Distribution of Pulls: Years 1996 Through 2006

By Truman Prevatt

While the data comes from official AERC records – the analysis and conclusions below are not officially sanctioned by the AERC.

1. Introduction: In this effort the trends in the distribution of pulls is undertaken to cover rides from 1996 through 2006. This expands on an effort done earlier that covered 1996 through 2003. However, with the inclusion of RO-L/M and three more years it was time to update the results. The data covers over 210,000 records. The data are analyzed in several different ways, but in general the four categories of interest are Limited Distance (LD), S50 (single day 50 to 95 mile rides), P50 (Pioneer or multiday 50 to 95 mile rides), one Day 100 mile rides and two day 100 mile rides. Little analysis is performed on the 2 Day 100 mile rides since they have for the most part have numbers that are too small to support statistically significant conclusions.

In the LD and S50, comparison is performed on both the members and non-members. This was not done P50 and 1 Day 100, again because the numbers were too small for statistically valid results.

There was a change in the 2002 season when the Rider Option (RO) pull code was changed. Two new pull codes were added: RO-L [RO-M] to specify that although the rider chose to pull the horse and the horse was fit to continue the rider pulled because of a rider perceived lameness [metabolic] issue. In general the addition of these pull codes should not change the distribution of the lame (L) and metabolic (M) pulls. The effect should be to spread the RO options over three subcategories, RO, RO-L and RO-M. The reason for this was to isolate the horses that were being pulled by the rider to a specific reason for the pull. Namely if the rider was sick, injured, etc. then it was an RO pull. If on the other hand if the rider had some concern about the horse but decided to pull although the vet could not pull the horse – then the RO-M/L pull code would signify that pull.

Two new codes were also added that year. These were Surface Factors (SF) and Disqualified (DQ). It turns out that both these pull factors have proved to be insignificant in the data for this analysis since for the most part since combined they seldom account for percent of the pulls.

The most interesting and important information is related to the pulls that fall in the categories:

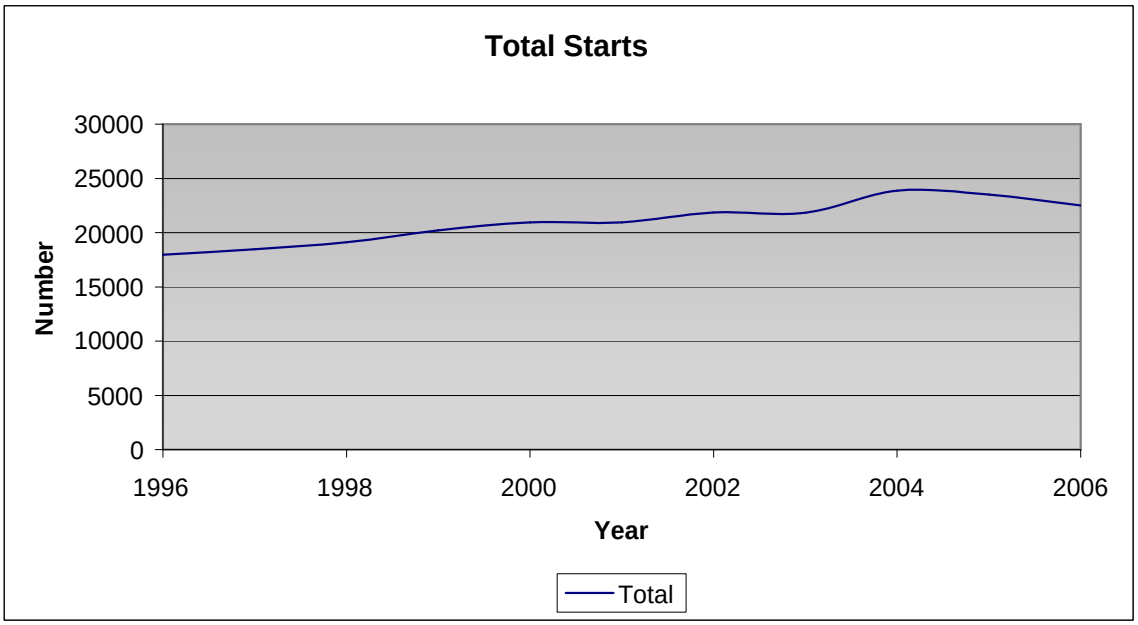
- Metabolic – not fit to continue the ride for metabolic reasons,
- Lame - not fit to continue the ride for lameness reasons,
- RO – Rider Option – horse fit to continue,
- OT – Overtime.

In 2002 the RO category was broken into three:

- RO – Rider pulls for reasons associated only with the rider,
- RO-L - Rider pulls (horse fit to continue) but associated with lameness concerns,
- RO-M - Rider pulls (horse fit to continue) but associated with metabolic concerns.

These have been analyzed in several ways in several different categories and sub categories that will be discussed in subsequent sections.

2. Trends In and Start Demographics: Figure 2.1 shows the overall trends in starts over the past 11 years.



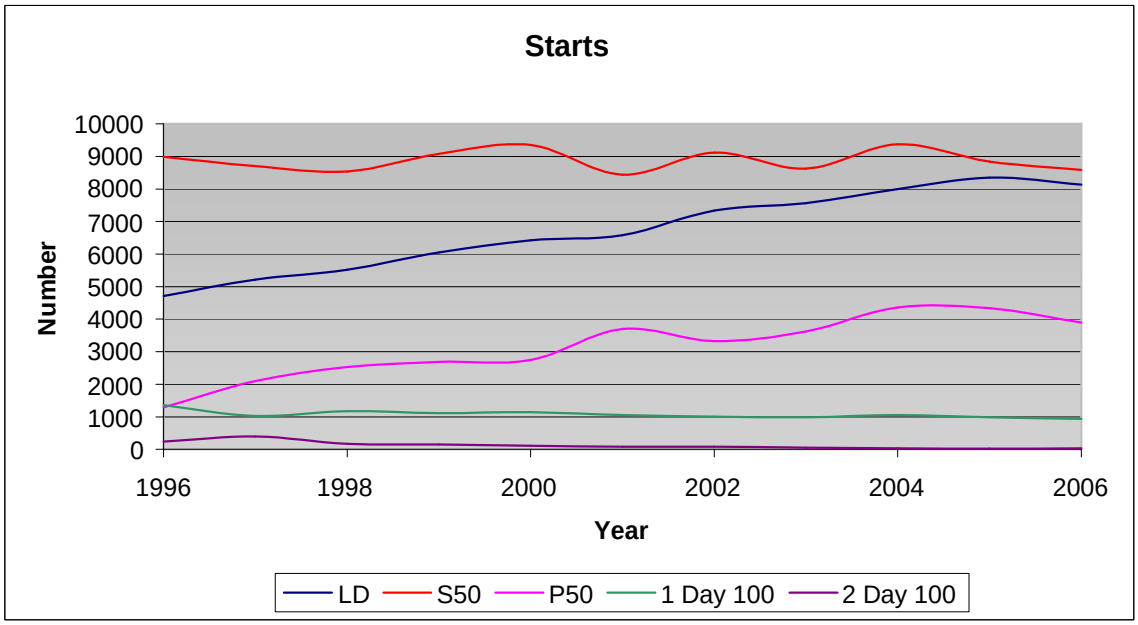
Total Starts
Figure 2.1

There was fairly rapid growth from 1996 through 2004. The past two years have been flat. There is probably a multitude of reasons for that which is out scope of this study. In Figure 2.2, the starts are broken out to the major ride categories listed above.

As has been well documented, the LD and Pioneer rides have grown while the single day 50 mile rides have remained flat. The 2 Day 100 has practically disappeared. The one day 100 has been experiencing a steady decline since 1997 (the peak in the period of the current database) with 2006 about 5% down over 2005 and 2005 about 5% down over 2004. It is clear from this chart that LD and multiday starts account for the growth in the total starts that the AERC has experienced over the past 11 years. The fastest growth has been in the LD's.

