scores vary when a large sample is averaged. It has to be stressed that these calculations apply only to Newfoundland dogs—other breeds may show different variations.

Figure 8. Percentages of progeny with low scores (less than 11) and high scores (greater than 25) for various categories of parent scores.



The ranges illustrate a clear progression from left to right, with the percentage of dogs with high scores increasing and the percentage of good hips decreasing at an even faster rate. Predictably, the best results are those for parents with scores less than 11, with a modest increase when the range is extended to 14.

There is a marked change when the parental range rises to 14-21 with an even more marked deterioration when parent scores are both greater than 21, when the proportion of low scoring progeny is half that of the next group.

It is not always possible to choose two parents with good hip scores, as suitable breeding stock may be limited and consideration must also be given to many other characteristics. The diagram shows that if one parent has a high score the results can be improved by choosing a mate with a much lower score.

5.2 Effect of parent scores on breed mean scores

While individual breeders may be concerned with the effects on progeny, breed committees might be equally concerned about the impact of different breeding policies on the breed mean score. Figure 9 shows the mean scores for the same five ranges of parental score, with the same pattern of results as in the previous section.

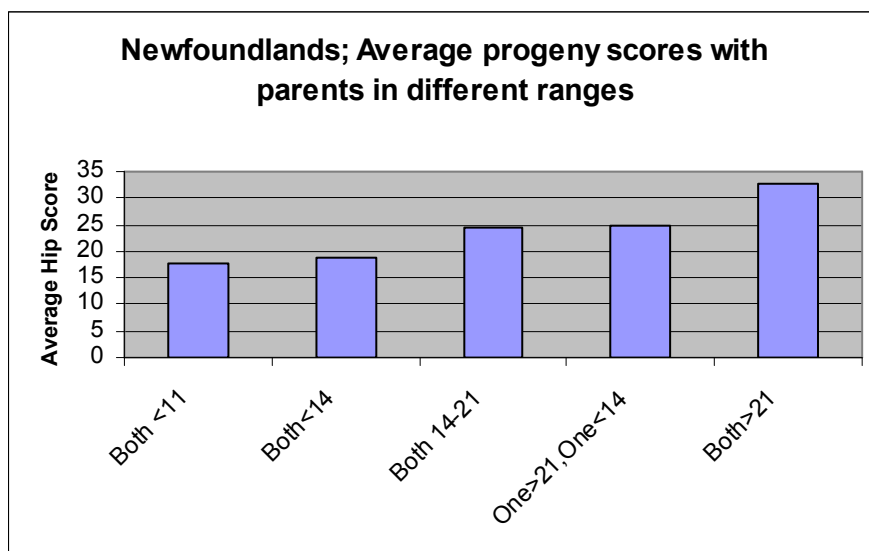


Figure 9. Average progeny scores for parents in five ranges of scores

It is recalled that the official BMS is 26, based on 4,000 scores. The mean score here is 24, based on 3,200 evaluations, This target value can be compared with the results shown in Figure 9:

The example at the extreme right show scores significantly greater than this target of 24. It is obvious that repeated breeding from dogs in this group would have the effect of raising the BMS over a period of time and must clearly be discouraged.

The two categories at the left of Figure 9 have average scores less than the target of 24, so that repeated breeding from these groups would tend to reduce the BMS over time. The parents in the range 14-21 (defined as borderline in Section 4.4) and the example with one high and one low score are marginally higher than the target of 24 but marginally less than the official BMS. Progeny from these two groups could have very little impact on the breed mean score.

In fact, none of these possibilities would have a great effect on the BMS. About 150 Newfoundland hips are evaluated each year. If these were restricted to progeny of parents both scoring less than 11 the BMS would drop by less than 0.3 points. The result is dominated by the 4,000 scores already in the database, not by the 150 scores added in any one year.

This result explains why most breeds find it very difficult to reduce the BMS in spite of strong recommendations to breeders to use low scoring stock.

