Statistical Analysis of Hypercalcaemia Data related to Transferability

Anne Frølich and Bo Friis Nielsen

December 29, 2005

Abstract

In this report we describe statistical analyses related to a study of hypercalcaemia carried out in the Copenhagen area in the ten year period from 1984 to 1994. Results from the study have previously been published in a number of papers [3, 4, 5, 6, 7, 8, 9] and in various abstracts and posters at conferences during the late eighties and early nineties. In this report we give a more detailed description of many of the analyses and provide some new results primarily by simultaneous studies of several databases.

1 Introduction

This report comprises the data material and describe statistical tests used in the thesis. "Transferability of data and medical knowledge - A case on "diagnosing hypercalcaemia" based on [3, 4, 5, 6, 7, 8, 9]. We give an overview of the study populations in Section 2, present basic comparisons of the hypercalcaemia populations in Section 3, and describe the handling of missing values in 4. In Section 5 we give a more detailed description of the statistical analyses that formed the basis for the papers [7] and [6, 8] while some further analyses of discriminant functions are presented in Section 7. Finally, in Section 6 we give more details of the analyses in paper [5, 9].

2 Study populations

Four studies in Hvidovre and Herlev hospitals both affiliated with the University of Copenhagen, were carried out to assess various aspects of hypercalcaemia in hospital populations (Table 1). Three studies took place in Hvidovre 1,2,3 (Hvidovre 1984 [7], Hvidovre 1989 [8], Hvidovre 1994 [3], and one study in Herlev (1987 [6]. The studies comprised a total of 635 hypercalcaemia patients, and 28,900 patients in total. For the distribution of beds on specialties see Table 2. For mean and SD for the four datasets, see Table 3-6.

3 Comparison of hypercalcaemia populations

We wanted to study the homogeneity of the blood profiles for the four hypercalcaemia populations, as the four hypercalcaemia population data sets provided a unique opportunity to study the variability of hypercalcaemia populations over time and between sites.

Let $\mathbf{X}_{ijkl}$ denote blood measurements from patient l in population i and diagnose group j with gender k.

We assumed that data could be adequately described by a normal distribution as will be discussed in section 3.1. Formally $\mathbf{X}_{ijkl} \in N(\boldsymbol{\mu}_{ij}, \Sigma_{ij})$. It was therefore natural to perform the homogeneity analysis of the datasets following the scheme:

1. Test for multivariate normality
   - Within diagnoses and populations (Section 3.1), i.e. test the model $\mathbf{X}_{ijkl} \in N(\boldsymbol{\mu}_{ij}, \Sigma_{ij})$
2. Tests for identical diagnoses between populations
   - Compare covariance structures $\Sigma_{ij} = \Sigma_j$ (Section 3.2)
   - Compare mean vectors for identical diagnosis between populations and genders $\boldsymbol{\mu}_{ijk} = \boldsymbol{\mu}_j$. (Section 3.3)
3. Tests for different diagnoses within populations
   - Compare covariance structures $\Sigma_{ij} = \Sigma_i$ (Section 3.4)
   - Compare mean vectors for different diagnosis and genders within populations $\boldsymbol{\mu}_{ijk} = \boldsymbol{\mu}_i$. (Section 3.5)

We wanted to restrict the number of variable combinations with a special interest in variables known to be of importance for differentiation between hypercalcaemia-associated diagnoses. It was decided not to distinguish between genders when investigating for multivariate normality and covariance structure in spite of the differentiation according to gender in our earlier work. The reason for this was twofold. Gender differences found in the analyses of the individual data sets were not strongly significant and secondly, the number of observations especially for men would be prohibitively low for many subgroups when testing for multivariate normality and common covariance matrices. Nevertheless, we included gender in the mean comparisons, as we knew that some variables vary between genders in normals (haemoglobin and albumin). Only variables measured in all four populations including, calcium, phosphate, urea, ALP, LD, albumin, protein, hemoglobin, and age were considered. Calcium, phosphate, ALP, albumin, and chloride were evidently important as supported by earlier studies (Table 15, thesis). However, chloride is not measured routinely in Danish hospitals and was not measured in our studies. Further, the nature of the data sets as measurements in patients with elevated serum calcium excluded this variable from a normal distribution hypothesis from a theoretical point of view. Thus it was decided to exclude calcium from the basic set of variables. Strictly, the truncation mechanism for serum calcium would also affect the normality for other variables correlated to serum calcium. However, this issue was not addressed. The three remaining variables phosphate, ALP, and albumin were included in all analyses and will be referred to as the basic variables. Efforts to restrict the variable selection further were not fruitful since no solid criteria for ranking among the remaining variables common to all four datasets were found. In general all covariance and mean analyses were performed including the three basic variables, and then all possible subsets of variables with four, five, and sometimes six variables.

3.1 Multivariate normality

A number of goodness of fit tests (GOF) for multivariate normality (MVN) has been suggested [13, 19]. An extensive survey is given in Romeu and Ozturk [19]. These authors in turn proposed a new scheme for classification of GOF-MVN tests and a new test.

There is a general understanding that no single MVN-GOF test is superior to other tests which is quite natural due to the generality and complexity of the alternative hypothesis of non-normality. Romeu and Ozturk compared a number of different methods by extensive Monte Carlo sampling from different test distributions. They concluded that the power of their method is quite competitive in general. Most MVN-GOF tests including the test proposed by Romeu and Ozturk rely on asymptotic properties of the test statistic implying that claimed levels for type I errors might not be valid at least not for smaller sample sizes. There is no validation of the type I error in [19]. In a recent study Mecklin and Mundfrom [13] also compare a number of MVN-GOF test. They find that the probability of type I errors of several tests including the test of Romeu and Ozturk is quite different from the asymptotic value. They proceed in comparing MVN-GOF tests with experimentally "correct" type I error probabilities. Their main recommendation is to use the Henze-Zirkler test. Based on this study, we decided to rely on the Henze-Zirkler test statistic as a convenient choice since it is readily available when using SAS proc model. This procedure also calculates the Mardia skewness and Mardia Kurtosis test statistics, both of which were found by Mecklin and Mundfrom to have well defined type I error level, and were recommended as supplementary tests [13]. The issue of which deviation from normality has the most severe consequences were not addressed in any of the studies examined. Mecklin and Mundfrom Mardia et al. (1979) and DeCarlo (1997) for the general conclusion that mean value tests are more sensitive to skewness problems while variance tests are more sensitive to kurtosis problems.

3.1.1 Results

Univariate test results (Shapiro-Wilks test) are reported in Table 8. The log normal distribution fits the distribution of urea, ALP, protein, and LD better than the normal. The two distributions fit the phosphate, alb, and hemoglobin distributions equally well, while the normal distribution fits the age distribution better than the log normal. Table 9 presents significance levels for the three MVN-GOF tests for untransformed and log-transformed data. The log-transformed data are generally better described by the multivariate normal distribution than the non-transformed data. The hypothesis of a multivariate normal distribution can be accepted for most sub-groups apart from the group of malignant patients in Herlev. Mardia's skewness test seems to be the most sensitive. The power of the tests was highest for the group of malignant patients in Herlev being the largest group. In order to improve the test power we pooled the patient groups in Hvidovre with measurement values standardized with population mean and standard deviation. The results of the MVN-GOF tests are presented in Table 10. Based on the results in Table 8 it would be natural not to apply the log transformation to age. However, somewhat surprisingly the fit to the MVN was not as good (not reported in tables) as those obtained in Tables 9 and 10.

