

CANINE HIP SCORES

Research Papers by John Morgan, Bsc Hons (London)

Summary of essential results, Version 1.2, May 2010

INTRODUCTION

1. To put these remarks into context I should point out that my wife, Angela, is an extremely experienced and successful dog breeder (Quensha). Although her kennel is small, rarely containing as many as 10 dogs, she and her partner Richard have made up 10 Champions in five different breeds. Between them they certainly know how to pick a dog and how the intelligent use of genetics can be used to advantage in a breeding programme. On the other hand, I am not much involved with the dogs, although I have lived with them for more than three decades. My own interest is in Science and Mathematics, especially Statistics. This is not merely a hobby; my statistical analyses are still the scientific basis for the launch programme of one of the world's leading satellite organizations.
2. In a small way I aim to bring a little rocket science to this sensitive and poorly understood issue. I was drawn into the debate in 2005, when Angela was widely criticized for using a dog with a hip score of 27:25 with her own low scoring bitch. I certainly wanted to support her during this difficult time, but could not feel fully committed until I knew a little more about the subject. I therefore made an extensive statistical analysis of the hip scores of 1,700 English Setters, 18,000 Golden Retrievers and 43,000 Labradors.

BVA/KC HIP SCORES

3. The results were unequivocal; the distribution of scores across all three breeds is remarkably similar. This could not happen by chance so this demonstrates that variation in measured hip scores is certainly genetic in origin, with environment playing only a small role. However the heritability is very low and the genetic factor appears only when averages are taken over large numbers of individual animals. Parents with very low scores can have progeny with very high scores while high scores in parents do not give rise automatically to high scores in their progeny. In mathematical terms the statistical correlation between either parent and its progeny has the low value of around 15%. This has to be compared with the 100% needed to predict the score of the progeny with perfect accuracy.
4. Those who know a little about genetics are prone to say at this point that "ah, but these things skip a generation". Well it is only necessary to read one book on genetics to find that this may only be true in the simplest of scenarios, with a single gene causing a specific effect, but it is certainly not true in the case of polygenic conditions caused by several genes. When this argument fails there are those who then go on to say, "Ah, well, you have to study the grandparents, the parents' siblings, the progeny siblings and the progeny of the progeny before you get the full, true picture". This advice is to be found in many well respected books, but only serves to further confuse the issue, as no hint is given on how to interpret such

information. It is clearly true that if you cast your net wide enough you may find an apparent link, but statisticians would dismiss this along with the claim that the increase in sales of bicycles between the World Wars was the cause of the increased admissions to mental hospitals over the same period.

5. To test the generation skipping potential, in 2009 I extended my study to focus on those English Setters for which hip scores were known for both parents as well as for all four grandparents. Absolutely no pattern could be observed in these results. High scores could arise when all 6 of the parents and grandparents had impeccable scores, while high hip scores for older generations normally resulted in low scoring progeny. In spite of the acknowledged genetic causes, progeny scores can be, as everyone knows, apparently completely random.
6. As a result of my studies I became convinced that hip scores are only one aspect to be taken into account in a breeding programme. First, there is clear and conclusive evidence that high scores in the parents are not transmitted directly to their offspring. Secondly, and perhaps more to the point, there is hardly any evidence to connect high hip scores with any observable distress or suffering by the dog. Everyone seems to know of a dog with a high hip score but impeccable movement who can jump five bar gates, and there are some who admit to seeing dogs with scores of 0:0 so stiff that they can hardly walk around the show ring.

