3c) tree20: All clinical data
cp=0.01, endgroup=8, MissClassErr=23.4%

ALC_Lympho >= 20

no — yes

−MRD
31 / 37
18%

vhmut = 1

gender = M

−MRD
15 / 19
9%

WBC >= 125

plat < 125

−MRD
8 / 10
5%

plat >= 94

−MRD
7 / 9
4%

+MRD
11 / 15
7%

+MRD
10 / 14
7%

haem < 10

plat >= 186

WBC >= 62

−MRD
12 / 19
9%

haem < 14

−MRD
8 / 10
5%

+MRD
28 / 40
19%

+MRD
15 / 18
9%

+MRD
15 / 18
9%

Tree has quite a good Missclassification error, but does not make much sense clinically
Note:vhmut=1 −> 98−100% mutated (which is good?

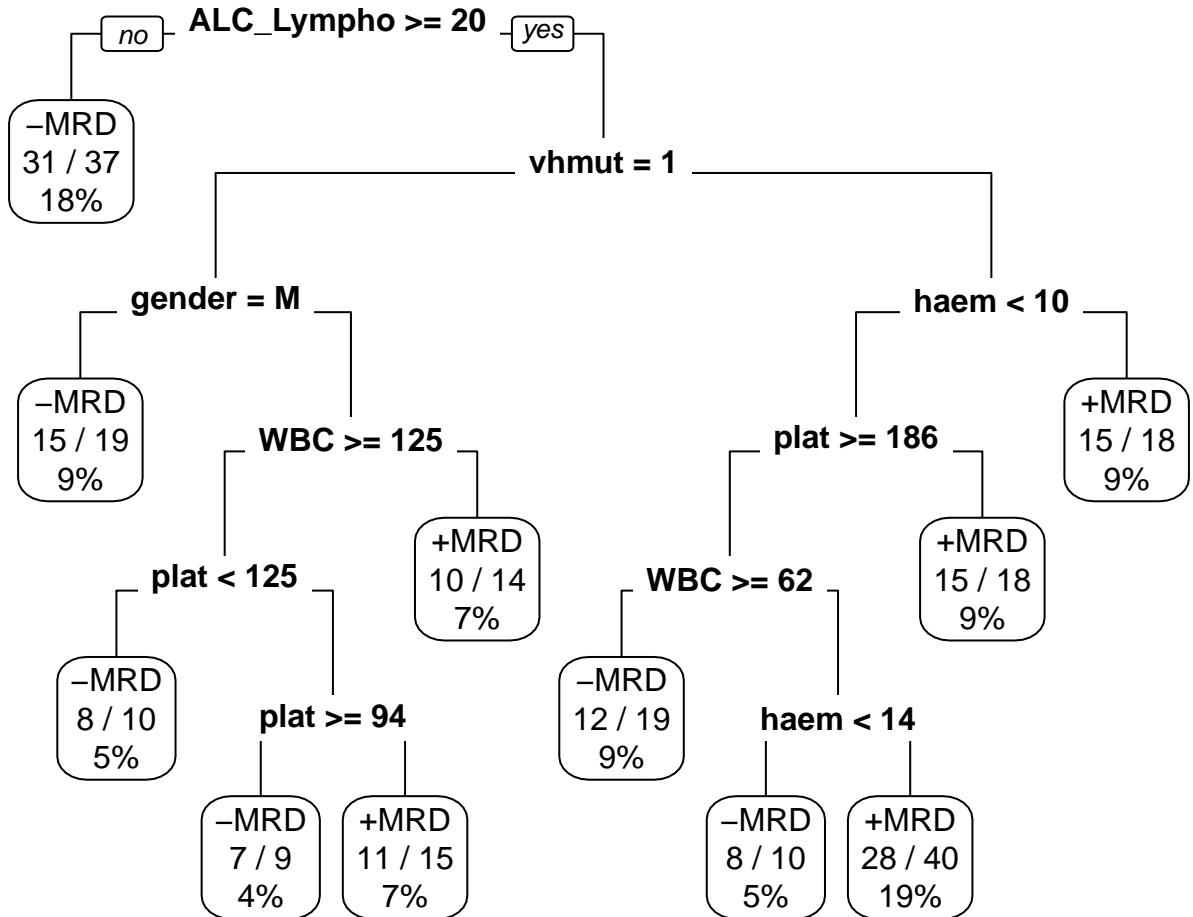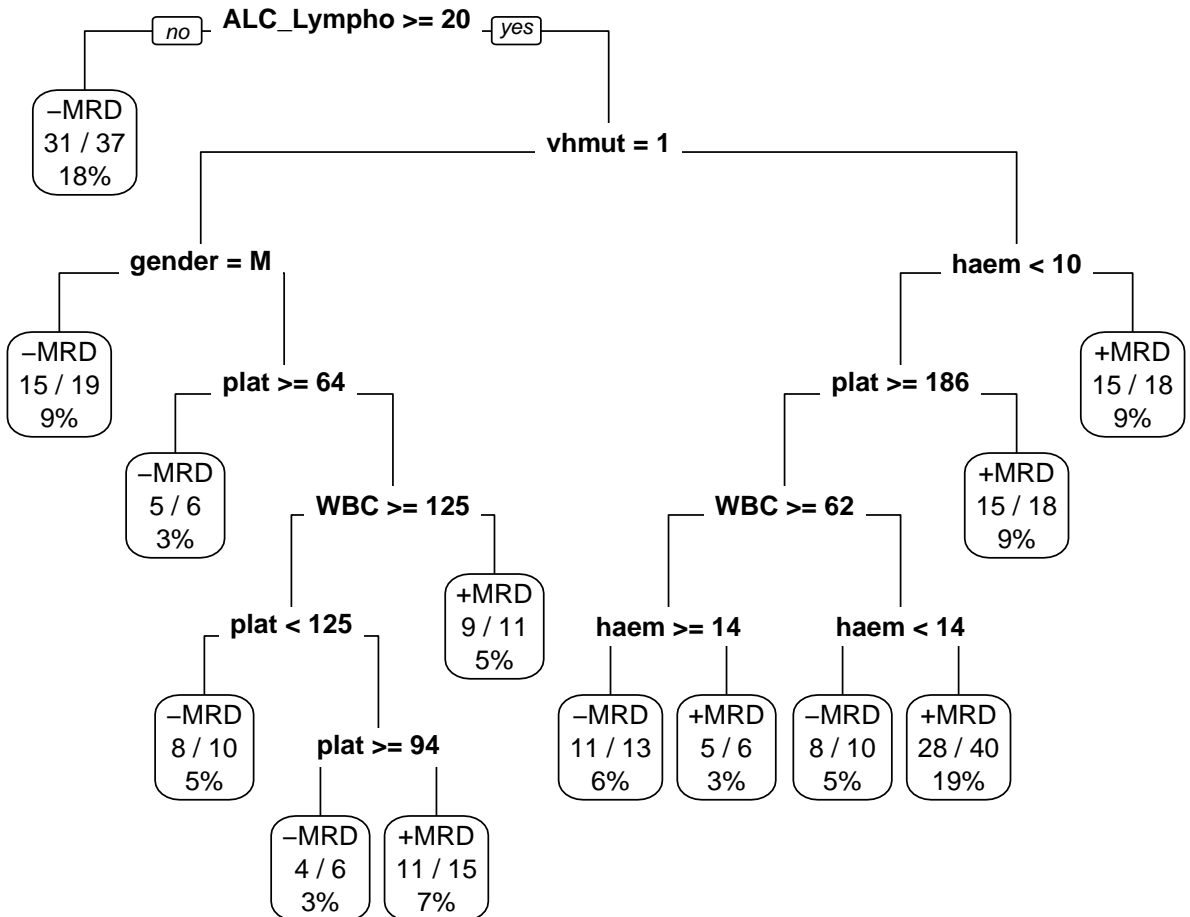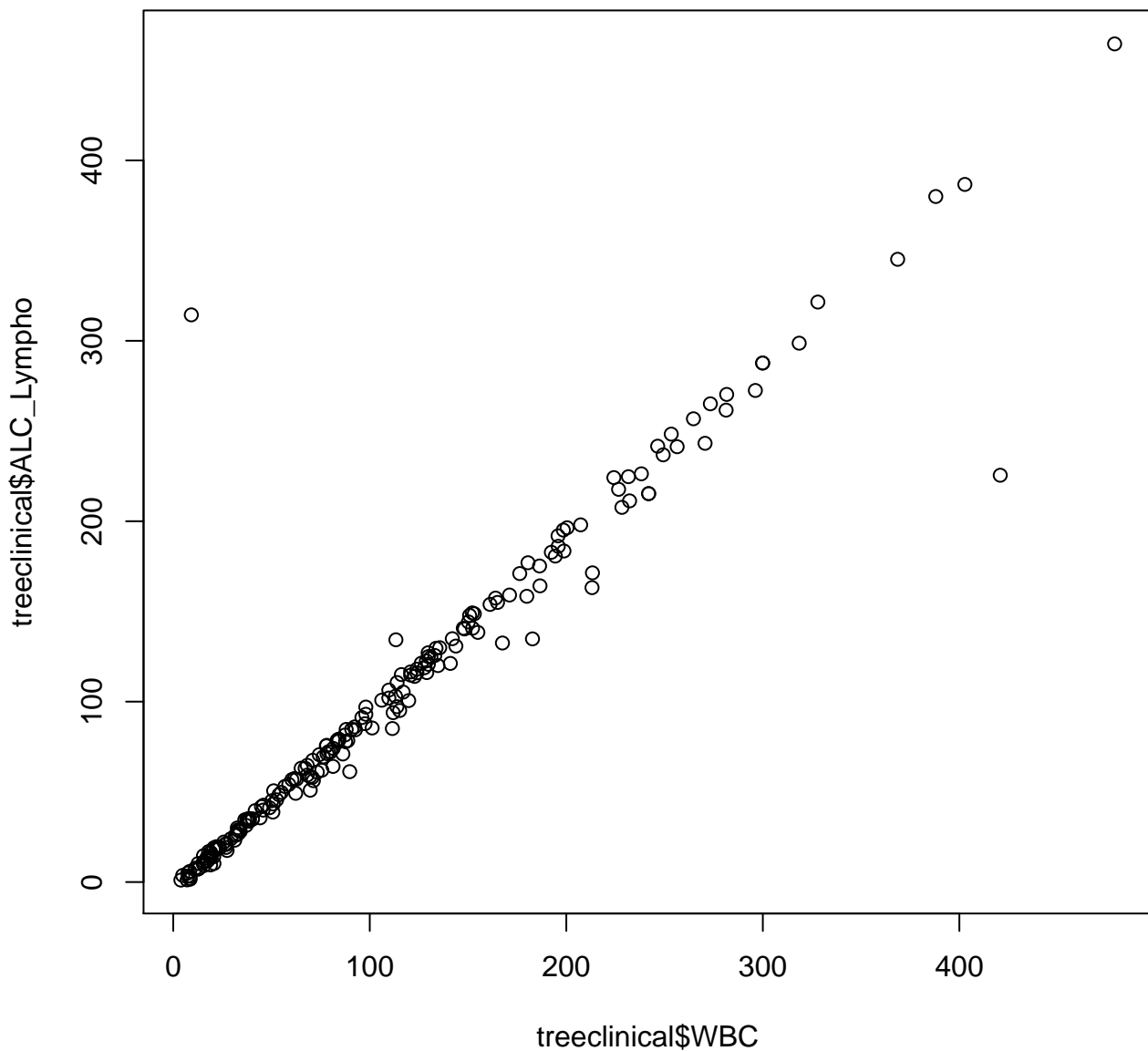# 3c) tree21:All clinical data
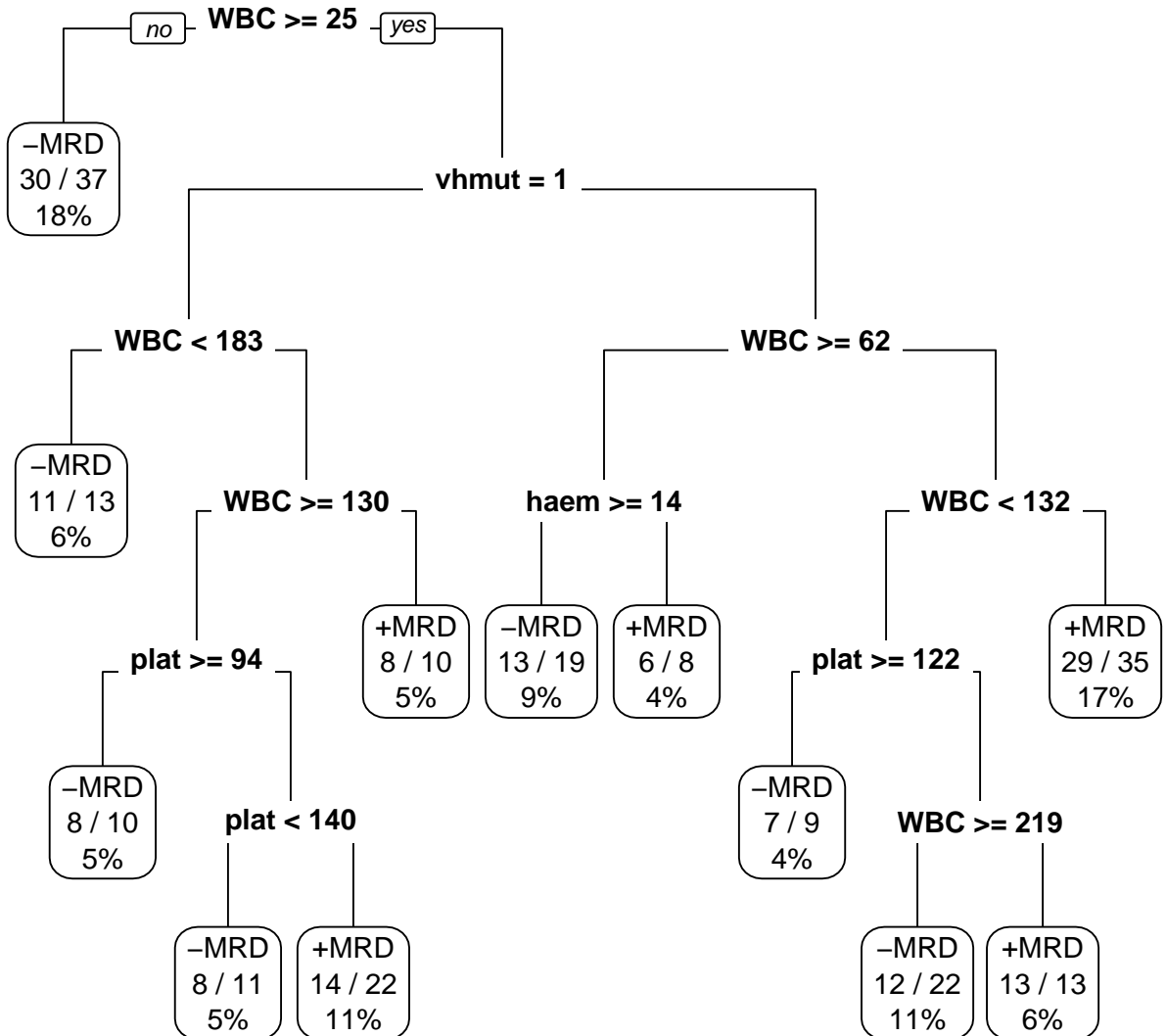# cp=0.01, endgroup=6, MissClassErr=21%