The power of the tests for the larger diagnosis groups, PHPT, and malignancy are evidently stronger due to the higher number of observations. For all miscellaneous groups and most groups of transients conclusions are probably rather weak. However, several tests for non-transformed variables were strongly significant, even for the smaller groups. No outliers were evident for logarithmically transformed data. Thus we conclude that no gross deviation from normality is present for the log-transformed data.

3.2 Covariance structure within diagnoses

Equivalence of distributions means equivalence of covariance matrices and mean vectors when assuming multivariate normality. At least in principle, comparison of mean vectors depends on whether covariance matrices are identical or not, and one can even question the soundness of comparing mean vectors if one concludes that covariance matrices differ. It was known from Frølich et al that changing measurement inaccuracy and imprecision was a fact for several routine analyses during time in Hvidovre and Herlev, as well as inaccuracy between Hvidovre and Herlev [9]. It might be theoretically expected that these inherent inaccuracy problems of the datasets would be less manifest when comparing covariance matrices as the impact of inaccuracy problem may be diminished. We also performed comparison of covariance matrices solely among the Hvidovre populations (these results are not reported in tables). Covariance matrices within the four diagnosis groups were compared for a number of subsets of variables. The three basic variables were always included. Tests were performed for all possible subsets of three to seven variables, and were only made for logarithmically transformed variables. The generalized Bartlett's test is the standard choice for comparison of covariance matrices as described in Rencher [18] p. 225. Some authors have argued that this test is far from optimal [1, 23] and alternative tests have been proposed but none of these seems to have obtained widespread acceptance. The type I error or size of the generalized Bartlett's test is known to be highly sensitive to the normal hypothesis [1, 17, 20], which appears to be one of the basic problems of this test. However, since the normality hypothesis was accepted for most subsets of variables, and since none of the alternative tests proposed has been widely applied, we considered it reasonable to rely on the standard test for equality of covariances. Some other tests have better power when testing against specific alternatives. In our case there is no obvious alternative of what the alternative might be and an omnibus test like the generalized Bartlett's test seems most appropriate. The test was performed using SAS Proc Discrim.

3.2.1 Results

In Table 12 we report test results for all subsets with up to 5 variables as results for 6 and 7 variables gave little or no extra information, i.e. one could deduce the results for the tests from the results with 3-5 variables. Problems with homogeneity, whenever present, seem to be associated with single variables rather than patterns. The PHPT groups did generally not deviate significantly from the hypothesis of equal covariance. However, when LD was included most tests failed at a 1% level or lower (Table 12). Less consistently some variable subsets including calcium caused problems too. This was primarily manifest for comparisons within Hvidovre populations (not documented in the report). For malignancy the subset including the three basic variables, and protein was highly significant also when considering the Hvidovre populations only. Not surprisingly all other subsets including protein were significant too. When comparing all four populations a few other combinations including calcium showed a modest significance. For miscellaneous the tests failed for combinations including calcium and urea. The significance was less profound but still present when comparing covariance matrices only between Hvidovre populations and might be ascribed to the low number of patients in the second and third Hvidovre population. Since the size (type I error probability) can be highly inflated when the normality assumption is violated one can speculate this as the reason for problems occurring with calcium. For transients identity of covariance matrices was rejected at approximately 0.1 % for the basic variable subset. Consequently the test failed for most other combinations generally with very low p-values. We did not investigate whether this could be ascribed to a single of the three basic variables.

3.2.2 Conclusion

In conclusion the covariance structures for the diagnoses PHPT and malignancy seem reasonably homogeneous between hospitals while this is less obvious for miscellaneous and questionable for transients. Technically we accept the hypothesis $\Sigma_{ij} = \Sigma_j$.

3.3 Comparison of mean vectors within diagnoses

We wanted to assess the degree of comparability within diagnoses between hospitals. This was less relevant for transient where in-homogeneity of covariance matrices were shown. The comparison of mean vectors for identical diagnoses between populations, and genders were performed for all possible subsets including the three basic variables. The model used was unbalanced two-way manova with gender and population as factors. The hope was that data primarily expressed true biologically variation such that we should find few or none population effects. However, due to the nature of the miscellaneous and transient groups we were aware that this might not be true in these cases.

3.3.1 Results

For malignancy there was no significant interaction or gender effects (Table 13). The three basic variables showed a significant population effect of 0.2%. Both ALP and albumin individually showed significant differences. The population effects were significant throughout though strengthened whenever either calcium or protein were present. For PHPT we found no interaction effects and no effects for the three basic variables (Table 14). The four variable subsets with lnhgb and lnage had 1% gender effects, while lnpro and lnage had approximately 1% population effects. These effects were obviously also present for 5 variable subsets and amplified whenever two of these were both present; e.g. the 5 variable subset with lnhgb and lnage. The 5 variable subset with lncat and lnld had a 2% population effect. No combinations, which could not be derived from these conclusions occurred for 6 and 7 variable subsets.

For the group of miscellaneous we found six significant effects all with p-values above 0.1% (Table 15). Two of these effects were population effects including urea, protein and haemoglobin while the remaining effects were gender-effects including protein, and age. For the group of transient most interaction effects were significant with p-values in the range between 0.1% and 1% (Table 16). We did not investigate whether this could be ascribed to a single of the three basic variables.

3.3.2 Conclusion

Comparing populations by diagnoses showed a high degree of comparability that is somewhat surprising taking results on inaccuracy and imprecision from the donor study into consideration (see section 6).

3.4 Covariance structures between diagnoses within populations

There is a good theoretical reason, to believe that the covariance matrices for different diagnosis are different within the populations as the pathogenesis for the hypercalcaemia are different. The statistical procedures for this part of the study were similar to the procedure used when comparing populations within diagnoses. The test results presented in Table 17 for subsets with up to 5 variables strongly rejects the hypothesis that covariance matrices should be identical between diagnoses.

3.4.1 Results

Only four variable combinations for the first Hvidovre population had p-values above 0.1 %. A few combinations including hemoglobin were insignificant. All combinations for the second Hvidovre population were significant at the 0.1% level. The third Hvidovre population with least patients had a p-value just below 0.1% for the subset of basic variables, and six combinations with p-values between 1% and 10%, all other were highly significant. For the Herlev population two combinations including haemoglobin had p-values between 0.1% and 1% and two combinations had a p-value around 3%.

3.4.2 Conclusion

The covariance matrices of the diagnosis groups were significantly different as expected. The overall variability as measured by the determinant of the covariance matrices was highest for cancer. We did not do the calculations comparing the three other diagnoses excluding cancer but we are convinced from the general magnitudes of the determinants of the covariance matrices that the deviation between these three groups would still be substantially higher than the variation within diagnosis groups between populations.

3.5 Mean structure between diagnoses within populations

We compared the mean structure of the diagnoses within each population. All subsets were highly significant. Significance levels were higher but still low for the third Hvidovre population most likely due to the low number of patients.