6. THE TRIPLE GENE THEORY

After the complexities of Section 4, and the practicalities of Section 5, it may come as light relief to consider the genetic background. It is generally accepted that CHD is polygenic in origin, meaning that it is caused by interactions of more than one gene. Explanations usually add that the genetic factors are then modified by environmental conditions, including nutrition, trauma, exercise, among other factors. It seems that the specific genes have not been identified but that does not exclude a more general analysis using basic genetic theory. It was surprising to discover that a simple genetic model using classical Mendelian¹ theory could explain all of the characteristics of observed progeny scores as illustrated in previous sections.

Well-established genetic theory indicates that the characteristics of an animal or plant are controlled by a large number of gene-pairs, where one gene in each pair is inherited from each parent. Each gene is said to be either Dominant or Recessive. These names arise because if just one of the gene pair is Dominant then the trait exhibits the associated characteristic (assumed to be a low hip score in this example), whereas both genes of the pair have to be Recessive for the animal to show the recessive characteristic (a high hip score).

On the assumption that three gene pairs acting together are associated with hip quality, then it can be shown that these can combine in 64 different ways. 27 of these cases have no recessive pairs and therefore show the dominant characteristic of good hips, with differences between individuals arising naturally through environmental factors. The theory also confirms that even these low-scoring cases are carriers of the same genetic factors which cause high hip scores. This is why even low scoring parents can give rise to high scoring progeny as illustrated already in Figures 3 and 8.

10 of the 64 cases (15.6%) include two or three recessive gene pairs, so these must be expected to show high hip scores. This is to be compared with the OFA statistics where the all-breed average for dysplastic dogs is also 15%. The theory indicates that if both parents fall into this category the progeny are also very likely to have high scores, with a substantially reduced chance of having very low scores. This is illustrated in Figure 9, which shows that high scoring parents have about a quarter of low scoring progeny compared with parents with lowest scores.

A further 27 instances have exactly one recessive pair of genes, which should show intermediate characteristics between these two extremes. This corresponds in part to the overlap area between normal and CHD results shown in Figures 6 & 7.

This short section demonstrates that there is a sound theoretical basis for the statement that hip scores are to a large extent controlled by genetic factors, and are not simply wildly random as often seems to be the case. A fuller explanation with more details is available from the author of this paper.

¹Gregor Mendel was a 19th century Augustinian priest who developed classical models of inheritance, based on the study of the colour of flowers in pea plants. His work was dormant for half a century until re-discovered in the early part of the 20th century, leading to an explosion of interest and research

7. TRENDS IN NEWFOUNDLAND HIP SCORES

7.1 An apparent decrease over time.

Figure 10 shows the values of the annual mean hip score (shown as blue diamonds) and the calculated trend line (least squares linear regression), together with its equation. Although there is a high degree of variability, the R^2 value shown on the chart is high enough to confirm that the downward trend is mathematically realistic, suggesting that the mean score has decreased by about 7 points during this period. However, this is not the whole story, as the following paragraphs demonstrate.

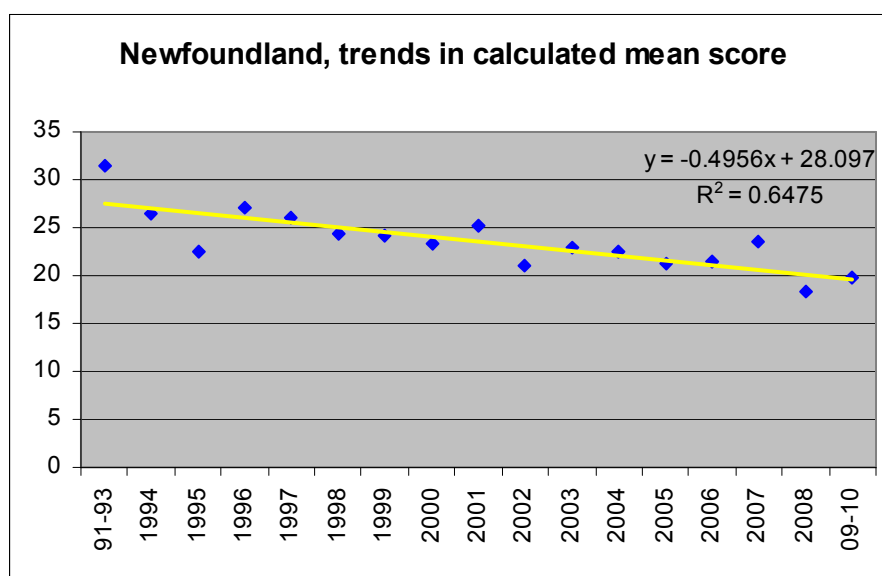


Figure 10.
The trends in calculated annual mean hip scores for Newfoundlands over the period 1991 to 2010 (both part years). Not all of the apparent downward trend may be real.

