Duchenne Muscular Dystrophy Data Set


(2) Until 1987, the specific genetic deficiency that causes Duchenne’s Muscular Dystrophy (DMD) was unknown. Given the debilitating nature of this X-linked disease (especially in boys), efforts were made to diagnose mothers who were potential carriers of the genetic defect. Previous work demonstrated that carriers show elevated creatine kinase (CK) activity and elevated serum levels of hemopexin (H). Unfortunately, protocols for diagnosis solely examined CK activity as an indicator of carrier status, which was only capable of distinguishing 65% of carriers from “normals.” Thompson et al. sought to determine whether an assessment of both CK activity and H levels would provide a better indication of carrier status. The study found that H levels alone distinguish only 27% of carriers, but in combination with CK activity distinguish 83% of carriers, allowing the researchers to conclude that the combination of assays is more effective in determining carrier status.

(3) There were no treatment arms, per say, in the study as all women (whether normal, carrier, or suspected-carrier) had serum samples drawn (in some cases, multiple serum samples were drawn from the same woman). H levels and CK activity were measured for all serum samples (in some cases, multiple measurements taken from the same sample).

(4) The study is necessarily observational as the researchers could not impose DMD carrier status on a woman and then proceed to measure CK activity and [H]. The study involved 220 women age 20-40 (inclusive) and some unspecified number of women younger than 20 and another unspecified number over 50 years old. The unspecified numbers of women provided 50 total serum samples. For those individuals aged 20-40, there were 104 non-carriers (164 serum samples), 23 known DMD carriers (45 serum samples), and 93 potential carriers (93 serum samples). The potential carriers consisted of 51 mothers with one affected son and no family history of DMD, and 42 sisters with one affected brother. No information is given as to why the individuals chose to donate blood to the study. No blinding or randomization was mentioned. The study was funded by Muscular Dystrophy Association of Canada, the National Science and Engineering Research Council, and the University of Toronto’s Statistical Consulting Service.

(5) Although the researchers do not explicitly state that any individuals were removed from the study, several “ghosts” do exist, namely a normal who appears (104→105) in Table 1 and nowhere else and an obligate carrier who disappears (23→22) in Figures 2&3.

(6) Variables:
- **table_no**: The table number in a book of tables, in this case table number 38
- **id**: Measurement identification number (1-209)
- **obs**: Number of measurements of taken from a given patient
- **hosp_id**: Patient identification number
- **age**: Patient’s age (in years)
- **month**: The month that a given serum sample was drawn (1-12)
- **year**: The year that a given serum sample was drawn (19XX)
- **CK**: Creatine Kinase activity (international units/mL)*
- **H**: Hemopexin concentration (mg protein/100 mL, mg%)*
- **PK**: Pyruvate Kinase activity (international units/mL)*
- **LD**: Lactate Dehydrogenase activity (international units/mL)*
- **group**: The carrier status of women (“carrier” or “normal”)

*Data are sometimes based on averages of multiple measurements on serum samples from a given patient

(7) It is interesting that Thompson et al. assume that one-third of all DMD cases are the result of spontaneous mutation in a given male. Based on the assumption that the allelic frequency of DMD is constant in a population despite the fact that boys with DMD have decreased fitness, an assumption of spontaneous mutation rate as high as one-third is actually reasonable (see Haldane JBS (1935) The Rate of Spontaneous Mutation Of A Human Gene. *J. Gen.* 31: 317-326).

### Summary Statistics:

<table>
<thead>
<tr>
<th></th>
<th>CK (IU/L)</th>
<th>H (mg%)</th>
<th>PK (IU/L)</th>
<th>LD (IU/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>92.26</td>
<td>84.29</td>
<td>16.07</td>
<td>198.6</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>15</td>
<td>9</td>
<td>2.8</td>
<td>66</td>
</tr>
<tr>
<td><strong>1st Quart.</strong></td>
<td>30</td>
<td>78</td>
<td>10.3</td>
<td>148.2</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>41</td>
<td>86</td>
<td>13.7</td>
<td>177</td>
</tr>
<tr>
<td><strong>3rd Quart.</strong></td>
<td>73</td>
<td>93.2</td>
<td>17.4</td>
<td>231.8</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>1288</td>
<td>118</td>
<td>110</td>
<td>593</td>
</tr>
</tbody>
</table>

Table 2: Mean & Standard Deviation for CK, H, PK, and LD (Separated into Normals and Carriers)-
There were 75 Carriers and 134 Normals.

<table>
<thead>
<tr>
<th></th>
<th>CK (IU/L)</th>
<th>H (mg%)</th>
<th>PK (IU/L)</th>
<th>LD (IU/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean (Normals)</strong></td>
<td>39.13</td>
<td>82.94</td>
<td>12.14</td>
<td>164.57</td>
</tr>
<tr>
<td><strong>Std. Dev. (Normals)</strong></td>
<td>18.32</td>
<td>12.32</td>
<td>4.32</td>
<td>41.37</td>
</tr>
<tr>
<td><strong>Mean (Carriers)</strong></td>
<td>187.19</td>
<td>86.67</td>
<td>23.93</td>
<td>256.19</td>
</tr>
<tr>
<td><strong>Std. Dev. (Carriers)</strong></td>
<td>225.52</td>
<td>23.16</td>
<td>17.21</td>
<td>81.11</td>
</tr>
</tbody>
</table>
Descriptive Graphics:

**Figure 1: Creatine Kinase Activity Separated into Carriers and Normals**

It is interesting to note the scale difference between carriers and normals. While both have the majority of CK activities from 0-200 IU/L, the carriers have a fair portion well above 200 IU/L.

**Figure 2: [Hemopexin] Separated into Carriers and Normals**

It is important to notice that while both carriers and normals have a large number of individuals with [H]>80mg%, almost all Carriers fall within the range, while the normals also have a fair portion with [H]<80mg%.
Inference:

T-tests on the difference between CK activity and [H] between carriers and normals.

CK Activity:
$H_0$: True difference in means is equal to zero
$H_a$: True difference in means (carriers-normals) is greater than zero
$t=5.675, \text{df} = 74.547, p\text{-value}=1.254\times10^{-7}$. We say that if we assume the null hypothesis is true we would only expect to see our data or more extreme $1.254\times10^{-7}$% of the time. Therefore, we can reject the null hypothesis and concluded that the true difference in means is, indeed, greater than zero (i.e. carriers have a greater CK activity)

[H]:
$H_0$: True difference in means is equal to zero
$H_a$: True difference in means (carriers-normals) is greater than zero
$t=1.2962, \text{df}=97.944, p\text{-value}=0.09898$. We say that if we assume the null hypothesis is true we would only expect to see our data or more extreme 9.898% of the time. Therefore, we do not have enough evidence to reject the null hypothesis and so conclude that the true difference in means is equal to zero (i.e. carriers and normals have the same mean [H])