3.6 Statistical tools

The statistical analysis was performed by means of the SAS and the BMDP packages (SAS Institute Inc., North Carolina, USA) (BMDP Statistical Software Inc., California, USA).

4 Replacement of missing data/imputation

Several observations were incomplete and it was tested whether the missing data showed a certain pattern, which was not the case. Imputation methods were still in an early state when the first part of this work was done in the mid-eighties. Among readily available methods was imputation by regression on non-missing variables as discussed by Solberg and Rudin, and available through the BMDP-AM program. In the first Hvidovre study missing values were imputed using the BMDP AM program with option all-value, a regression method that were used for all further analysis. Imputations were performed within each diagnosis group, and done separately for each gender. Computer intensive methods for data imputation has now reached a mature state with availability in commercial software. Little and Rudin [12] recommend multiple imputation (MI) based on likelihood inference. The MI method offers the opportunity to investigate sensitivity with respect to the imputations. However, it is important that the model used in the imputatin procedure has a reasonable fit to data since the approach is parametric, based on maximum likelihood estimation. As we had concluded that data could be described by a multivariate normal distribution we decided to apply MI for all datasets. Our analyses were performed using SAS Proc MI (multiple imputation).

5 Hypercalcaemia discriminant functions

Our first development of hypercalcaemia discriminant functions are described in [7]. However, the technical description in these papers are quite brief. Here we provide more details and some additional comments and analyses.

In this section we describe the development and evaluation of the Hvidovre 1 discriminant function.

5.1 Development of hypercalcaemia discriminant functions in Hvidovre 1

It was decided early to develop discriminant functions for each gender. See section 5.3 for a discussion of this. The classification method of choice was linear discriminant analysis (LDA) as described in most textbooks on multivariate statistical methods [18] p. 215. Classification based on LDA is mostly expected to be a reliably and robust method [21] and [10] p. 89. LDA is principally a parametric method with certain assumptions necessary for P-values and distance measures to be valid.

5.1.1 Selection of discriminant variables

Discriminant variables were selected by a stepwise procedure with forward selection. The variables with the highest potential to separate groups were identified including one variable at a time retaining only those variables that led to a significant improvement in the differentiation between diagnosis-groups. Stepwise selections of variables were performed using the BMDP 7M program, and later recalculated using SAS Proc Stepdisc with similar results. In the BMDP 7M program entrance and exit levels are specified by F-values rather than P-values. The F-to-enter and F-to-exit values were set to 3.1 and 2.1 respectively. This was in accordance with significance levels of 5 % and 7.5 % respectively, as the same variables would be selected with these values, as verified by SAS proc Stepdisc. The entry step and significance level at entry are presented in Table 19 for women and men respectively. We chose to use identical prior probabilities, as there was no unique way to assign prior probabilities since the prevalence of diseases associated with hypercalcaemia varied considerably, and also vary significantly for the two populations in question (Table 15, Thesis). We therefore considered it the most reliable choice to assign equal probabilities to each of the groups. This is not an uncommon approach see [11] p. 111.

The discriminant functions are given in Table 20. In Table 19 we also present results for a discriminant function applied for both genders. In the light of the actually small differences between the genders, one can question the optimality of developing separate functions for each gender and we do not recommend it.

5.1.2 Performance evaluation of discriminant functions

The best assessment of discriminant functions performance is obtained by applying the function on a new and similarly defined population as the original population. However the number of patients in the two patient populations was not sufficient to divide the populations to assess the discrimination power by the most correct approach. A common alternative approach is to assess discriminant functions by cross-validation or the hold-out method. Each observation is sequentially removed from the sample, and classified by the functions estimated from the remaining observations. The rate of misclassified observations is used as an estimate of the true error rate of the classification function [21]. Since missing values were estimated, the method was not a true cross-validation as all observations contributed to the estimation of missing values. Most likely this effect tends to create an optimistic estimate of the misclassification rate. We believe that the effect of estimated missing values on the estimated misclassification rates to be small. We have not been able to find any in-depth discussion of this issue despite the common nature of the problem. The results from the Hvidovre 1 function in Hvidovre 1 is given in Table 21. Diagnostic performance was assessed by various methods Table 24, see Table 18 for definitions (Table 12, thesis).

5.2 Development of hypercalcaemia discriminant functions in Herlev

The intention with the Herlev study was to assess the performance of the Hvidovre functions when applied in a different setting. However, the clinical problem in the Hvidovre 1 study was not relevant in Herlev due to the low number of patients with transient hypercalcaemia.

In [6] we compared discriminant functions from Hvidovre 1 and Herlev respectively developed to separate malignancy from non-malignancy. For the Hvidovre functions we still used step 1 but now the groups of PHPT, miscellaneous and transient formed the aggregate group of non-malignancy, while the other group obviously was the group of malignant patients.

Step 1 in Herlev was performed as discrimination between malignancy, PHPT, and miscellaneous, while step 2 had the same classification as in Hvidovre 1. The same procedure as in Hvidovre 1 was applied in Herlev for the development of step 1. Stepwise selection of variables and development of discriminant functions with identical priors. The summary of the stepwise selection is given in Table 19. There was a slight inconsistency in the development of the two functions as the F-to-enter and F-to-remove values were 2.1 and 1.8 respectively for the Herlev functions, corresponding to p-to-enter values of approximately 10%. The discriminant functions are presented in Table 23. The diagnostic performance is shown in Table 24.

5.3 Functions without gender separation

The function estimated on all early patients (both genders) is based on the variables given in Table 19. The performance of the function is comparable to the performence of the functions developed for each gender (Table 25 and Table 24).

6 Simultaneous measurements at Herlev and Hvidovre

Early in the process of collecting data for [6] it became clear that there might be problems with inaccuracy and imprecision between Hvidovre and Herlev hospitals as the diagnostic patterns were clearly different.

A study was set up in Herlev hospital, including 183 blood donors to assess comparability of routine measurements in Hvidovre and Herlev hospitals. The measurements were taken on 15 different days during a four-months period. Two blood samples were taken in the same needle puncture and were randomly allocated for routine measurement analyses in Hvidovre and Herlev hospitals. The blood analyses were analyzed on identical machinery as for the hypercalcaemia populations. This section gives a more in-depth description of the statistical analysis that formed the basis for [9]. The overall goal for the analyses were to compare the results from blood measurements in Herlev and Hvidovre hospitals in order to identify and quantify differences in inaccuracy and imprecision between the two sites.

The Danish Society for Clinical Chemistry regularly distributes control material for external quality assessment to all Danish hospital laboratories. The means and coefficients of variation were available and analyzed in parallel with the donor material. These data were compared to results obtained from the donor data.

6.1 Mathematical model for donor material

Before the analyses we define the following quantities

$X_{Herlev,ij}$ The measurement of a value in Herlev hospital or the natural logarithm of the value. i refer to a specific patient, while j refers to a specific day. Of course there will be only one value of j for each i but several values of i for each j.

$X_{Hvidovre,i,j}$ The corresponding measurement at Hvidovre.

μ The overall mean of the serum value in the population.

δ_{Herlev} Systematic measurement error in Herlev hospital.

$\delta_{Hvidovre}$ Systematic measurement error in Hvidovre hospital.