DISCUSSIONS AT THE KENNEL CLUB

7. I was so concerned about these findings that I sent discreet copies of my papers to a few influential people who I thought should be made aware of the facts I had uncovered. As a result I was invited to a meeting at the Kennel Club with several genetic experts, including Dr Malcolm Willis, who has done more than any other individual to promote routine hip scoring. (To her mild annoyance the invitation was not extended to Angela, who is of course the partner passionately involved in the dog world. How come I got to Clarges Street before her?).
8. I was shocked to discover that nobody at the meeting could point me in the direction of substantial research that linked high hip scores with an actual condition such as arthritis in later years. Dr Willis had some information on GSDs, but even this was somewhat small scale. In view of current controversies, nobody could claim that GSD studies on hips are relevant to the wider canine world. I have suggested that the KC conducts some research into this issue, perhaps by tracking a group of dogs throughout their lives to determine how hip scores affect their later health. The BVA have hip scored over 200,000 dogs during the past three decades. In our experience the total cost for each test including vet's fees is over £200, so this activity has cost us some £40m over the years. Surely some of this money should be used for basic relevant research?
9. Discussion then focused on the role of hip scores in a breeding programme. No minutes were kept or announcement made, but my own notes on the meeting (checked and confirmed by two other interested parties) indicate that all participants, including Dr Willis, completely agreed that hip scores were only one of the many aspects to be taken into account when considering a mating. Other conditions for which heritability is greater (such as some skin conditions) should

be of equal or greater concern, as should general conformation to the breed standard. There is no point in attempting to breed to a low hip score only to find you still have a high score and a dog that does not adhere to the standard either!

10. This meeting entirely vindicated (unfortunately not publicly) Angela's approach to a highly successful breeding programme. If she had the choice of two dogs with similar characteristics then clearly she would tend to choose the one with a lower hip score, as this at least shortens the odds against extremely high hip scores in the progeny. In other cases a high hip score would have to be evaluated against other valuable criteria. Her instinct paid off, and the proof is as they say, in the eating. None of the first or second generation progeny of the controversial mating has a high hip score. The latest (Score 5:5) moves like a dream, hardly touching the ground when in full flight, and impressing even non doggy people like myself. Fit for purpose? Undoubtedly. Does she adhere to the Breed Standard? Eight CCs awarded well before her third birthday would seem to suggest so!
11. My 2005 studies indicated that because little of the observable genetic effect is passed directly between generations it would be very difficult to improve the hip scores by selective breeding. My analyses showed that significant improvements from generation to generation could only be achieved by consistently breeding only from parents with extremely low scores. For English Setters I found that the average progeny score of 16.5 when the scores of both parents were disregarded fell to just 14.9 when both parents scored less than 10, as widely recommended at the time. It seems hardly worth the effort, especially as selecting English Setter parents with scores of less than 10 would reject 88% of the breeding population. This loss of genetic diversity from the gene pool is clearly unacceptable. All for a measurement that is not firmly linked to any ailment!

THE AMERICAN EXPERIENCE

12. Much of the above is derived from the details of thousands of individual animals in the three gundog breeds as supplied by the KC. The KC were kind enough to supply (for a fee) an update in 2010 with many more data which I look forward to analyzing in due course. I should also note that I have also looked carefully at the summary statistics available from the USA. The *Orthopedic Foundation for Animals (OFA)* uses a similar radiograph based scoring system. The OFA do not publish the same type of summary statistics as the BVA, but does list the percentage of dysplastic dogs in each breed. Furthermore the OFA website suggests that in their view "Dysplastic" corresponds to a BVA measurement of "above 25". This also corresponds to a discontinuity that I found in the distribution curves for each of the three gundog breeds. In the USA statistics some 15% of all dogs are dysplastic according to this definition (it varies by breed). This number also corresponds to the percentage likely to be dysplastic from consideration of a simple genetic model, as discussed later. All this suggests that the magic number of 25 is indeed an indication of where the problems start.
13. The OFA statistics of dysplastic dogs are similar to the percentage of scores above 25 found for the three gundog breeds I have studied. Furthermore, when all of the breeds measured by OFA are ranked in order of hip score alongside those from the BVA/KC there is a close

correspondence in the ordering of the two lists, suggesting that the two measurement techniques are fairly compatible and that there are generally no large differences between US and UK strains. This fact was to prove useful in further studies, where the USA statistics confirmed that the findings I had somewhat laboriously obtained for three gundog breeds in the UK were by and large representative of the canine world as a whole. Most if not all breeds have the same genetic tendency to widely varying hip scores, with the vast majority (85% overall) classed as in the normal range.