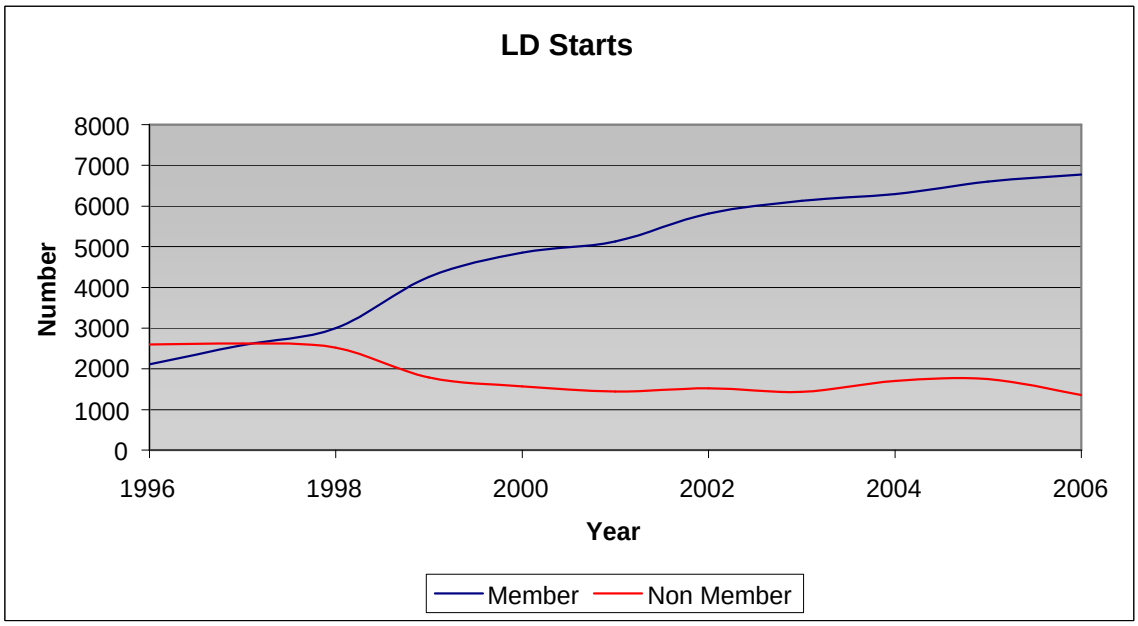
The 2 Day 100 format was popular in the early part of the period because it could be used for ROC qualification. The last ROC was 1997. There were over 550 2 Day 100 in '96

and '97 combined. The last few years the 2 Day 100 has accounted for 20 to 30 starts each year.

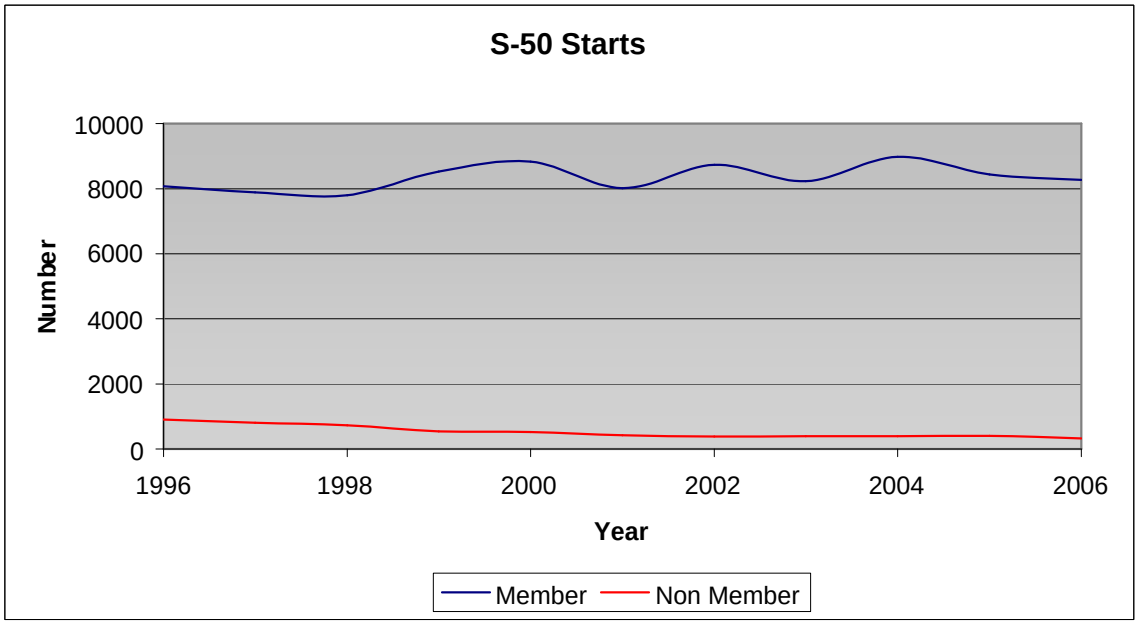


Starts Per Ride Category
Figure 2.2

Referring to Figures 2.3 and 2.4 the member starts and non-member starts for the LD (Figure 2.3) and S50 (Figure2.4) are shown. There were not a sufficient number of non-member starts in the Pioneer or 1 Day 100 categories to provide a reliable statistical base. The P50 had less than 40 to 50 non-member starts per year and the 1 Day 100 had on the order of 10 to 20 non-members per year.



LD Starts Member vs. Non-Member
Figure 2.3

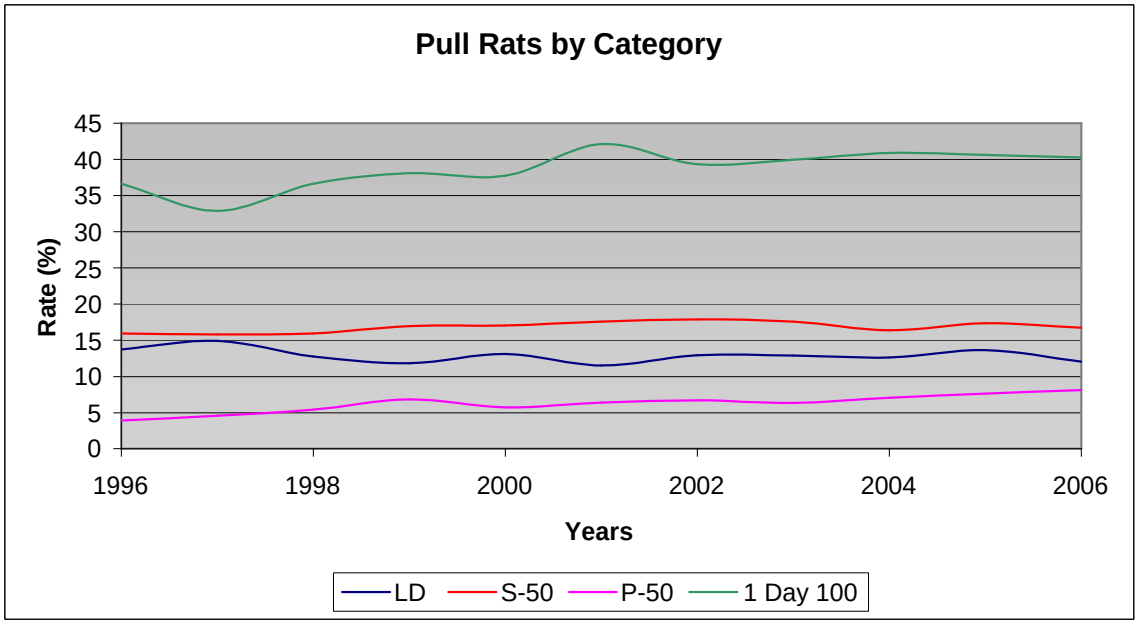


Single Day 50 Starts Member vs. Non-Member
Figure 2.4

What is clear is the demographics of the LD have changed significantly over the past 11 years. In 1996 more starters were non-member (2700) than member (2000). That rapidly changed as the member starts (6800) rapidly surpassed the non-member starts (1400) by 2006. The largest number of non-member starts is in the LD – which would be expected since that is the entry point for most new riders.