P_i The biological true value of a specific sample's deviation from μ obtained from patient i.

$D_{Herlev,j}$ The day to day variation at Herlev.

$\epsilon_{Herlev,i,j}$ The intra assay variation corresponding to repeated measurements in one patient at a specific date, at Herlev.

$D_{Hvidovre,j}$ The day to day variation at Hvidovre.

$\epsilon_{Hvidovre,i,j}$ As for $\epsilon_{Herlev,i,j}$.

$$X_{ij} = \mu + \delta_{Herlev} + P_i + D_{j,Herlev} + \epsilon_{Herlev,i,j}$$

$$Y_{ij} = \mu + \delta_{Hvidovre} + P_i + D_{j,Hvidovre} + \epsilon_{Hvidovre,i,j}$$

$$P_i \in N(0, \sigma^2_B)$$

$$D_{j,Herlev} \in N(0, \sigma^2_{D,Herlev})$$

We can derive:

$$\epsilon_{Herlev,i,j} \in N(0, \sigma^2_{R,Herlev})$$

$$D_{j,Hvidovre} \in N(0, \sigma^2_{D,Hvidovre})$$

$$\epsilon_{Hvidovre,i,j} \in N(0, \sigma^2_{R,Hvidovre})$$

$$Z_{i,j} = X_{Herlev,i,j} - X_{Hvidovre,i,j} = D_{Herlev,j} - D_{Hvidovre,j} + \epsilon_{Herlev} - \epsilon_{Hvidovre}$$

$$Z_{i,j} \in N(\delta_{Herlev} - \delta_{Hvidovre}, \sigma^2_{Herlev,D} + \sigma^2_{Hvidovre,D} + \sigma^2_{Herlev,R} + \sigma^2_{Hvidovre,R})$$

6.2 Distribution tests

Usually serum values from a normal population are assumed to be normally or log normally distributed. This assumption was tested by the Sharpiro-Wilks test. Most of the analyses could be assumed be to normal or lognormal distributed. The distribution of bilirubin and AST could not be normalized in any of the hospitals. It was concluded that none of the distributions were grossly abnormal. However, differences or log ratios could not be adequately described by normal distribution in general, as they showed a heavy tailed distribution pattern (heavier than normal). This was seen as first indication of a time effect as this could be explained by presence of the D-effect i.e. $\sigma^2_D > 0$.

6.3 Background for the results in [9]

Several analyses pointed towards the presence of a changing inaccuracy over time. These analyses are reported in [9]. However, it is generally not fully transparent which statistical analyses were performed in this paper. The control material provided a basis for the estimation of $\sigma^2_{Herlev,R}$ and $\sigma^2_{Hvidovre,R}$ as the control material did not include biological variation. However, one could necessarily expect the equivalence of residual variances for human material and the control material of animal origin. The estimates of the residual variances based on control material was reported in Table 1 of [9].

Presence of day-to-day variation, - or better inter assay variation - i.e. the presence of P_i terms in the model above, was tested in the donor material using one-way analysis of variance with random effects (Table III [9]). The tests were performed for the two hospitals and for the differences. Obviously significant effects were more likely to be manifest for the differences as the biological variation were removed from these. Additionally the differences contains the combined inter assay variation from both Herlev and Hvidovre if present.

Analyses for day-to-day variation in the control material were also performed by one-way analyses with random effects and reported in Table V [9]. This table also included tests for stability of imprecision, i.e. an inter assay dependency of $\sigma^2_{Herlev,R}$ and $\sigma^2_{Hvidovre,R}$ was tested using Bartletts test. This effect could actually be caused by the standard inter assay variation as the control measurements for each month could belong to different batches/assays.

The imprecision of the control measurement in the two hospitals were compared by an F-test, reported in the last column of Table V in [9].

A test for inaccuracy $\delta_{Herlev} = \delta_{Hvidovre}$ were reported in Table IV, while comparison of bias estimates from the donor and control material was reported in Table VI.

6.4 Conclusions

For the purpose of the thesis work it is of particular interest to note that the inaccuracy was high for calcium as well as for albumin and that although the two hospitals had different reference intervals, this was not reflected in the results. This result and its consequences is discussed in section 7.3 in the Thesis.

7 Discrimination between Malignancy and PHPT: Transferability of discriminant functions

In this section we describe results when only discriminating between malignancy and PHPT. These results have not been published elsewhere. We wanted to investigate whether we could obtain better performing discriminant functions when only discriminating between the two larger and more well defined groups of malignancy and PHPT. Of special interest was the sensitivity to transfer, as functions developed on these two groups might be more robust. An additional motivation was that a number of studies reported in the literature investigates this question.

7.1 Variable selection in datasets with imputation

In the Hvidovre 1 dataset albumin, LD, and protein were selected by the stepwise procedure in all datasets, phosphate in most, and AST, urea, and bicarbonate in few (Table 26). Albumin, phosphate, and LD were selected in Hvidovre 2 and Hvidovre total, while albumin and bilirubin were selected in Hvidovre 3. In Herlev haemoglobin, LD, ALP, and albumin were selected in all datasets, globulin, and urea in few.

For the combined dataset albumin, LD, phosphate, haemoglobin and ALP were selected in most datasets, and age in few. Albumin was selected in all training sets, while phosphate and LD were selected in all training sets but Hvidovre 3. Perhaps most striking is the choice of LD in all 4 basic - and thus the two combined - datasets, as LD has not earlier been reported as an important hypercalcaemia discriminant variable. Discriminant functions were estimated for all imputed datasets based on the variables in the last column of Table 26. The variables entered the discriminant functions in a consistent way in all imputations. That is low values of albumin and haemoglobin and high values of LD, phosphate, ALP and protein indicated malignancy. Protein was only present in the Hvidovre 1 function and is probably the least general of the selected variables.

7.2 Performance on complete versus non-complete cases

We compared the performance of the discriminant functions on cases complete in discriminant variables and all patients for the Hvidovre 1 and Herlev populations. The results are reported in Table 27. One should note that the complete cases comprise a sub-group of all patients. Even with this in mind it is clear that there is no gross difference between the performance for complete and completed cases.

7.3 Performance when using priors

We performed a number of comparisons between functions with identical and proportional priors as reported in Table 28. The functions with proportional priors perform generally best for the Hvidovre 1 and Herlev populations with more malignant than PHPT patients while the functions with identical priors generally perform better on the Hvidovre 2 and Hvidovre 3 populations. While the two types have more or less the same performance for the two combined populations.

7.4 Performance when using quadratic (and priors)

We developed quadratic discriminant functions for the two combined datasets having a sufficient number of observations. From Table 29 there is no apparent difference between LDA and QDA. Finally in 30 we compare LDA and QDA discriminant functions with proportional and identical priors. No function seem to have striking superiority.

7.5 Sensitivity with respect to imputations

The imputation procedure gave the possibility of getting some insight in the sensitivity of the discriminant functions. In Table 31 we give maximum and minimum values for a number of discriminant functions.

There are no striking differences for diagnostic accuracy, while the variations in sensitivity and PPV are somewhat larger. The latter result is close to obvious as a higher than average sensitivity for malignancy will tend to be associated with a lower than average sensitivity for PHPT.