This tree has quite a good Missclassification error, but is quite complicated.
One idea would be to use either Lymphos OR WBC and platelets OR haem
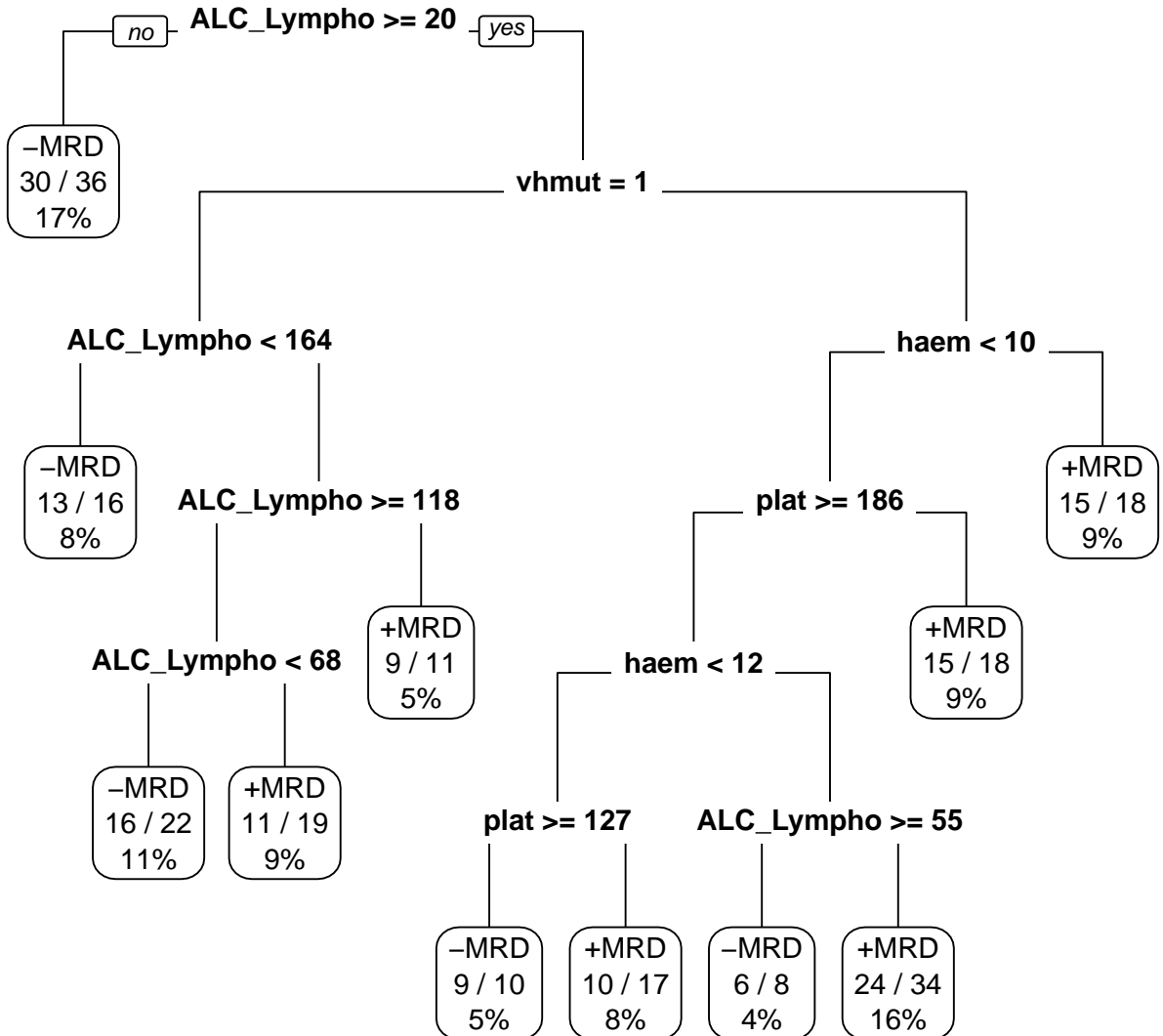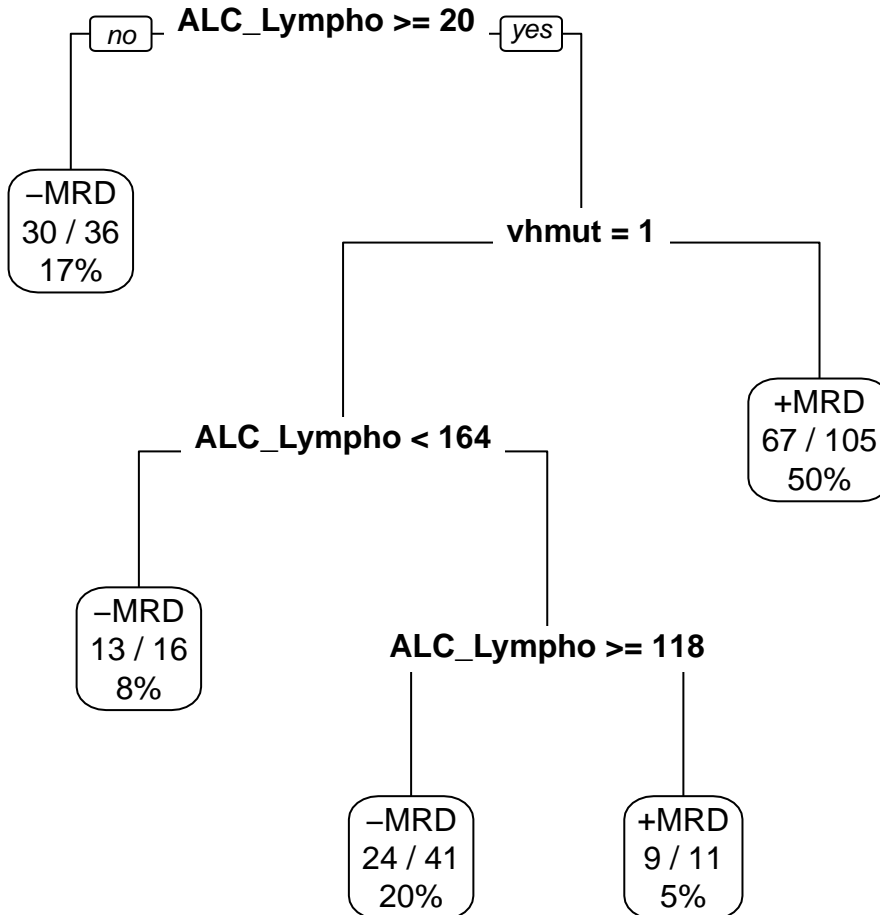(See next trees)