7.2 Some doubt is cast

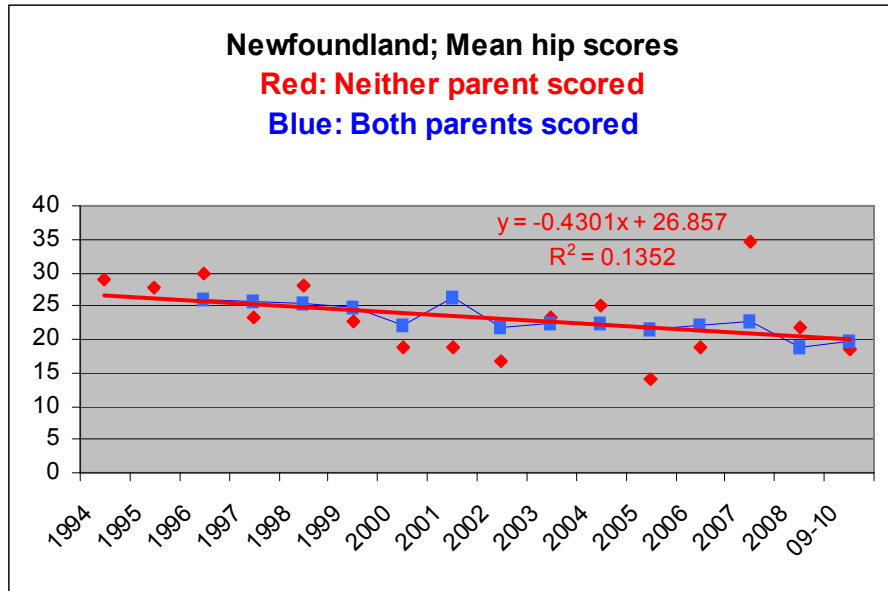
Many people in this breed and others have expressed concern about the probable under-reporting of high hip scores. This may arise when experienced breeders examine a radiograph and decide not to submit it to the BVA for scoring. It may also arise if experienced radiographers use extreme methods (such as multiple radiographs or manual restraint of the dog) to achieve “good” results.

All of this is quite legal (except possibly the use of manual restraint for the radiograph) but makes it difficult to draw firm conclusions about the data, as the inevitable result is to bias the statistics towards low unrepresentative values.

The extent of such practices are unknown, but a clue exists in diagrams such as Figure 11, which shows the trends for Newfoundlands when both parents had been scored, and when neither parent had been scored.

The downward trend with both parents scored (blue squares) is to be expected, as this is similar to the trend for all dogs as shown in Figure 11. However, the downward trend when neither parent has been scored (red diamonds and trend line) is entirely unexpected.

Figure 11. Trends in Newfoundland Hip Scores when neither parent had been scored (red) and where both parents had been scored (blue). The decline when neither parent has been scored is unexpected and needs some explanation.



It has to be said that the annual values (red diamonds) are highly variable and the R^2 value (a measure of the quality of the trend line) is very low. Based on this evidence alone the statistical basis for any conclusion is very weak. However, visually a downward trend is apparent, the slope of the trend line when both parents are scored is such an exact match that it cannot be shown on the same diagram, and the effect shows up with greater clarity in breeds where there are more evaluations. All this leads to greater confidence in the calculated trend line. This additional evidence suggests that the downward trend shown by the red line is probably real. If it is indeed real, and the slope is the best estimate we can make, what is the explanation?

This is a highly significant discovery.

If hip scores decline from year to year even when neither parent has been scored then selective breeding based on hip scores cannot be the cause of the decline.

The most obvious explanation is the failure to score radiographs thought to show high hip scores, as this would bias the averages towards low values.