8 PTH as a discriminant variable discriminating between malignancy and PHPT

By the introduction of the PTH assay that measures the biologically active intact parathyroid hormone (PTH), about fifteen years ago, several of the differential diagnostic problems related to hypercalcemia decreased [15, 16]. The sensitivity of the intact PTH assay has been shown to be about 96% when combined with total serum calcium concentration [2]. Although the intact PTH assay has improved the diagnosis of hypercalcemia diagnostic limitations still remain. Some patients with PHPT have normal or only marginally elevated PTH that make diagnosing difficult [14]. In addition about 10% of patients with malignant hypercalcemia also have coexisting PHPT [22, 24]. We wanted to evaluate the hypothesis that serum concentrations of intact PTH is an important discriminant variable in discriminant functions differentiating between patients with hypercalcemia caused by malignancy or by PHPT. We compared the diagnostic performance of a discriminant function estimated from intact PTH and routine variables to a function estimated from routine blood variables only. Intact serum PTH concentration was measured in a subset of Hvidovre 3 patients; thirteen patients with malignancy, and eighteen patients with PHPT (Table 6). Discriminant functions were estimated by stepwise procedures as described in Section 5. The following variables were identified for the discriminant function with PTH as a possible discriminant variable; PTH, creatinine, potassium, urea, and total CO_2, and the function estimated from routine variables only identified; albumin, bilirubin, protein, and LD, as discriminant variables.

The diagnostic performance of the two discriminant functions with and without intact PTH was similar (Table 32). This is a surprising result as it was thought that intact PTH would add diagnostic performance to the discriminant function with PTH as a discriminant variable compared to the function based on routine variables only. However, one should be somewhat cautious drawing strong conclusions from the result due to the low number of observations in the study population. Nevertheless, we conclude that inclusion of intact PTH as a discriminant variable does not improve the diagnostic power of a discriminant function differentiating between malignancy and PHPT in this patient population.

9 Reducing the impact of transferability factors

The donorstudy revealed a significant measurement inaccuracy between Herlev and Hvidvore hospitals. The impact of this difference would presumably be even higher as the reference intervals were different for serum calcium. The reference value in Herlev was higher, while the donor material pointed towards a higher level at Hvidovre (Table VI in [9]). A third factor that might effect inclusion in the project was the different formulas for albumin correction of serum calcium.

	Calcium (cor)
Herlev	calcium - 0.00123 (albumin - 616)
Hvidovre	calcium - 0.0011 (albumin - 600)

For the Hvidovre 1 patients we estimated the calcium (cor) value in Herlev by adding the estimated mean inaccuracy for albumin of 48 to the albumin value and subtracted the estimated calcium inaccuracy of 0.07(Table VI in [8]). Thus, estimated Herlev calcium (cor) values for Hvidovre patients were calculated by the formula

Calcium (cor) = total-calcium - 0.07 - 0.00123 (albumin - 568)

Equivalently one could also calculate the albumin corrected value by adjusting the Hvidovre calcium (cor) value

calcium (cor) Herlev = calcium (cor) Hvidovre + 0.0315 + 0.00013 albumin(Hvidovre).

The composition of Hvidovre patients at when applying the Herlev inclusion levels when using this value of albumin corrected serum calcium is reported in Table 22.

References

[1] Dennis D. Boos and Cavell Brownie. Comparing Variances and Other Measures of Dispersion. *Statistical Science*, 19(4):571–578, 2004.

[2] B. Carnaille, C. Oudar, F. Pattou,J. Quievrux, and C. Proye. Improvements in parathyroid surgery in the intact 1-84 PTH assay era. *Aust N Z J Surg*, 68:112-6, 1998.

[3] A. Frølich. Combined reports on serum calcium and discriminant functions increase the diagnostic rate of hypercalcaemia. *Scand J Clin Lab Invest*, 57:725–730, 1997.

[4] A. Frølich. Prevalence of hypercalcaemia in normal and in hospital populations. *Danish Med Bull*, 45:436–439, 1998.

[5] A. Frølich, P. McNair, and I. Transbøl. Awareness of hypercalcaemia in a hospital population. *Scand J Clin Lab Invest*, 51:3–41, 1991.

[6] A. Frølich and B. F. Nielsen. Transfer of hypercalcaemia discriminant functions between local hospitals. *Int.J. Biomed. Comput.*, 41:167–173, 1996.

[7] A. Frølich, B. F. Nielsen, K. Conradsen, and P. McNair. Filtering clinically significant hypercalcaemia from non-signifcant hypercalcaemia at the laboratory level. *Scand J Clin Lab Invest*, 53:215–223, 1993.

[8] A. Frølich, B. F. Nielsen, K. Conradsen, P. McNair, and I. Transbøl. Within centre evaluation of hypercalcaemia discriminant functions 5 years after their development. *Int.J. Biomed. Comput.*, 40:235–240, 1996.

[9] A. Frølich, B. F. Nielsen, J. Nielsen, K. Conradsen, and P. McNair. Do local hospitals meet the analytical goals for the use of common reference intervals. *Scand J Clin Lab Invest*, 54:169–176, 1994.

[10] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer Series in Statistics. Springer, New York, 2001.

[11] Carl J Huberty. *Applied Discriminant Analysis.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1994.

[12] Roderick Little and Donald Rudin. *Statistical Analysis with Missing Data.* John Wiley & Sons, 2. edition, 2002.

[13] Cristopher J. Mecklin and Daniel J. Mundfrom. On Using Asymptotic Critical Values in Testing for Multivariate Normality. Downloaded from interstat.stat.vt.edu/InterStat/ARTICLES/ 2003/articles/J03001.pdf, jan 2003.

[14] C. Mischis-Troussard,P. Goudet, B. Verges, P. Cougard,C. Tavernier, and JF. Maillefert. Primary hyperparathyroidism with normal serum intact parathyroid hormone levels. *Q J Med*, 93:365–367, 2000.

[15] SR. Nussbaum, DJ. Zahradnik, JR. Lavigne, GL. Brennan, K. Nozawa-Ung, LY. Kim, HT. Keutmann, CA. Wang, JT Jr. Potts, and GV. Segre. Highly sensitive two-site immunoradiometric assay of parathyrin, and its clinical utility in evaluating patients with hypercalcemia. *Clin Chem*, 33:1364–7136, 1987.

[16] SR. Nussbaum, and JT Jr. Potts. Immunoassays for parathyroid hormone 1-84 in the diagnosis of hyperparathyroidism. *J Bone Miner Res*, 6 Suppl:S43–50, 1991.

[17] Peter C. O'Brien. Robust Procedures for Testing Equality of Covariance Matrices. *Biometrics*, 48:819–827, sep 1992.

[18] Alvin C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2002.

[19] Jorge L. Romeu and Aydin Ozturk. A new graphical test for multivariate normality. *American Journal of Mathematical and Management Sciences*, 16(1& 2):5–48, 1996.

[20] James R. Schott. Some tests for equality of covariance matrices. *Journal of Statistical Planning and Inference*, 94:25–36, 2001.

[21] Helge E. Solberg. Discriminant analysis. *Crit. Rev. Clin IV*, pages 209–242, 1978.