WBC and Lymphos are highly correlated

3c) tree22:All clinical data, using only WBC
cp=0.01, endgroup=8, MissClassErr=23.9%

**3c) tree23:All clinical data, using only Lymphos
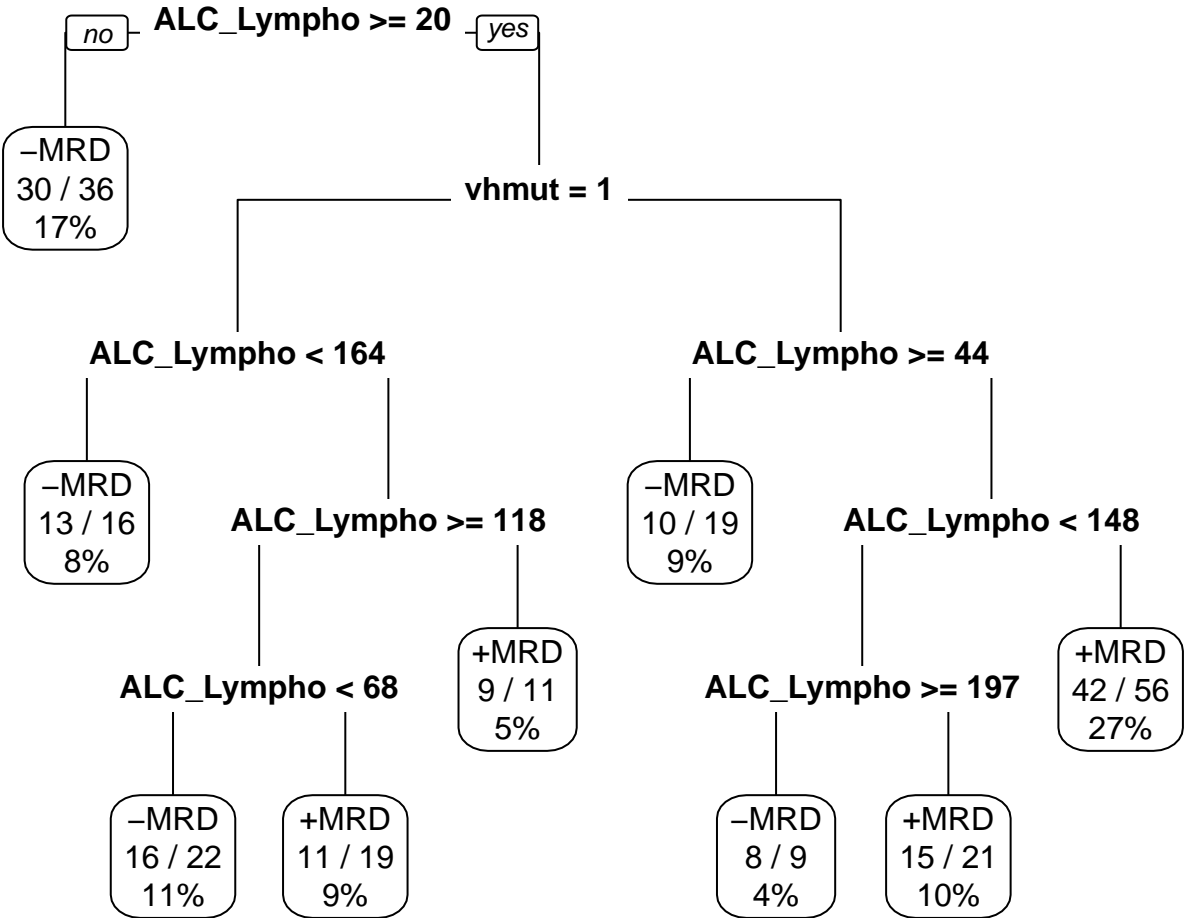cp=0.01, endgroup=8, MissClassErr=24.4%**