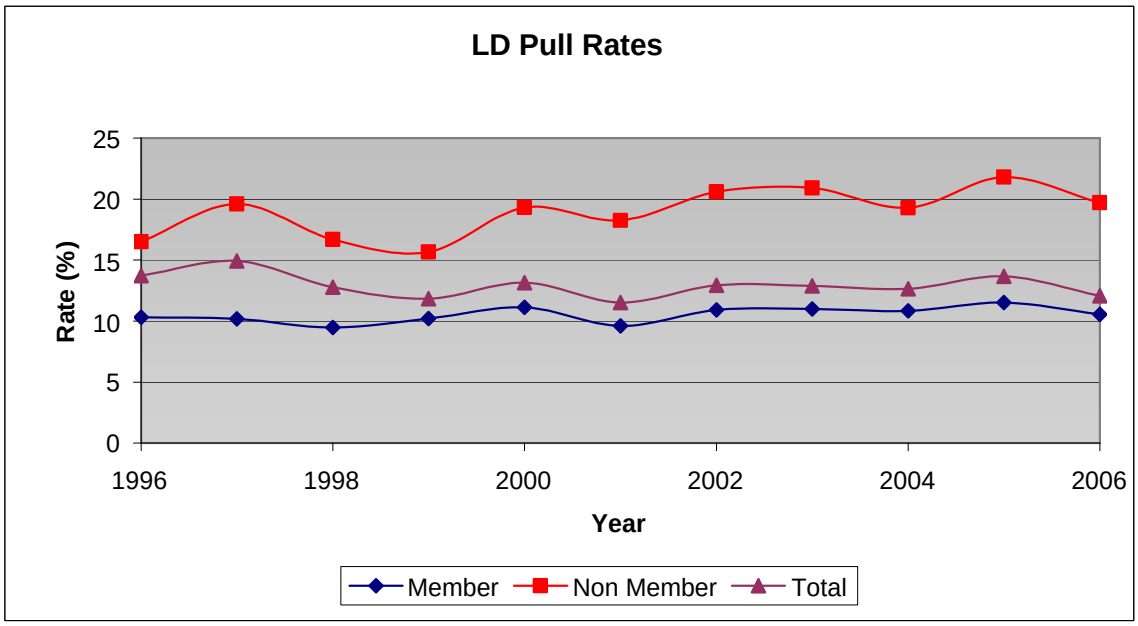
What is interesting is there is a significant number of non-member starts in the S50. To be sure it is not the same level as for the LD but significant none the less. For both cases there are sufficient non-member starts to compare the pull rates between members and not members. This will be discussed in the next section.

3. Pull Rates: As one would expect the pull rates are different for different events. The overall pull rate is shown in Figure 3.1. It is not surprising that the pull rate of the 1 Day 100 is 2.5 times that of a S50 and 3 times that of an LD. What is a little surprising is the pull rate for a P50 is less than the LD – in fact significantly less. The overall pull rate in the P50 has undergone steady increase. While the seeming discrepancy in the pull rate of the Pioneer rides would make an interesting topic, it is not in scope of this study.

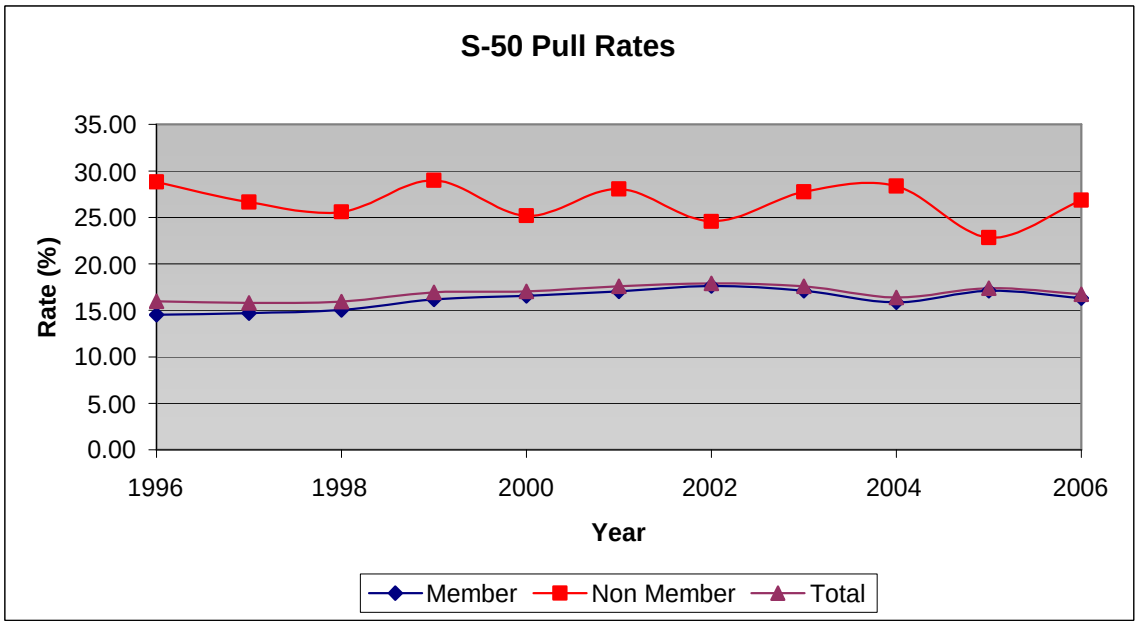


Overall Pull Rate
Figure 3.1

In Figures 3.2 and 3.3 the pull rates for the LD and single day 50 are broken down into total, member and non-member. The interesting thing to notice is in both the LD and single day 50 the pull rate of non-members is almost twice that of the members- with the single day 50 non-member pull rate hovering between 25% to 30% for the non-member starts. That is one in every four non-member starter in a 50 will not finish. For the LD the non-member pull rate has grown and is hovering to 20% or one in five non-member LD starters will not finish.