It is not unrealistic to believe that this practice may have grown year-by-year as more and more breeders become aware of this possibility, explaining the apparent long-term trend

The parent scores are not known; the parents have not been selected because of their hip scores, nobody can know their hip scores as this is only possible with a radiograph, so there is no obvious reason for these scores to decrease over time. The un-scored parents are representative of the vast majority of Newfoundlands that are registered but not scored, so the decrease indicated by the red trend line is a puzzle.

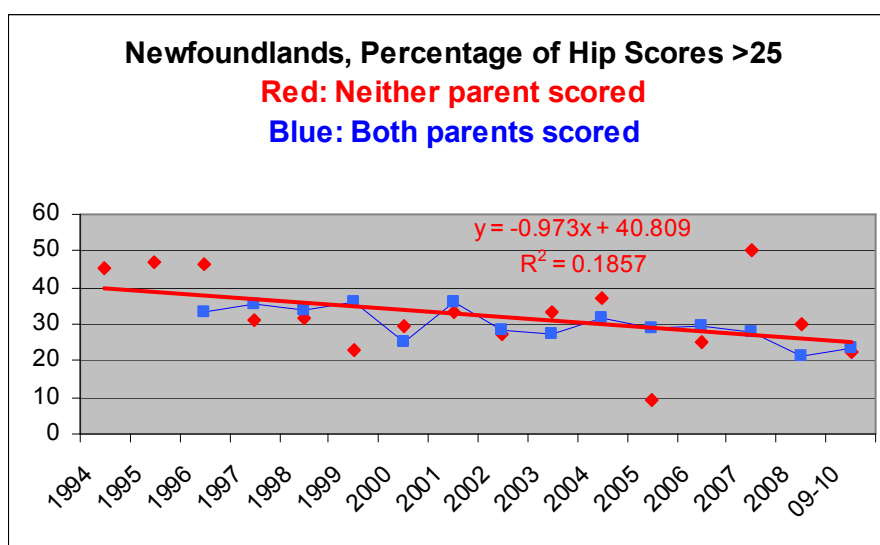
One possible explanation for this decline is the under-reporting of high hip scores. This is likely to affect dogs with both groups to the same extent, which is why the two trends are so similar. If this is the case, then the under-reporting has increased over the years to depress the calculated annual hip score by 6.5 points in 2009-2010. This is close to the apparent decrease in hip scores (7 points) shown in Figure 10, leading to the conclusion that hip scores have probably not decreased at all during this period.

7.3 Impact on scores greater than 25

The two previous charts have shown the annual variation in mean hip scores. Figure 12 shows that the percentage of scores greater than 25 also decreases when neither parent has been scored (red) as well as when both parents have been scored (blue). As in the previous example, it is difficult to understand why the scores should decrease when neither parent has been scored.

The under-reporting of high hip scores is again the prime suspect for this effect.

Figure 12. Trends in the percentage of progeny scores higher than 25 when both parents have been scored (blue) and when neither parent has been scored (red)



As can be seen from the chart, the annual values are highly variable from year to year, and the R^2 value is very low. This weakens the statistical case, but other evidence, as discussed for trends in the mean scores, suggests that the downward trend still looks to be a reasonable assumption, implying under-reporting of high hip scores which rises year by year to about 15% by 2010.

7.4 Newfoundland Trend Summary

The evidence suggests that the downward trends in Newfoundland hip scores shown in Figure 10 are probably largely illusory and arise from the under-reporting of high hip scores, which appears to have risen over the years to about 15% of all scores.

This conclusion is not an absolutely certain result as the numbers of cases each year are low and the variability high. This is not ideal from the statistical point of view, but is consistent with the results in other breeds with far higher numbers of cases.

This conclusion is also supported by the calculations made in section 5.2, which indicated that even if all breeders had selected stock with very low hip scores (which does not happen) there could only be a marginal decrease in mean scores from year to year.

8. BREEDING GUIDELINES

It must be stressed that the hip scores are just one aspect of the complex considerations that lead to the choice of parents for an individual mating. This paper has not discussed the clinical implications of hip scores because it has not been possible to find a single scientific study of the relationship between hip scores and problems in later life.