ARE BREED MEAN SCORES ACTUALLY IMPROVING?

14. In spite of the above there are papers in the dog world that congratulate a given breed on having improved its average hip score. Well done! In the UK press such remarks are based on an apparent decrease of the breed mean score, year on year, as published by the BVA and KC. Generally such reductions are very small, as one would expect, but are they real? Here the statistician has to step up to the line with some very firm remarks. First, the breed mean score could only be completely accurate if ALL dogs in the breed were scored. We know this is not the case. Second, in spite of this statement, a very good approximation to the breed mean score could be obtained by scoring a sufficiently large representative sample taken without prejudice from the entire breed. Here the important keyword is representative. If, say, the KC decided that every 10th registered dog in a given breed would be scored, then so long as there were several hundred registrations in a given year, then the average of that one in ten sample could be relied on statistically as a good estimate of the breed mean score taken as a whole.
15. We know that hip scoring is not done in a random way across the entire breed. Pet dogs are rarely scored. Because of the expense, even a show breeder may only score the animals selected especially for further breeding use. This may mean that animals selected for hip scoring may be exceptionally good examples and therefore less likely to have high hip scores. This tendency may be amplified by experienced breeders having the dog radiographed and choosing not to submit poor examples for scoring. Furthermore experienced breeders seem to favour certain radiographers who “produce good results”, whatever that may mean. Regrettably there are no statistics on such practices, for which there is ample anecdotal evidence. It is difficult to prove these suggestions, but in the course of studies I did find that there is a strong tendency for higher average hip scores to be associated with lower numbers of animals in the breed
16. These factors would all help to push down the statistics, which would include a bias which would not really reflect the underlying breed mean score. It would be helpful if the KC would insist that all radiographs were submitted for scoring, not just those selected by the breeder. It would also be normal to have routine checks on the accuracy of the scores – in the USA the OFA ensures that a selection of scores are re-examined by another panel and publishes the summary results. Radiographers are also “health checked” to ensure they follow procedures. Similar schemes in the UK would help eliminate doubts over the process. Until such measures are introduced any evidence of year-on-year decreases in a breed should be treated with great caution.

17. Having said that, it should be pointed out that in any case the breed mean score, the arithmetic average of all scores, is a very poor way of describing the general breed characteristics, because the small number of very high scores distort this arithmetic average. From a statistical perspective a better measure would be either the mode (the most frequently observed score) or the median (half the scores above this value, half below), or even better, the percentage of dogs scoring more than 25.

WHAT IS NORMAL?

18. The use of the word “normal” in a statistics paper is significant and usually has a specific meaning, describing the way random samples of almost any measurement are distributed around a mean value. A glance at the distribution curves for hip scores shows a classic “normal” curve for scores up to about 25 and then a departure from the established idea of normality above that value. The inescapable conclusion is that dogs with scores above 25 are not “normal” in some sense and have a genetic condition that somehow predisposes them to high hip scores. Is this something they “catch” like a virus, or do all dogs have the same genetic make up which is somehow not activated in 85% of case? The converse must also be true; dogs with scores less than about 25 have to be classed as “normal”, with scores depending on natural variation between animals. So far my remarks have been limited to the statistical analyses of the three selected gundog breeds with some support from the USA statistics. The results across all breeds are so similar that the effect of environment must be very limited and there can be no other conclusion than that genetics play a large part in hip scores, but if so how can they appear to be so random?