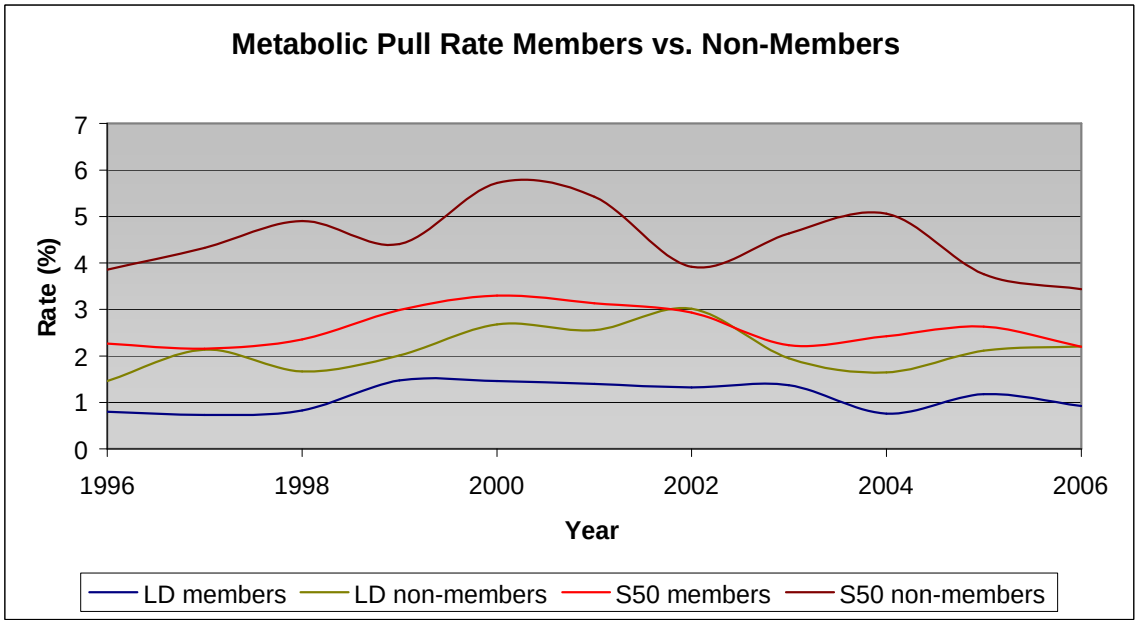


Pull Rates LD Member vs. Non Member
Figure 3.2



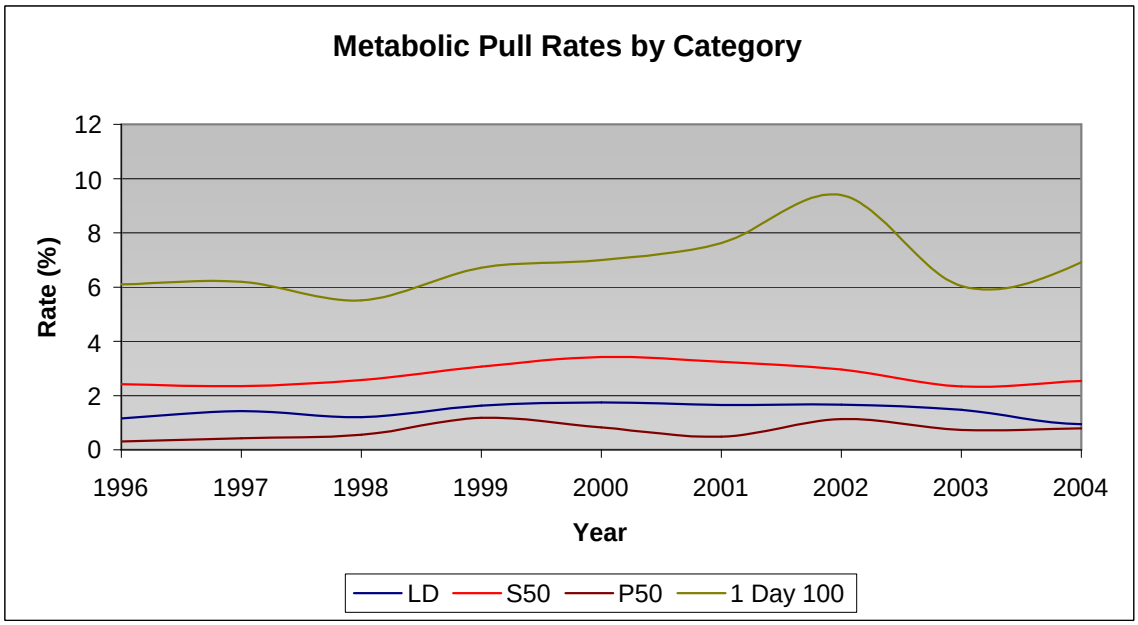
Single Day 50 Pull Rates Member vs. Non-Member
Figure 3.3

Figure 3.4 is the member and non-member metabolic pull rates. Again a non-member will have a metabolic pull at twice the rate of the members. There is another interesting thing that seems to be showing (which will be discussed later). There seems to be a change in the distributions (drop in the metabolic pull rates) the last four years of the period.



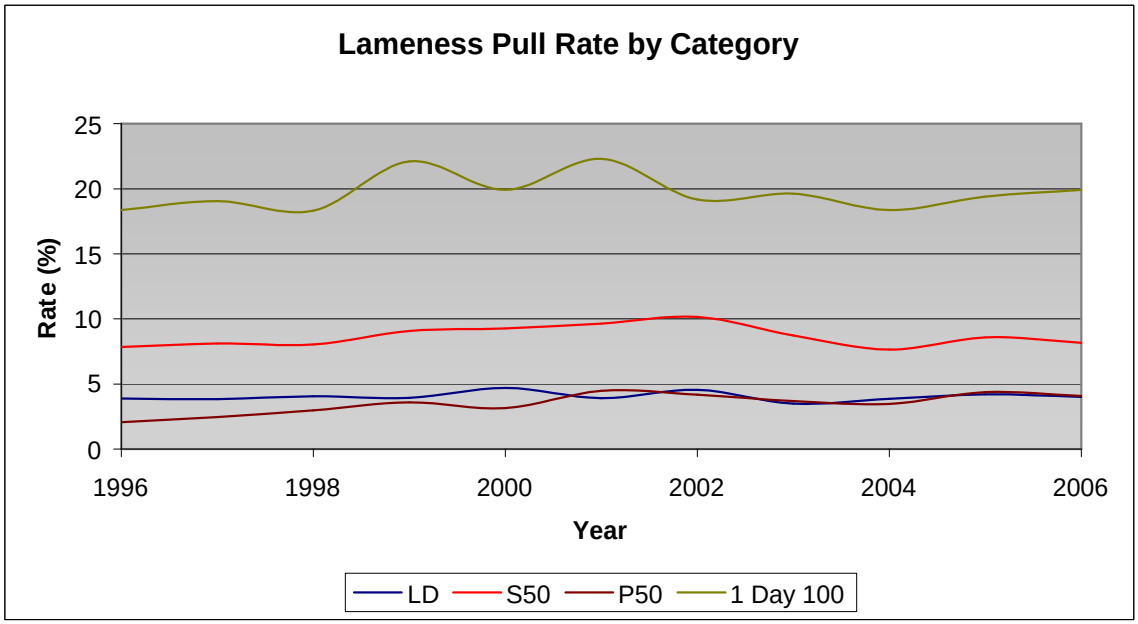
Metabolic Pull Rates
Figure 3.4

Figure 3.5 is the metabolic pull rates for the ride categories. The 1 Day 100 is running about 2 times that of the single day 50 which is the next highest. The Pioneer 50 rides has the lowest, however, they seem to have caught up to the LD rides. Again there seems to be the same drop in the metabolic pulls in the last four years.



Metabolic Pulls by Ride Category
Figure 3.5

Figure 3.6 shows the lameness pulls by ride categories. No surprises here – the 100's are about 2.5 times higher followed by the 50's. The LD and Pioneer lameness pulls are running at about the same rate.