This may be due to a lack of research talents of the author or the failure of veterinary research journals to publicise relevant papers on the internet, but for whatever reason it is crucial to recall that this paper is about hip scores as numbers, not as indications of hip disease.

It is the author's opinion that any breeding decision must give priority to the actual health of the parents and their conformity to the breed standard, not to the hip score. If the parents or close relatives do suffer from any clinical signs of hip disease then hip scores might be given higher weight. However, it is important to realise that low scoring parents can have high scoring progeny. If health and breed standards are ignored by giving priority to hip scores then the result could be progeny with high hip scores that do not conform to the breed standard either. Be aware!

That being said, Section 5 provides evidence that on average the hip scores of progeny can be reduced by choosing parents with low hip scores. This is fully consistent with the advice given by the BMA as quoted in Section 1.3. In general the lower the parental scores the higher the probability of good hips, but three ranges have been identified which help quantify the relevant probabilities, as shown already in Figure 8 and Figure 9.

If both parents have scores of less than 14 (considered to be normal) then about one in five of their progeny would be expected to have scores higher than 25, while about half are likely to have scores of less than 11.

When both parents have scores greater than 21 (considered to represent CHD) nearly half of all progeny should be expected to have scores higher than 25, with the proportion of low scores (less than 11) falling to one in seven. It seems obvious that cases when both parents are greater than 21 should be avoided.

If both parents are in the range 14-21 they fall in the transitional or overlap region between normal and CHD hip scores. Within this range the lower the scores the better, but overall the proportions of progeny with good hips (less than 11) or high scores (greater than 25) are both slightly less than one in three. These same proportions also apply if one parent has a high score (over 21) and one parent has a low score (less than 14).

Of course, these detailed findings apply only to Newfoundland dogs; in other breeds different proportions may apply.

It is also worth repeating the point that the mathematics demonstrate (Section 5.2) that it is very difficult to reduce the Breed Mean Score by selective breeding. The value of the BMS is determined primarily by the 4,000 dogs that have already been scored, not by the scores of the 150 or so that are evaluated in any particular year.

9. CONCLUSIONS

Hip Scores from over 3,200 Newfoundland dogs evaluated during the period September 1991 to July 2010 were analysed using standard statistical methods, leading to several important conclusions:

- a. When both parents have been scored the average scores of their progeny is about 5 points lower than in cases where neither parent had been scored. This is a clear indication of the value of the hip scoring scheme.
- b. The statistics confirm that low scoring parents can have progeny with very high scores, and it is impossible to predict the scores of individual progeny from the parental scores. This is consistent with genetic considerations and it is not necessary to ascribe high scores to trauma or other environmental factors.
- c. Consideration of the distribution of hip scores leads to a definition of normal hips as falling in the range up to 21, with cases of CHD occurring in the range from 14. The overlap range from 14 to 21 is transitional or borderline, with increasing proportions of dogs appearing to have CHD.
- d. Parents in these three ranges have significantly different chances of having progeny with high hip scores, defined as scores higher than 25. With parents both having scores less than 11 the proportion of progeny with high hip scores is about one in five. In the transitional range 14-21 it is about one in three. This proportion also applies with one high scoring parent paired to a low scoring parent. When both parents have scores greater than 21 about half of their progeny must be expected to have high scores (greater than 25)
- e. It is mathematically impossible to reduce the Breed Mean Score by any significant amount in any one year, when only about 150 dogs are scored, because the BMS is dominated by the 4,000 previous evaluations.
- f. There is evidence that the apparent decline in hip scores over the years is probably due to under reporting of high hip scores, rather than a real decline. The apparent decline may be due to radiographs being taken but not submitted to the BVA for evaluation because the outcome is likely to be a high score. The proportion of missing scores appears to have risen each year since 1997. By 2010 as many as 15% of all high scores may be missing from the statistics because of this and similar practices. (Section 7)
- g. All of the observations and conclusions are consistent with basic genetic theory based on the assumption that hip scores are affected by three interacting gene pairs (Section 6).
- h. It is necessary to repeat that these outcomes are a reflection of a statistical analysis of hip scores, not clinical conditions. Furthermore, the statistics have been used to infer the most likely interpretation and as ever the conclusions might be subject to error.