THE GENETIC CONNECTION

19. In 2009 I started to study basic genetic theory to try to determine if this could explain the conundrum. Of course, as soon as I started excitedly to propound my latest findings I was told that I was now totally ignoring environmental factors! Well there is no simple answer to that, except by keeping a huge kennel and subjecting the dogs to different regimes over a long period of time and many generations. I am not that interested! But I was surprised to find out how well a simple genetic model using classical Mendelian¹ theory could explain the observed near random distribution of observed progeny scores. I have to point out that modern geneticists have gone well beyond the classical Mendelian approach, but I am not a biological scientist so bear with me for this very simplistic treatment.
20. 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 (a low hip score in this example), whereas both genes have to be Recessive for the animal to show the recessive characteristic (a high hip score). A standard terminology is to use upper case for

¹ Gregor 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.

dominant genes, lower case for recessive genes. Thus AA is a dominant pair, aa is a recessive pair. There are four possible ways in which a single gene pair can be inherited, with three of them giving rise to the dominant trait (good hips say). These three pairs are normally written as AA, Aa and aA. The last two show the characteristics of good hips but are carriers of the version associated with bad hips. The final combination is the fully recessive gene pair aa. This is the only variant of the four that shows the recessive characteristic (bad hips).

21. A first analysis suggests that one or two relevant gene pairs in a dog are insufficient to describe the variations that occur in hip scores. A minimum of three gene pairs may be enough to describe the genetic basis of the observed hip statistics. We know nothing about these three genes of course, but consideration of the way in which they might combine and interact is instructive. The way to do this is well documented in basic genetic texts, and it is not necessary to go into details here. If one writes down all of the possible genetic permutations for three genes, A, B, C, starting with AABBCC and ending with aabbcc one finds a total of 64 possible variations.
22. If these 64 possible variations are examined in detail, three groups emerge which should have well defined outcomes.

- a) The first group consists of examples where all three gene pairs are dominant (one example, AABBCC) and 26 examples where there is at most one recessive gene in each pair (AaBBCC, aAbBcC, ...). According to genetic theory these 27 examples show the general characteristics of Good Hips, and there is some evidence to suggest that these correspond approximately to scores in the range 0 – 18 (OFA grades of Excellent, Good and Fair).
- b) The next major group includes those with exactly one recessive pair (27 examples, each including aa or bb or cc). If a full complement of three gene-pairs are needed to produce fully dysplastic hips then a single recessive pair might be regarded as a transitional group, or “borderline” in the OFA terminology, corresponding to hip scores in the range 19 – 25.

Notice that these two groups, with 54 examples, represent 84% of the total. This corresponds very closely to what we find in practice, with the same percentage of dogs overall falling into the “normal” category of hip scores.

- c) Finally, at the other end of the scale, the group consists of those gene pairs which are either fully recessive (one example aabbcc) or have at least two fully recessive pairs (9 examples, of type aa+bb, aa+cc and bb+cc). Dogs with two or three recessive pairs would be expected to have bad hips with hip scores over 25. The ten possibilities in this group represent 15.6% of the full list of 64. Now recall that in the USA the OFA lists the percentage of Dysplastic dogs in each breed. If an average is taken across all 153 breeds we find that the percentage of Dysplastic dogs is 15.3%. Now to me 15.6% (genetic theory) and 15.3% (OFA statistics) are pretty close! Certainly close enough to suggest that the triple-gene theory is at least plausible.
23. It turns out that one can take this theoretical excursion a little bit further and work out what happens in various classes of matings. Here it should be noticed that although group (a) is

associated with good hips, 26 of the 27 possibilities included hidden carriers for high hip scores. Only one out of 64 possibilities is entirely free of the “high score” gene, all the others can give rise to high scoring progeny. As an example, suppose a dog with genetic make-up AaBbCc is mated to a bitch having the same genetic make up, AaBbCc. Both should have low hip scores but there is no way of knowing the precise configuration of genes. Now, each parent contributes one gene to each of the gene pairs in the progeny, which obviously could end up with a configuration AABBCC (good hips) or aabbcc (bad hips), or anything in between. In other words, mating good hips with good hips is no guarantee of progeny with good hips, a theoretical result confirmed in practice as is well known.