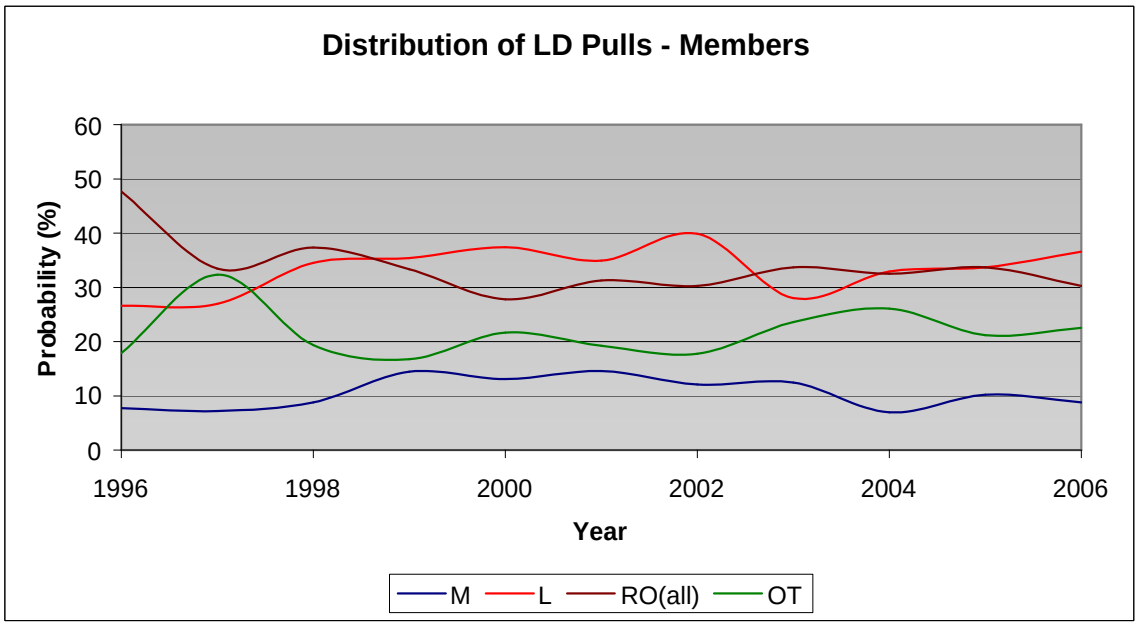


Lameness Pulls by Ride Category
Figure 3.6

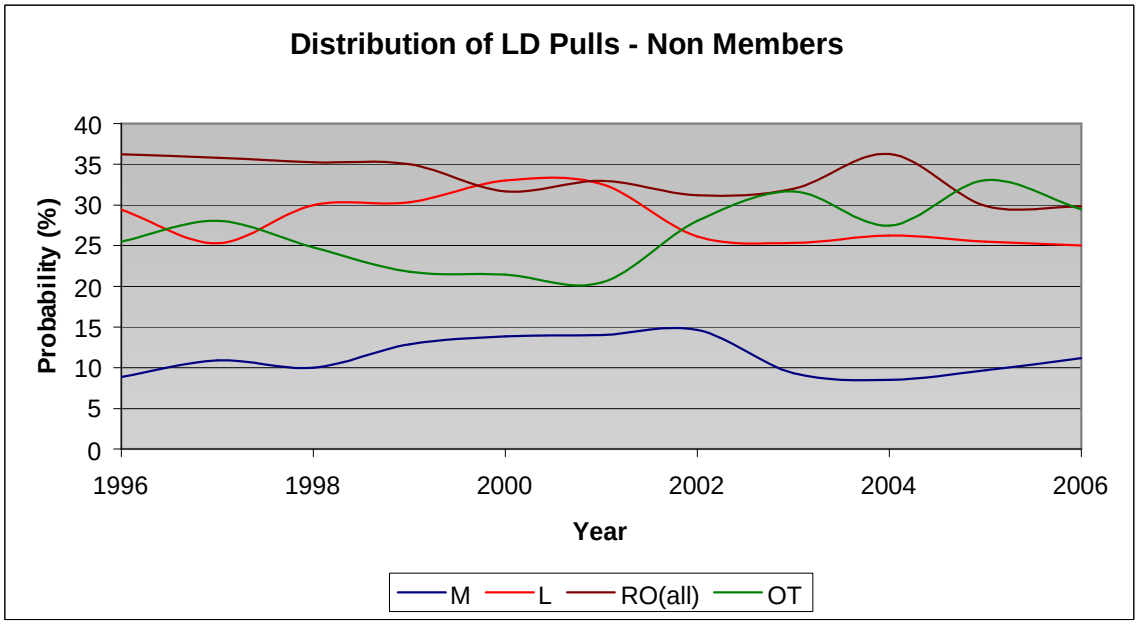
4. Distribution of Pulls: More interesting than the raw pull rate is the conditional probabilities associated with the pulls. That is given a pull, what is the probability that pull will be for a given factor. There is a simple transformation between the raw pull rate for a factor and the distribution for those that want, however, the distribution of the pulls will be considered for the remainder of this study. For this section for the last four years all the RO pull subcategories are combined to RO so it can be compared to the previous years. This is valid if the hypothesis on RO pulls as stated in the introduction is valid.

In Figures 4.1 and 4.2 the distribution of LD pulls for member and non-member are broken out. What is clear is (as will be seen later) the LD pulls are significantly different than the longer distances. The big difference is OT (over time) pulls. The LD OT pulls for members is around 20% over the period. For the non-members it has ramped up to approximately 30% to become the competitive (with lameness and RO) as the most likely reason for a pull for the non-member in the LD.

For the members the most likely pull is either RO or lameness with OT third. In both cases metabolic pulls are the least likely.



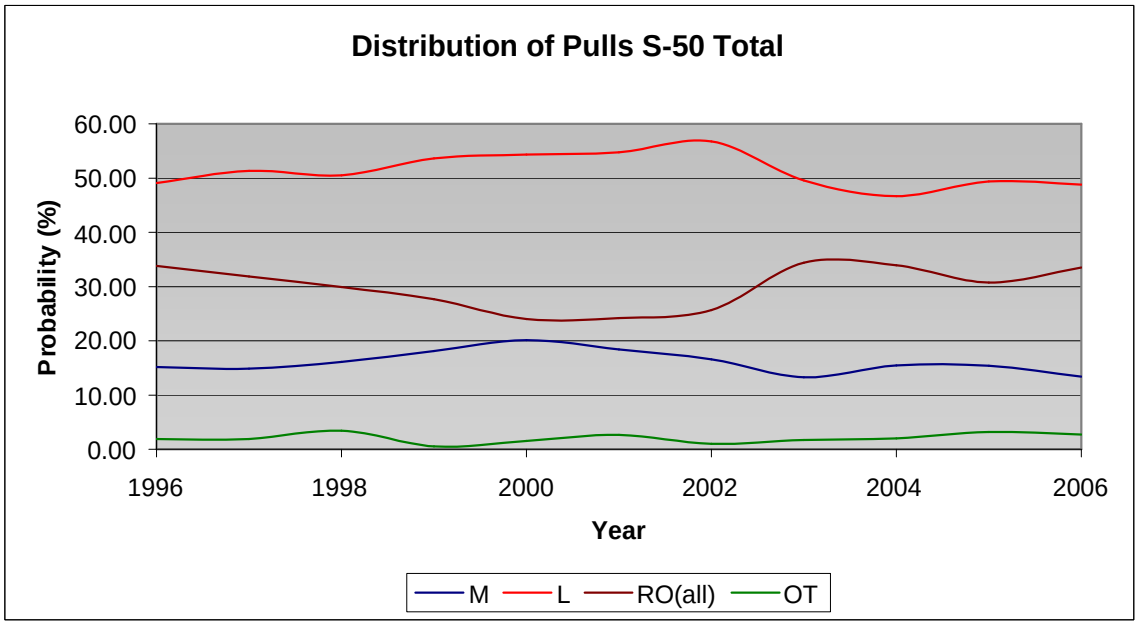
Pull Distribution LD Members
Figure 4.1



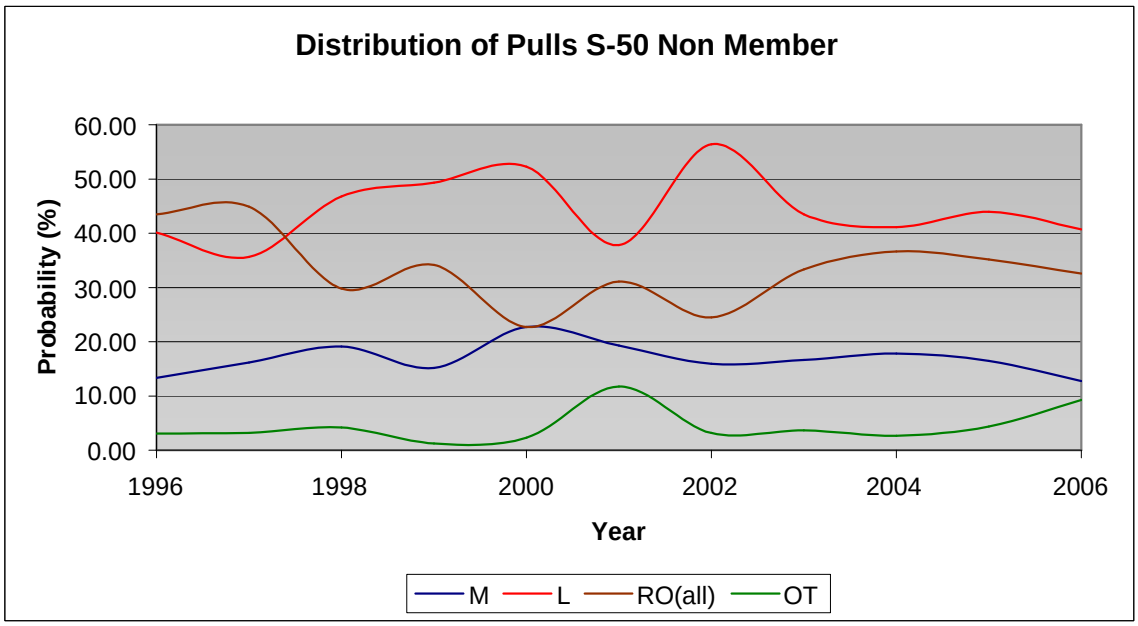
Pull Distribution LD Non-Members
Figure 4.2

For the single day 50, since the member pulls are so close to the total pulls the distribution of the total pulls (Figure 4.3) and non-member pulls (Figure 4.4) are shown. What is clear in Figure 4.3 is OT is not a big factor in 50 mile rides. While it is higher for non-members it is no-where the factor for pulls as in the LD. While the distributions are somewhat different the ranking for pulls goes from Lameness – first, RO- second,

Metabolic – third and OT – fourth. The probability of a metabolic pull is higher for the non-member than the member.