3c) tree24:All clinical data, using only Lymphos
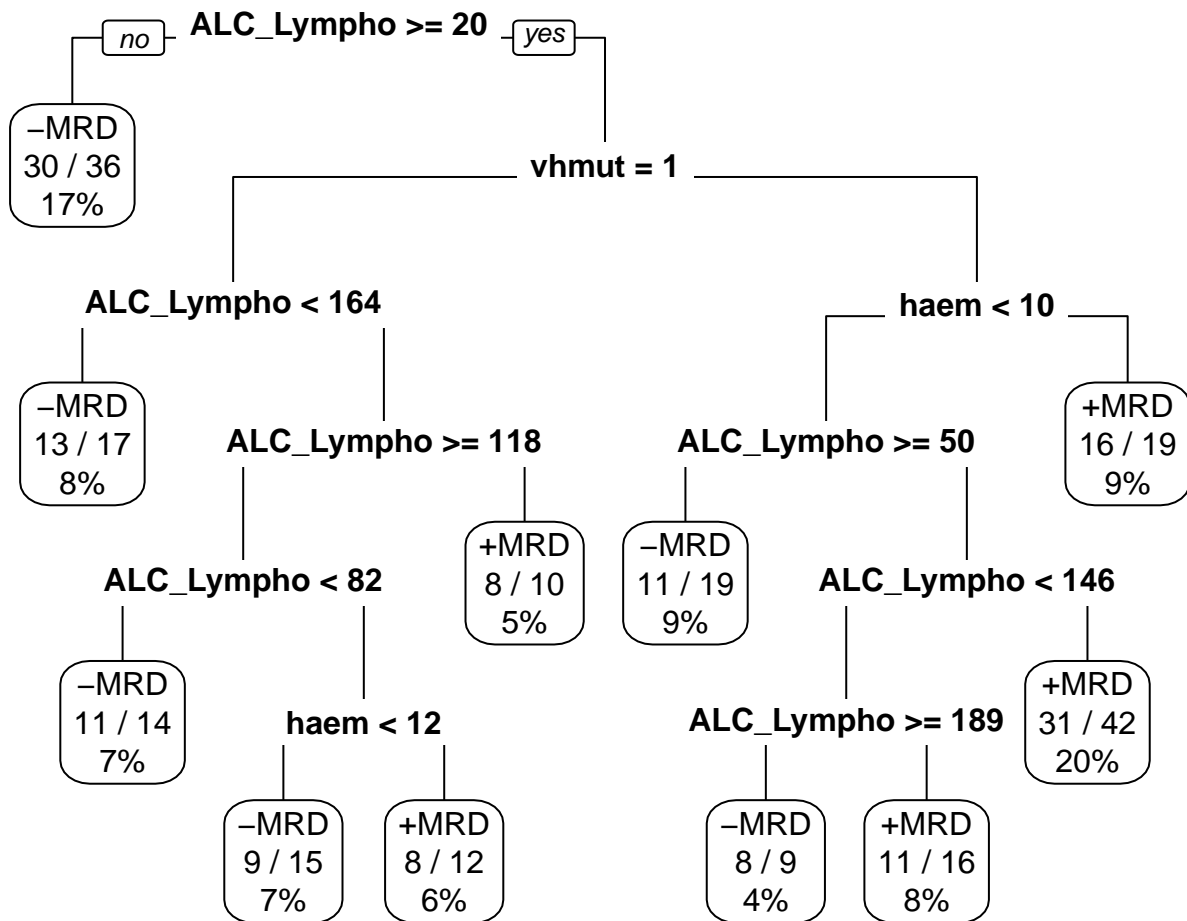# cp=0.03, endgroup=8, MissClassErr=32%

Trying to make the tree less complex results in increase in missclassification.
Note that platelets and haem are not important anymore.

3c) tree25:All clinical data, using only Lymphos and only plat
cp=0.01, endgroup=8, MissClassErr=26.3%