At the other extreme, if a dysplastic dog with genetic makeup of aabbCC (say) is mated to a normal bitch with a score in the range 0-18, with genetic makeup AaBbCc then the progeny could have a genetic pattern AaBbCC, in other words “normal”, in group (a). This helps to explain why high scoring parents can give rise to normal progeny.

24. This simple genetic model confirms what is obvious in practice, that selective breeding based on hip scores does not eliminate high hip scores in progeny but can reduce to the odds on high hip scores.
25. To summarise, there are now three pieces of information that suggests that the magic number of 25 is what separates the good from the bad. These are:

- a. There is a discontinuity in the observed distributions of hip scores at around 25; below that value the shape of the distribution looks like a classic “Normal” or “Gaussian” distribution. Above that number the distribution of scores is more random.
- b. The *Orthopedic Foundation for Animals* in the USA suggests the equivalent of the same value as the threshold for dysplasia, meaning that their experts support this view.

(I have not been able to find any KC/BVA advice as to which hip scores define the onset of hip dysplasia. It would be helpful if they would. The impression I received was that a score of zero was excellent, but with a score of 1 or higher the poor animal suffered from an incurable complaint. This cannot be the case! It seems far more reasonable to regard the variety of scores from 0 – 18 as normal variations amongst healthy animals.)

- c. The simple genetic model suggests that 15% of dogs should be dysplastic, and the OFA statistics show that about 15% are indeed dysplastic, with scores higher than the equivalent of 25.

Of course, this is only meant to be applicable in the statistical sense across all breeds. There are variations within breeds and in different periods, especially where it has been possible to reduce hip scores through the use of dogs available in a large gene pool.

CONCLUSIONS

26. Genetic theory suggests that it would be impossible to eliminate the possibility of hip dysplasia entirely through selective breeding. Even if all dogs would be scored, the relevant genes are so prevalent that reduction in hip scores would be extremely difficult and would have a severe negative effect on the gene-pool in many breeds, as has been demonstrated in practice. Theory and observation are in good agreement and there is a rational explanation for the apparently random nature of hip scores. Now all we have to do is to work out a proper statistical approach to hip scoring in general.
27. As a first approach ALL dogs should be scored as a condition of registration. Dr Willis has recently been quoted in a similar conclusion, so we do not disagree about this. Alternatively a truly random sample of all dogs should be scored. If this is not done the statistics we have already are not very reliable or helpful. In either case, the statistical summary provided by the BVA should be modified to provide information that is more useful than those listed at present (which would not be the first choice of a statistician), and should include the percentage of dogs in a breed with hip scores greater than 25.
28. The good news is that there is little apparent or documented correlation of hip scores with the well-being and health of the individual animal, so perhaps we need not bother at all! It would be helpful if the KC would devote some of the fee used for hip scoring to make a well publicized study of the relationship between high hip scores and later health. It also means that when making breeding decisions a simple number (the hip score), describing just one of a thousand characteristics of a dog, is no substitute for the human eye and a deep understanding of the canine condition. Angela was right after all!
29. Breeding advice? Well it seems obvious that hip scores should NOT be the primary consideration. There is ample evidence to show that high scoring parents can have low scoring progeny, and vice versa so it is pointless to try to take into account one high hip score in one of the parents. Instead, choose sound stock on both sides and try to match deficiencies on one side with corresponding positive attributes on the other, all shown over several generations, and always with regard to the breed standard. Search diligently for any of the less obvious faults such as skin problems.
30. Finally, have a look at the hip scores of as many close relatives of the shortlisted stock as possible. In general 15%, or 1 in 7 dogs (depending on breed), would naturally have a score greater than 25, so do not be surprised to find such examples. There are no further discontinuities in the distributions above 25, so in terms of heritability or genetic characteristics these considerations apply to any total score beyond 25. If you find a significantly greater proportion than 1 in 7 (it depends on the breed) then you might have cause for concern, but with 1 in 7 or better, hey ho, that's normal. Ignore!

John Morgan
BSc Hons (London)
May 2010