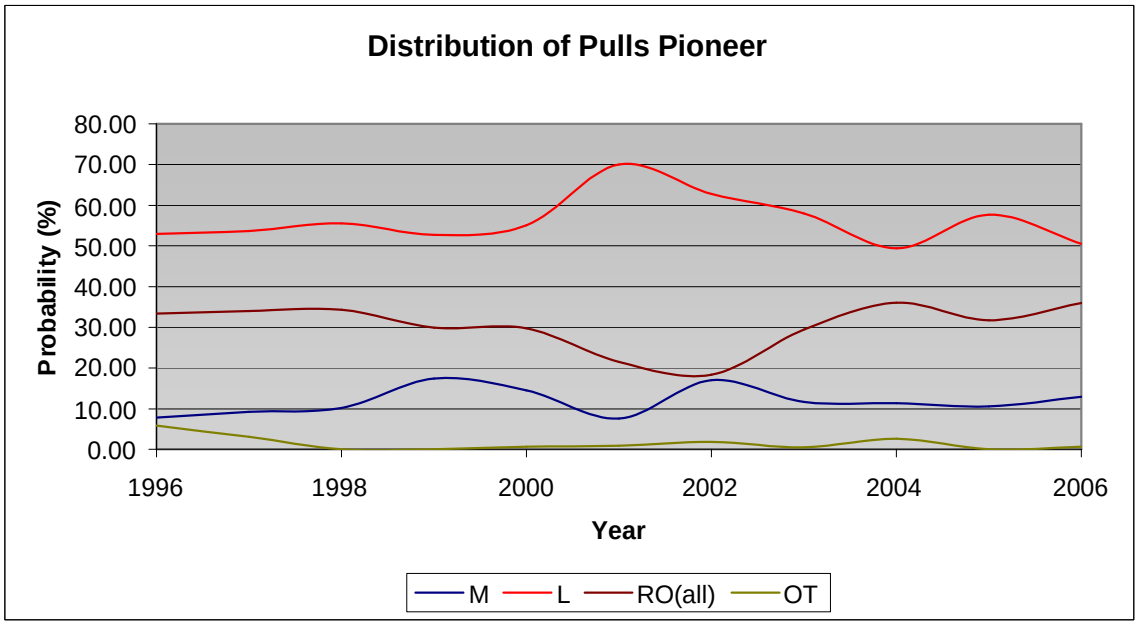


Pull Distribution Single Day 50 Total
Figure 4.3

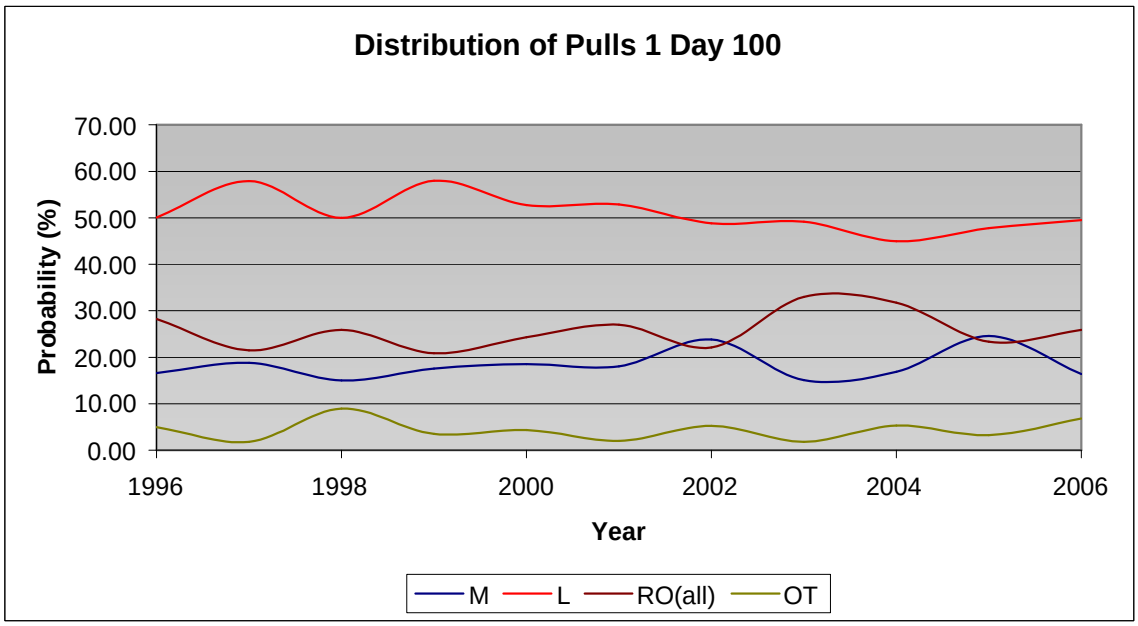


Pull Distribution Single Day 50 Non-Member
Figure 4.4

In Figure 4.5 the distribution of the pulls for a Pioneer (multiday) ride is shown. In the multiday ride the probability of lameness seems to be higher. Figure 4.6 is the distribution of pulls for the 1 Day 100 mile ride.



Pull Distribution Pioneer Rides
Figure 4.5



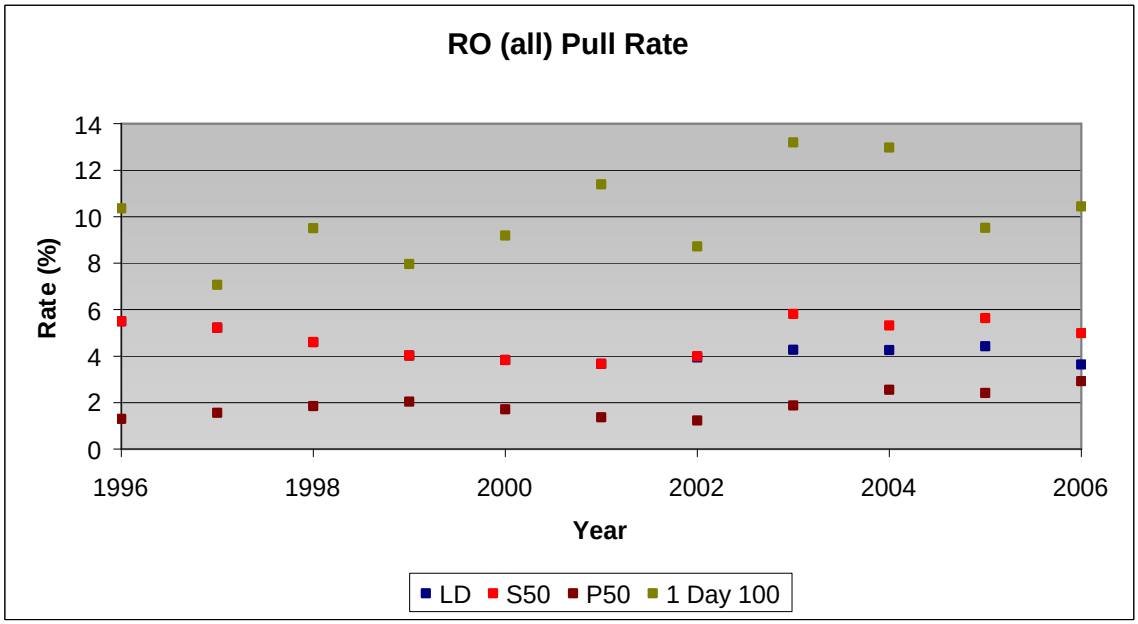
Pull Distribution 1 Day 100
Figure 4.6

The distribution of the pulls for the 1 Day 100 are similar – except the OT pulls are a little higher. The Tevis and OD together could account for that.

Across the board – except for the LD - the most likely reason for pull is lameness followed by RO (combined RO 2002 on) then metabolic and the OT. The LD is significantly different as shown in Figures 4.1 and 4.2.

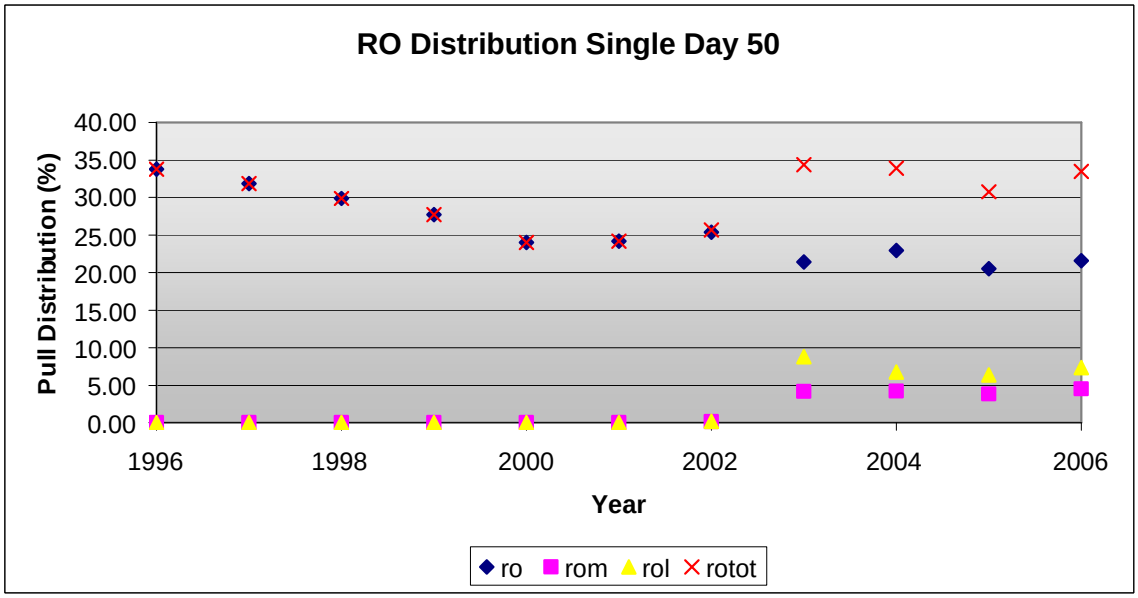
5. The RO Conundrum: As stated earlier there was a change in the 2002 season when the Rider Option (RO) pull code was changed. Two new pull codes were added: RO-L [RO-M] to specify that although the rider chose to pull the horse and the horse was fit to continue the rider pulled because of a perceived lameness [metabolic] issue.