ALC_Lympho >= 20
no / yes

−MRD
30 / 36
17%

vhmut = 1

ALC_Lympho < 164

ALC_Lympho >= 44

−MRD
13 / 16
8%

ALC_Lympho >= 118

−MRD
10 / 19
9%

ALC_Lympho < 148

+MRD
9 / 11
5%

ALC_Lympho < 68

ALC_Lympho >= 197

+MRD
42 / 56
27%

−MRD
16 / 22
11%

+MRD
11 / 19
9%

−MRD
8 / 9
4%

+MRD
15 / 21
10%

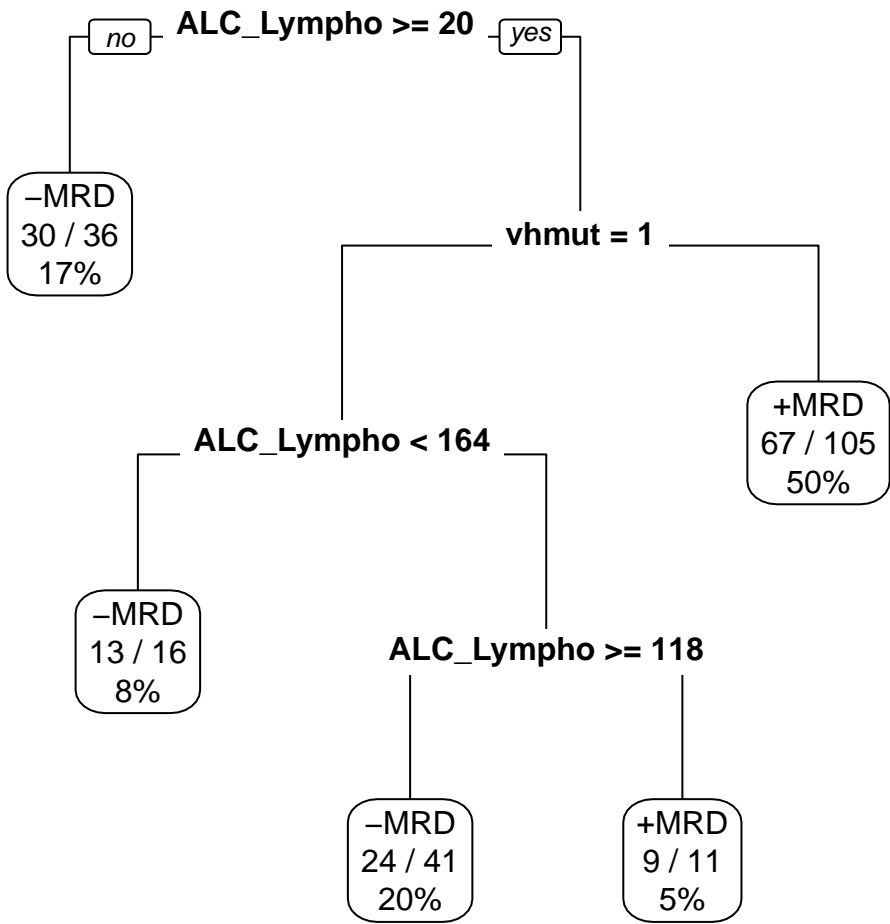Note that platelets don't seem to be important anymore once you take out haem.

3c) tree26:All clinical data, using only Lymphos and only haem
cp=0.01, endgroup=8, MissClassErr=25.4%

Note that haem doesn't seem to be important anymore once you take out platelets.

3c) tree27:All clinical data, using only Lymphos and only plat cp=0.03, endgroup=8, MissClassErr=32%