The addition of these pull codes should not change the distribution of the lame (L) and metabolic (M) pulls. The effect should be to spread the RO options over three subcategories, RO, RO-L and RO-M. This was the working hypothesis of this study and I believe the working hypothesis of the BOD when these changes went into effect. Figure 5.1 shows the RO pulls (where all RO pulls were combined for the last four years) for each ride sub category. Given the hypothesis above the pull rate should be more or less consistent across the board. However, there appears to be an increase starting about 2003 in most ride categories. That is there seems to be a change in the underlying population distribution of pulls in the last four years compared to the previous (say) four years.



Rider Option Pull Rates
Figure 5.1

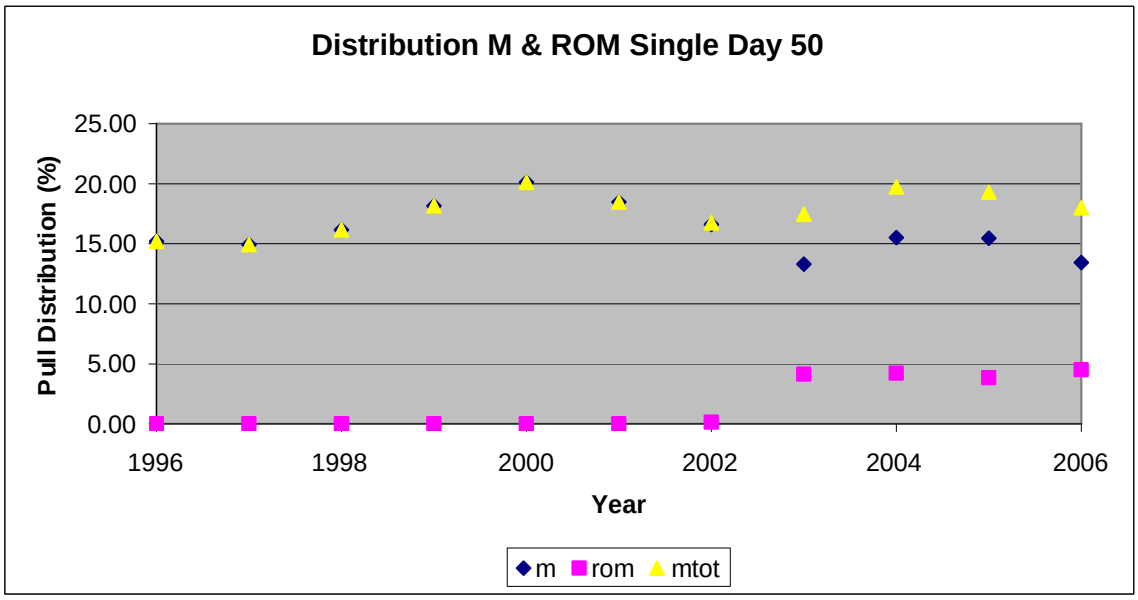
To explore further we look at the distribution of pulls with RO, RO-L/M broken out over the span of the 11 years. To best see the trends the distribution (namely the probability of a given pull given the horse was pulled) is used. In figure 5.2 the distribution of RO for single day 50 mile rides is shown. Here RO is the RO category (in the last four years rider only). This is a good ride category to analyze since it has been pretty constant at approximately 8500 starts per year over the 11 year period.



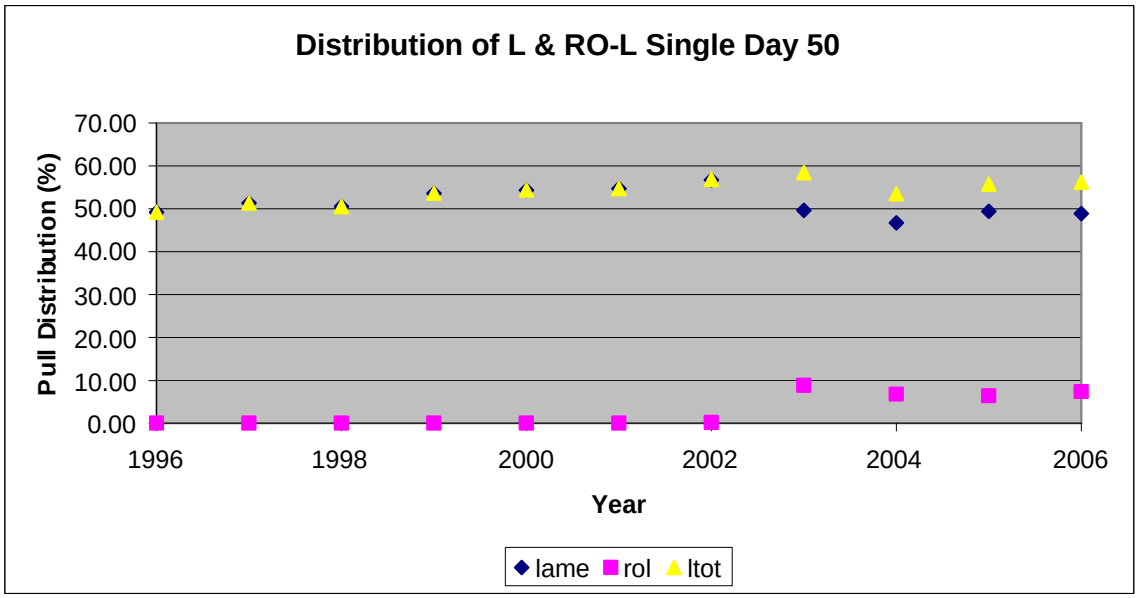
Distribution of RO
Figure 5.2

The trends are quite interesting. During the early part of the period the RO pulls made up about 35% of the pulls. With the effort of education and the desire to get a more accurate picture of why our horses don't finish the RO pull steadily declined to 2000 and from 2000 through 2003 were flat. Then comes the introduction of RO-M/L and the total RO (including RO-M/L) percent of the pulls went back to 1996 level and has seemed to stay there. However, the true RO (rider) has dropped in the last four years.

Consider to Figures 5.3 where the M, RO-M and the sum of M and RO-M proportions are plotted. What has happened in metabolic is a drop in the metabolic categories with some presumably becoming RO-M with the combined M and RO-M distribution being slightly greater after the introduction of RO-M than prior. The same hold true for the RO-L, Figure 5.4.