Trying to make the tree less complex results in increase in missclassification. Note that platelets don't seem to be important anymore once you take out haem.
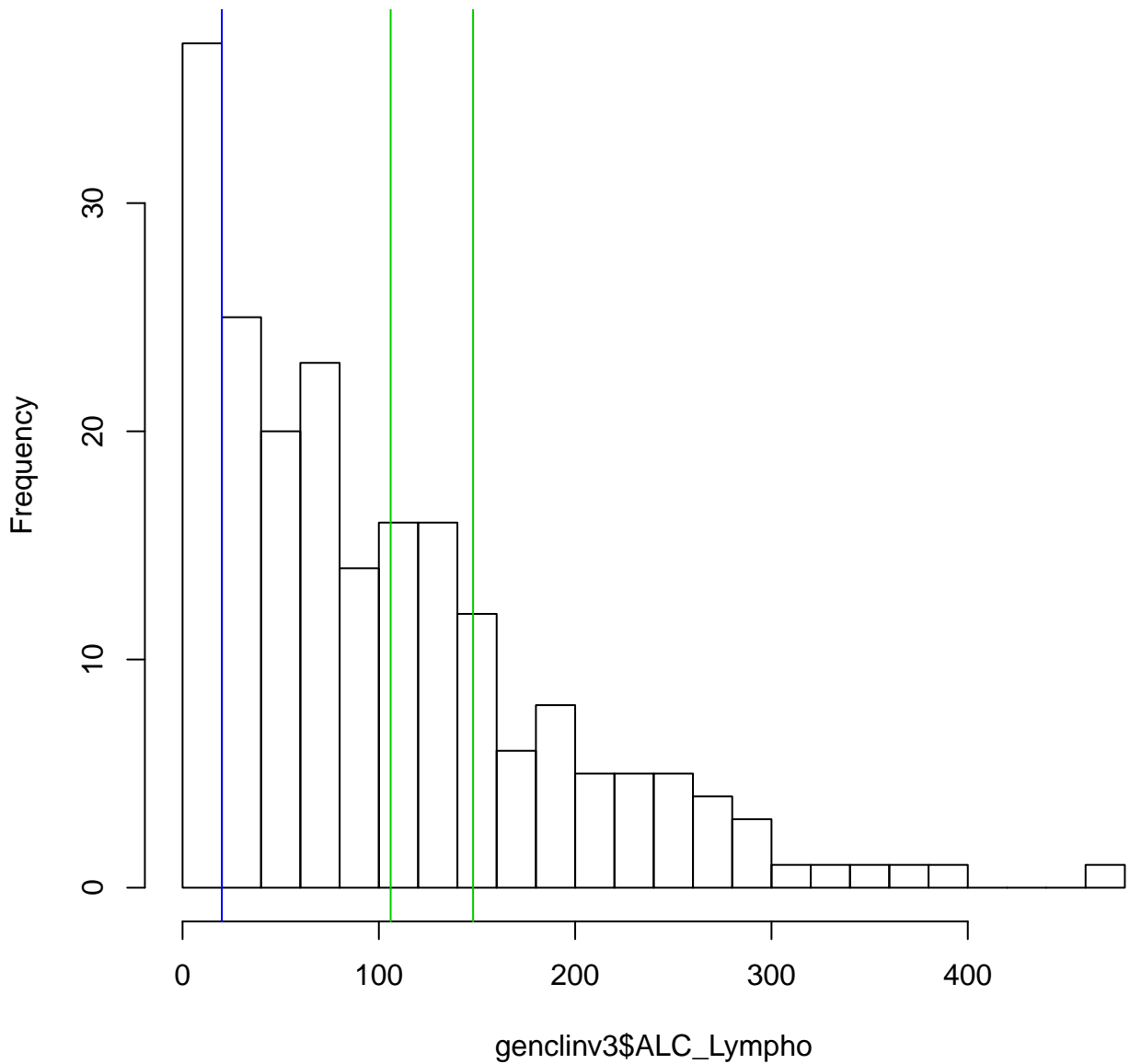
Question1:
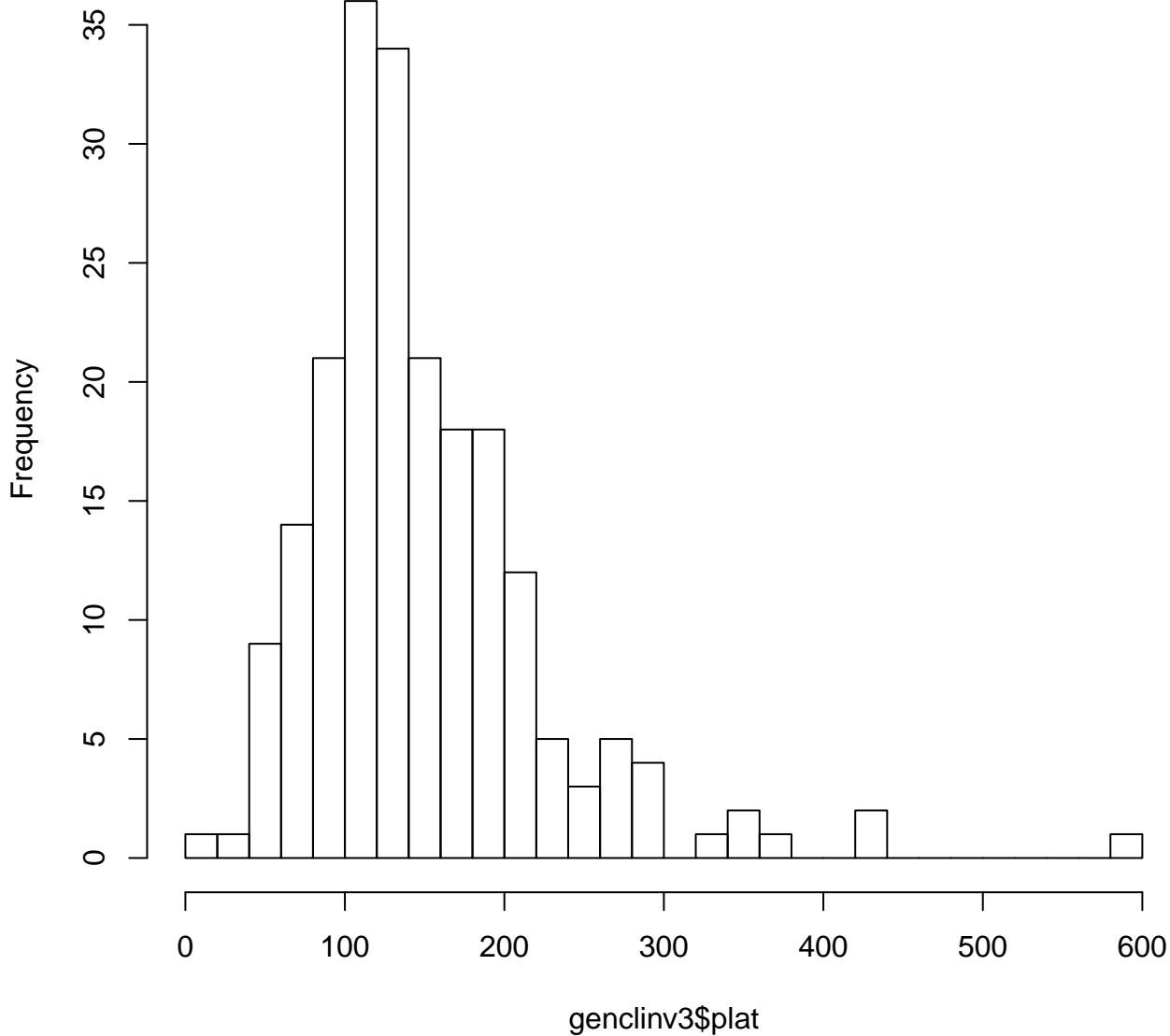Is there a way to group full blood count data?
Are there any common standards that are used for deciding if a blood count is abnormal?
I attach an overview of how the data is distributed.

Histogram of genclinv3$ALC_Lympho

Histogram of genclinv3$plat

Histogram of genclinv3$haem

Frequency

genclinv3$haem