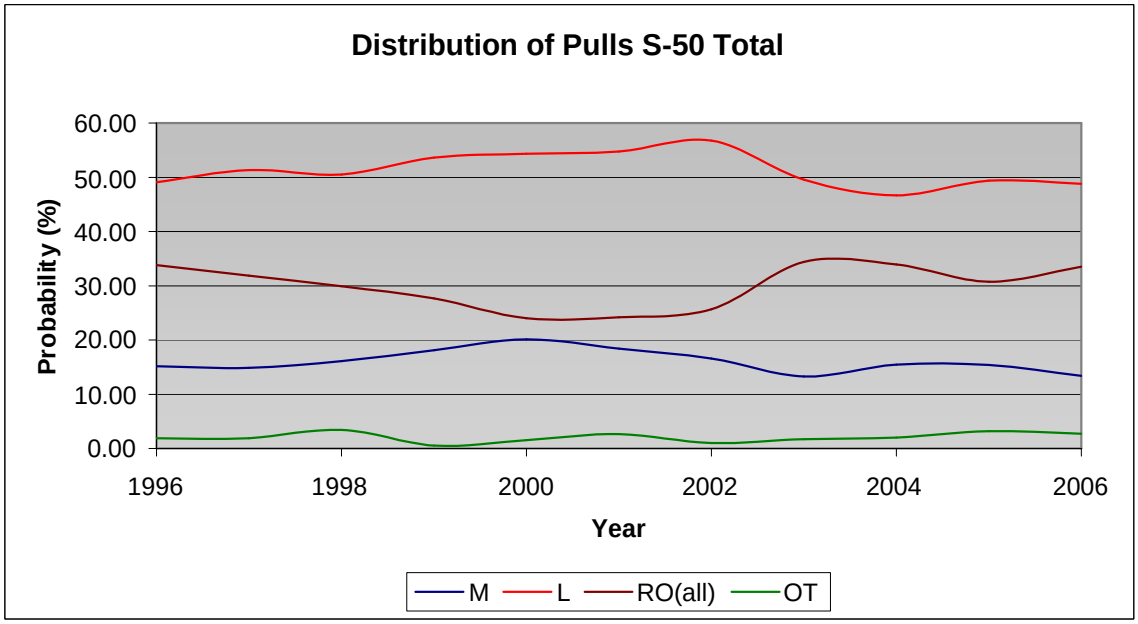
Distribution of M and RO-M
Figure 5.3



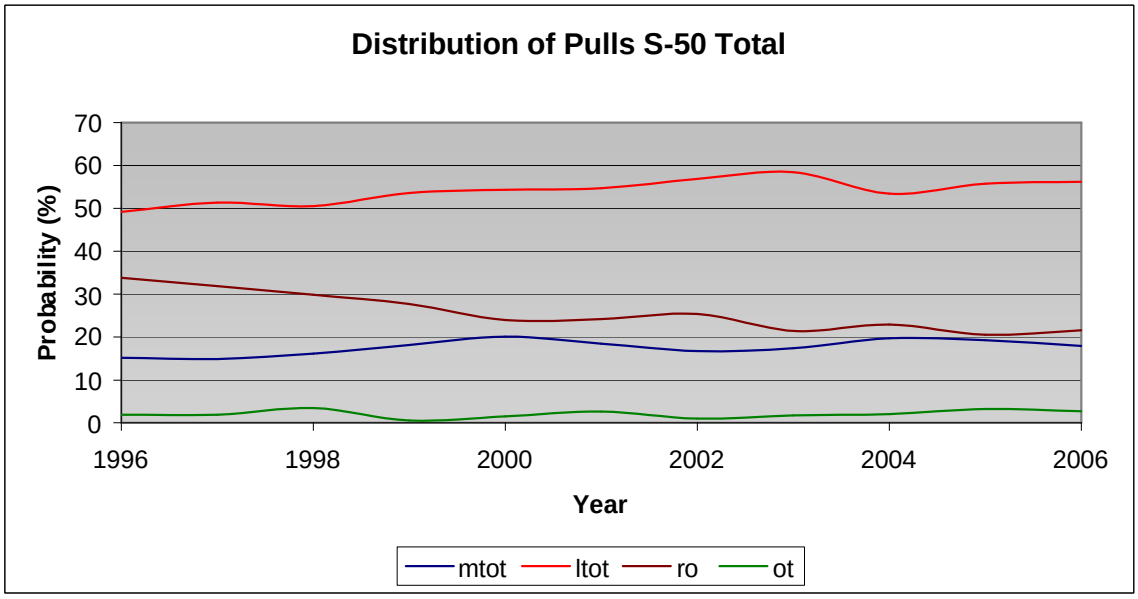
Distribution of Lame and RO-L
Figure 5.4

The data in the previous sections was presented by combining all the RO pulls after 2002. That is the RO categories after 2002 consisted of RO plus RO-M plus RO-L. However, the data presented in Figures 5.2, 5.3 and 5.4 indicate that that some of what might have been M/L pulls prior were being classified as RO-M/L. In Figure 5.6 the data in Figure 4.3 is represented with the RO-M, RO-L combined with the M, L pulls respectively. This hypothesis is that RO-M is a metabolic pull and RO-L is a lameness pull which up till 2003 had been "hidden" as a RO pull and that the pull distribution of metabolic and

lameness pulls had been under estimated in the past. Below Figure 4.3 is repeated as Figure 5.5 for reference.



Distribution of Pulls Combining RO-M/L with RO
Figure 5.5



Distribution of Pulls Combining RO-M/L with M/L
Figure 5.6

Figure 5.6 tends to give a better picture of the pull distribution. That is when RO-L is grouped with Lame there is a bump up in 2003 but before and after it seems stable. In

fact except for 2004 the lameness pulls shown in Figure 5.6 is a steady 55% to 57%. There is a similar observation for metabolic in Figure 5.6. After the bump up in 2003 it has been running steady at about 20%. In Figure 5.6 with the RO-M/L removed from the RO category the RO pull distribution has declined to a steady 20% in the end of the period. While the time period in only about 3 years the RO- M/L categories when combined with the M/L categories for calculation of trends tends to give a stable picture of the pulls. That is the lameness pull rate for a single day 50 is a steady 56%, for a metabolic pull it seems to be a steady 20% which is the same for an RO.

6. Observations: There are several things that jump out. While we still have a lot of day members riding (approximately 1,800 day member starts in 21,000 starts or 8.6%) the number of day members is significantly down from 1996 (approximately 3600 day member starts in 18,000 starts or 20. The vast majority of day members are in the LD which is to be expected. In 1996 57% of the LD starts were day member – in 2006 that was down to 17%.

The total starts have been flat after a long period of growth – fueled by growth in LD and 50 mile starts. The 50 mile rides have been transformed into Pioneer rides so that the growth in the 50 mile distance is seen as Pioneer alone with the number of single day 50 starts remaining constant over the period. The 100 mile starts have been steadily declining with 5% declines in 2005 over 2004 and again in 2006 over 2005. The two day 100 is about dead. The LD's and 50 mile rides (primarily Pioneer rides) have made up for all the growth in the sport since 1996.

The day rider is twice as likely to get pulled as the member. While our education of new riders seems to be working – I don't think one "new riders" meeting has much effect on the ride the next day. Maybe our new rider education needs to be the first year so new riders can learn at a progressive rate as they gain more experience in the sport.

An interesting statistic is the LD riders are about 8 times as likely to be pulled for over time as the 50 mile riders and about 6 times as likely as the 100 mile riders. That seems to say a large proportion of the LD riders are taking their time – not pressing their horses.

The overall pull rates have been basically constant over the period – particularly over the last 5 years. That is 40% of the 100 mile riders are pulled. For the 50 that number is 16% pulls and for the LD it is 13%. What that seems to say is pulls are an integral part of the sport and should be taken as such. A pull in a 100, 50 or LD is not a failure since on the average 40, 17 or 13 riders respectively out of every 100 riders starting will end up pulled.

An interesting discussion is "accuracy of the pulls" recorded in the data base. For the most part they seem to be accurate in that for the major categories (metabolic and lameness) the rates and the distribution of these categories in all the pulls have remained amazingly constant over the 11 year period covered by the current AERC data base. No data base will be completely 100% accurate.

Random errors in the data base will not bias the results. The higher the error rate - the more data it takes to draw a conclusion because the more random variability in the data. Random errors do not bias the results. Errors happen and that's fine. However, biasing the data presents a significant issue in interpretation of the data or using it for long term trend analysis.

That is where the results shown in Section 5 are interesting. It appears that despite the best intentions potential lameness/metabolic pulls were being reported as RO. The data presented in Section 5 tends to reject the hypothesis that the combined RO pulls after 2002 were the same as the RO pulls prior (Figure 5.2). Figure 5.3 points to RO-M/L pulls coming from two categories. First some RO pulls went to this category as seen by the drop in RO pulls in 2003 on. The second source was pulls that were either M or L. Some of what would have been M/L become RO-M/L. However, when the distribution is calculate under the assumption that a RO-M pull is because of metabolic issues and similar for an RO-L the distribution do not show the "change" at the boundary year for the introduction of RO-M/L (compare Figure 5.5 and 5.6).

These new categories appear to have taken some of the previous M and L pulls where riders may now be pulling their horses earlier than before. This can be seen in that after these codes were introduced, M and L pulls went down by about the numbers in the RO-M and RO-L codes. They also seem to have changed the RO pulls to where for the most part they are rider related. In many respects the addition and use of RO-L/M seems to be making the database more accurate and more useful for trend analysis.