[22] GJ. Strewler, AA. Budayr, OH. Clark, and RA. Nissenson. Production of parathyroid hormone by a malignant nonparathyroid tumor in a hypercalcemic patient. *J Clin Endocrinol Metab*, 76:1373–5,1993.

[23] M. L. Tiku and N. Balakrishnan. Testing equality of population variances the robust way. *Commun. Statist.-Theor. Meth.*, 13(17):2143–2159, 1984.

[24] J. Walls, WA. Ratcliffe, A. Howell, and NJ. Bundred. Parathyroid hormone and parathyroid hormone-related protein in the investigation of hypercalcemia in two hospital populations. *Clin Endocrinol*, 41:407–413, 1994.

A Abbreviations

caa	calcium (cor)
cat	calcium
cai	ionized calcium
pho	phosphate
cre	creatinine
ure	urea
ura	urate
alp	ALP
bil	bilirubin
ast	AST
ala	ALAT
ld	LD
bic	total-CO_2
alb	albumin
pro	protein
glo	globulin
hgb	hemoglobin
iro	iron
leu	leucocyte count
sod	sodium
pot	potassium
age	age

Hypercalcaemia Study Populations	Year of study	Hospital population	Number of hypercalcaemia patients	Hypercalcaemia prevalence %	Duration of study (months)
Hvidovre 1	1984	10,000	257	2.57	6
Hvidovre 2	1989	5,400	129	2.39	3
Hvidovre 3	1994	4,500	87	1.93	2
Hvidovre 3-intact PTH	1994	4,500	31	Not relevant	2
Herlev	1987	9,000	162	1.80	6
Total		28,900	635	2.20	17
			Number of		
			donors		
Donor study	1991		183		3

Table 1: The table shows Hvidovre and Herlev hospital populations, study year, size of hospital population, number of patients with hypercalcaemia, prevalence of hypercalcaemia, and duration of studies (months).

Hvidovre Hospital	Number of beds	
Year of study	1984-89	1994
Infection Medicine	0	35
Reumatology	37	58
Hematology	29	0
Cardiology	59	51
Hepatology	29	0
Endocrinology	29	58
Gastroentorology	29	65
Nephrology	54	21
Neuromedicine	73	58
Surgical departments	299	297
Gynecology and obstetrics	129	124
Children's care unit	75	70
Intensive care unit	10	10
Ear, nose and throat unit	50	0
Medical emergency department	39	39
Surgical eye department	0	22
Psychiatry	0	20
Part-time beds	0	42
Total beds	941	1,070

Herlev Hospital	Number of beds
Year of study	1987
Cardiology and endocrinology	88
Gastroentorology	87
Nephrology	51
Oncology	96
Surgical departments	216
Gynaecology and obstetrics	78
Intensive care unit	12
Medical emergency department	12
Total beds	640

Table 2: The table shows the distribution of beds according to speciality in Hvidovre and Herlev hospitals, respectively.

WOMEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	47	2.73	0.20	32	2.74	0.14	14	2.71	0.10	47	2.64	0.04
cat	46	2.58	0.20	32	2.72	0.16	14	2.59	0.13	47	2.62	0.08
pho	44	1.18	0.26	28	0.98	0.21	11	1.25	0.24	39	1.28	0.34
cre	42	115.90	58.16	31	94.16	50.53	14	105.21	38.16	42	109.74	48.57
ure	44	10.10	6.58	31	5.89	3.18	14	7.98	4.59	42	9.47	6.20
ura	35	0.41	0.16	27	0.33	0.09	10	0.35	0.11	34	0.45	0.16
ALP	42	460.19	554.16	32	340.41	330.29	12	326.83	211.41	43	300.63	148.18
bil	46	29.98	77.46	31	6.81	3.26	13	7.23	2.31	40	8.80	5.87
AST	41	41.71	48.23	32	28.19	15.00	14	72.29	170.01	43	30.42	27.14
LD	38	507.08	269.70	27	347.59	72.53	12	415.75	172.85	43	376.28	102.89
bic	44	26.98	4.77	32	27.22	3.30	14	26.00	4.79	44	28.48	8.12
alb	46	468.54	77.40	32	583.48	74.19	14	492.01	98.39	47	584.96	67.12
pro	44	64.52	8.80	30	65.87	5.06	13	63.69	7.04	41	72.66	5.57
hgb	36	7.27	1.02	24	8.16	0.86	12	8.02	0.84	34	8.57	1.01
iro	36	14.24	10.52	28	14.53	7.05	11	13.26	7.62	34	14.44	6.08
leu	32	19.78	28.97	23	7.17	2.33	10	9.05	3.93	33	8.55	2.59
age	47	66.53	13.67	32	67.44	9.40	14	52.00	25.61	47	62.19	13.39
MEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	31	2.87	0.33	11	2.72	0.16	10	2.69	0.06	17	2.66	0.12
cat	31	2.70	0.33	11	2.69	0.14	10	2.68	0.11	17	2.60	0.11
pho	29	1.10	0.26	11	0.91	0.13	8	1.17	0.33	11	1.21	0.39
cre	31	122.03	62.95	11	117.45	47.55	10	104.90	37.42	15	144.73	88.33
ure	29	11.16	8.07	10	6.57	3.05	9	7.46	4.09	16	11.46	6.81
ura	23	0.49	0.16	5	0.41	0.09	8	0.41	0.10	10	0.47	0.17
ALP	28	504.61	305.47	10	241.30	80.60	9	496.56	501.33	16	289.56	124.04
bil	30	13.80	12.20	11	8.09	3.27	10	8.00	3.50	17	16.24	15.01
AST	29	49.59	82.99	11	27.36	18.40	9	59.78	82.69	16	42.69	43.57
LD	27	587.63	727.50	11	323.00	30.97	9	384.67	149.36	12	378.67	88.08
bic	31	29.39	7.31	9	27.56	2.74	9	26.78	5.63	15	27.73	3.75
alb	31	441.03	71.71	11	571.00	54.55	10	597.40	85.60	17	548.29	92.60
pro	29	66.31	9.19	10	65.30	4.74	10	73.30	4.57	13	75.46	7.62
hgb	24	7.38	1.27	8	8.45	0.63	9	8.90	1.14	11	8.67	1.55
iro	18	16.94	15.72	8	15.76	10.87	8	12.78	5.82	11	13.74	11.46
leu	25	17.72	37.23	8	8.65	2.74	8	9.71	3.57	11	8.54	3.71
age	31	70.45	9.73	11	66.00	11.84	10	46.50	16.65	17	54.29	17.42

Table 3: Number (N), mean, and standard deviation (SD) for blood analyses and age measured in women and men in the Hvidovre 1 population.

WOMEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	60	2.99	0.41	25	2.88	0.19	9	2.83	0.13	3	2.78	0.05
cat	60	2.80	0.39	25	2.83	0.23	9	2.76	0.10	3	2.58	0.10
cai	46	1.59	0.25	23	1.58	0.14	6	1.52	0.11	3	1.22	0.12
pho	56	1.15	0.31	24	0.89	0.14	9	1.09	0.44	3	0.86	0.20
cre	59	102.59	43.62	25	83.88	22.90	9	80.44	53.08	3	108.67	42.52
ure	58	9.68	9.48	25	5.76	2.86	9	7.72	5.51	3	9.47	4.20
ura	52	0.49	0.39	23	0.37	0.09	8	0.39	0.26	2	0.42	0.20
ALP	59	696.75	588.81	23	290.52	169.87	9	253.11	117.25	3	967.67	843.76
bil	56	27.68	78.66	21	6.71	3.02	9	51.67	112.09	3	16.00	19.16
AST	55	64.13	62.33	23	35.52	67.04	9	36.11	16.53	3	49.00	57.30
ala	52	37.21	45.83	19	32.68	67.96	8	33.75	24.18	2	20.00	9.90
LD	59	807.14	826.95	25	350.84	70.43	8	544.88	247.53	3	852.00	738.56
bic	59	24.14	3.04	24	24.13	2.31	9	23.22	2.11	3	21.67	5.03
alb	60	479.35	99.49	25	594.80	86.50	9	575.67	49.72	3	472.33	114.03
pro	59	64.97	8.47	24	66.46	6.13	9	61.78	4.87	3	65.67	9.29
glo	58	32.16	8.52	23	25.98	5.50	9	22.06	6.16	3	33.03	2.74
hgb	53	7.23	1.26	17	8.38	0.85	8	7.78	0.95	3	7.20	1.06
iro	53	11.35	9.96	20	13.33	5.04	6	15.80	13.47	3	14.27	10.36
leu	53	10.39	5.51	18	10.54	6.25	7	12.30	6.80	3	14.13	7.36
sod	58	138.59	4.06	25	139.08	5.30	9	141.89	3.92	3	136.67	8.74
pot	55	4.04	0.65	24	4.38	0.50	9	4.17	0.34	3	3.93	0.58
age	60	64.02	10.18	25	66.88	14.68	9	53.20	13.24	3	62.47	14.97
MEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	26	2.99	0.33	4	2.78	0.07	5	3.03	0.45	5	2.73	0.04
cat	26	2.79	0.34	4	2.76	0.13	5	2.92	0.35	5	2.56	0.13
cai	15	1.60	0.23	4	1.49	0.10	5	1.69	0.22	5	1.30	0.02
pho	25	1.02	0.35	4	0.89	0.24	5	1.28	0.35	5	1.80	0.72
cre	26	117.92	71.49	4	112.25	20.69	3	146.67	62.00	5	134.80	87.25
ure	23	10.00	6.83	4	9.08	2.72	4	16.88	6.53	5	10.82	7.04
ura	24	0.42	0.18	4	0.62	0.24	4	0.39	0.16	5	0.57	0.18
ALP	25	744.96	978.02	3	224.00	87.79	5	507.00	329.14	4	251.75	79.90
bil	24	33.17	59.99	2	8.50	0.71	5	81.60	128.75	5	11.80	8.73
AST	26	56.58	83.30	4	21.00	19.63	5	114.00	141.49	5	23.80	18.90
ala	23	40.13	66.10	3	17.33	6.81	3	219.33	173.30	4	20.75	17.21
LD	26	631.46	377.48	4	395.75	153.31	5	445.80	151.49	3	304.00	79.87
bic	26	24.58	3.00	4	25.75	1.26	5	21.40	3.78	5	22.20	7.29
alb	26	470.35	78.08	4	615.00	72.66	5	545.00	90.99	5	496.40	115.95
pro	26	64.12	7.73	4	66.25	7.14	5	69.60	6.19	5	63.00	5.92
glo	24	31.31	5.89	4	23.83	6.62	5	32.00	9.60	5	28.74	6.47
hgb	22	7.19	1.16	4	8.85	1.74	3	7.10	1.95	4	6.68	1.59
iro	21	7.24	5.68	3	15.83	3.63	4	18.35	5.86	4	11.40	4.67
leu	20	12.03	5.68	4	7.95	2.28	3	13.40	7.94	4	13.50	6.22
sod	26	137.77	5.58	4	138.75	2.63	5	139.80	4.82	5	135.40	6.07
pot	25	4.08	0.51	4	4.18	0.25	4	4.48	0.52	5	4.84	0.76
age	26	62.10	11.96	4	64.91	8.86	5	41.27	21.21	5	65.11	6.14

Table 4: Number (N), mean, and standard deviation (SD) for blood analyses and age measured in women and men in the Herlev population.

WOMEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	16	2.82	0.31	26	2.69	0.10	7	2.76	0.20	6	2.63	0.06
cat	16	2.60	0.30	26	2.68	0.11	7	2.69	0.24	6	2.57	0.12
pho	16	1.09	0.25	26	0.94	0.21	7	1.50	0.55	6	1.30	0.17
ure	16	10.04	6.32	26	6.00	2.63	7	23.43	12.26	6	9.03	2.80
ALP	16	528.88	696.79	26	246.35	88.87	7	676.43	708.36	6	215.33	76.29
LD	16	983.06	1450.1	26	336.46	88.95	7	418.43	206.35	6	345.67	47.73
alb	16	398.13	82.11	26	587.31	52.89	7	534.86	78.79	6	551.50	104.86
pro	16	70.81	20.38	26	68.23	4.04	7	70.14	7.40	6	68.33	8.71
hgb	16	6.82	1.09	25	8.13	0.80	7	7.34	0.81	5	7.64	1.06
leu	16	11.66	6.68	25	7.45	2.08	7	9.67	4.49	5	8.66	2.29
age	16	73.75	8.11	26	69.23	10.06	7	66.57	11.03	6	73.50	6.16
MEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	14	2.87	0.40	2	2.67	0.11	4	2.60	0.04	8	2.60	0.02
cat	14	2.71	0.39	2	2.61	0.08	4	2.55	0.04	8	2.58	0.10
pho	14	1.10	0.27	2	1.00	0.28	4	1.28	0.31	8	1.07	0.19
ure	14	9.64	4.09	2	4.95	1.20	4	10.63	3.98	8	9.29	5.96
ALP	14	313.79	182.69	2	159.00	63.64	4	207.50	102.09	8	369.63	360.20
LD	14	705.71	1007.6	2	265.00	15.56	4	342.75	46.46	7	333.86	71.11
alb	14	459.71	120.42	2	542.00	25.46	4	551.00	63.18	8	580.38	89.10
pro	14	74.00	18.02	2	64.00	1.41	4	73.25	10.28	8	74.88	6.31
hgb	12	7.30	1.49	2	9.15	0.07	4	8.80	1.62	8	8.65	0.65
leu	12	8.45	4.59	2	8.65	2.76	4	12.78	2.96	8	9.71	1.18
age	14	65.86	11.57	2	60.50	17.68	4	52.25	10.14	8	58.88	10.44

Table 5: Number (N), mean, and standard deviation (SD) for blood analyses and age measured in women and men in the Hvidovre 2 population.

WOMEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	8	2.77	0.22	21	2.73	0.19	6	2.77	0.24	9	2.61	0.02
cat	8	2.60	0.24	21	2.72	0.20	6	2.69	0.32	9	2.63	0.12
cai	6	1.45	0.19	18	1.48	0.16	5	1.50	0.17	8	1.36	0.05
pho	8	1.19	0.58	21	0.95	0.15	6	1.18	0.20	9	1.16	0.19
pth	5	1.22	1.67	15	6.57	5.28	5	0.96	0.88	7	2.43	1.88
cre	8	99.88	58.33	21	81.67	22.49	6	112.67	48.84	9	87.67	44.38
ure	7	9.17	9.52	20	6.88	3.50	5	9.54	4.19	9	7.70	8.60
ura	8	0.39	0.22	21	0.35	0.08	6	0.38	0.12	9	0.36	0.09
ALP	8	547.88	531.31	21	279.86	167.85	6	239.00	93.45	9	272.22	173.54
bil	8	12.88	6.77	21	9.10	4.37	6	5.67	1.63	9	9.67	5.41
LD	8	717.75	406.46	21	328.71	123.28	6	289.83	32.86	9	456.22	188.55
bic	8	27.25	4.50	21	29.43	2.80	6	29.00	3.29	9	27.33	7.21
alb	8	451.25	85.54	21	593.43	62.21	6	573.17	98.43	9	616.67	94.36
pro	8	68.38	9.66	21	69.10	4.25	6	71.00	7.13	9	71.11	8.25
hgb	8	6.88	0.93	21	8.31	1.01	6	7.57	1.02	8	8.50	1.10
pot	8	4.08	0.47	20	4.02	0.44	6	3.98	0.31	9	3.97	0.55
sod	8	132.63	8.18	20	139.40	5.07	6	139.50	4.42	9	138.00	7.58
age	8	70.13	14.98	21	70.76	10.85	6	66.33	9.97	9	59.89	12.29
MEN	Malignancy			PHPT			Miscellaneous			Transient		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
caa	10	2.64	0.06	6	2.84	0.16	2	2.95	0.35	7	2.63	0.03
cat	10	2.50	0.12	6	2.83	0.19	2	3.01	0.41	7	2.51	0.10
cai	9	1.44	0.16	5	1.54	0.17	1	1.47	.	6	1.39	0.09
pho	10	1.13	0.23	6	1.01	0.10	2	1.03	0.83	7	1.27	0.28
pth	8	1.76	0.86	3	8.37	3.71	1	0.70	.	4	2.43	1.95
cre	10	119.80	76.61	6	94.33	12.97	2	267.50	47.38	7	109.14	24.07
ure	9	8.97	7.03	6	6.57	2.22	2	6.85	8.70	7	11.84	6.83
ura	10	0.37	0.17	6	0.40	0.06	2	0.26	0.25	7	0.50	0.05
ALP	10	452.40	313.44	6	295.17	79.56	2	219.50	119.50	7	450.86	308.45
bil	10	15.40	17.90	6	7.83	2.71	2	5.50	0.71	7	13.29	11.09
LD	10	580.00	564.55	6	247.17	28.44	2	453.50	139.30	7	503.43	521.56
bic	10	28.60	4.90	6	25.00	3.10	2	32.00	2.83	7	25.00	5.20
alb	10	468.50	110.44	6	593.50	97.08	2	650.50	58.69	7	495.14	106.34
pro	10	70.80	14.59	6	69.67	5.54	2	77.50	14.85	7	69.29	11.16
hgb	10	7.36	1.45	6	9.12	0.94	2	7.35	0.64	7	7.34	0.94
pot	10	4.48	0.87	6	4.32	0.42	2	3.20	1.13	7	4.19	0.59
sod	10	136.80	6.49	6	138.50	2.88	2	145.00	4.24	7	141.29	5.88
age	10	63.90	11.88	6	50.17	17.44	2	45.50	2.12	7	75.71	7.95

Table 6: Number (N), mean, and standard deviation (SD) for blood analyses and age measured in women and men in the Hvidovre 3 population.

		Herlev		Hvidovre	
	N	Mean	SD	Mean	SD
cat	163	2.33	0.10	2.40	0.11
pho	164	1.08	0.15	1.09	0.16
cre	158	81.20	14.92	81.86	13.58
ure	164	4.61	1.14	4.62	1.16
ura	160	0.33	0.09	0.31	0.08
ALP	159	154.11	47.13	152.29	51.69
bil	157	8.42	4.01	7.45	4.07
AST	154	19.94	10.89	19.86	9.44
LD	163	312.23	57.04	297.12	49.26
bic	163	23.04	1.79	27.15	2.74
alb	159	666.57	34.83	618.10	38.52
pro	159	67.25	3.31	66.48	3.21
iro	157	20.96	8.36	20.62	8.27
pot	164	4.04	0.32	3.95	0.28
sod	166	140.90	2.46	141.31	2.27

Table 7: Number (N), mean, and standard deviation (SD) for the donor material. For observations with valid measurements in both hospitals.

			alb	bas	fos	hgb	car	pro	ldh	age
Mali	Herl		0.3054	0.0001	0.0594	0.3090	0.0001	0.1935	0.0001	0.2182
		log	0.0311	0.1498	0.0003	0.0001	0.0015	0.6709	0.3633	0.0001
	Hvi1		0.1409	0.0001	0.5826	0.2068	0.0001	0.0001	0.0001	0.1174
		log	0.1630	0.0070	0.9396	0.0831	0.2013	0.0088	0.0694	0.0001
	Hvi2		0.3463	0.0001	0.0429	0.8619	0.0088	0.0001	0.0001	0.0775
		log	0.5293	0.0012	0.2162	0.4938	0.6365	0.0020	0.0017	0.0096
	Hvi3		0.1818	0.0001	0.0001	0.0655	0.0001	0.2592	0.0008	0.2513
		log	0.4983	0.5873	0.0143	0.1636	0.0185	0.8481	0.0478	0.0766
PHPT	Herl		0.2804	0.0728	0.9863	0.8801	0.1126	0.6459	0.2322	0.3256
		log	0.0562	0.2010	0.7986	0.5432	0.6167	0.4944	0.6073	0.0946
	Hvi1		0.9996	0.0001	0.5466	0.8862	0.0034	0.7172	0.5768	0.0751
		log	0.9658	0.0001	0.6862	0.7557	0.2437	0.5514	0.6060	0.0182
	Hvi2		0.1772	0.1790	0.6365	0.5078	0.0001	0.2487	0.3679	0.1420
		log	0.0937	0.5962	0.5314	0.2333	0.1479	0.1503	0.0189	0.0104
	Hvi3		0.7620	0.0001	0.8647	0.4346	0.0002	0.3350	0.0001	0.0108
		log	0.1795	0.1229	0.9181	0.3072	0.1186	0.1969	0.0101	0.0001
Misc	Herl		0.0277	0.0029	0.9124	0.7712	0.1733	0.9313	0.3776	0.7105
		log	0.0157	0.0787	0.4986	0.7466	0.3861	0.9580	0.8417	0.5981
	Hvi1		0.7828	0.0047	0.4197	0.2145	0.0007	0.6469	0.0004	0.1798
		log	0.6919	0.6339	0.1828	0.0915	0.2093	0.4344	0.0255	0.1517
	Hvi2		0.0664	0.0001	0.1210	0.1490	0.1071	0.2838	0.0001	0.2342
		log	0.0464	0.0612	0.3692	0.3440	0.9973	0.5019	0.0060	0.0926
	Hvi3		.	.	.	.	.	.	.	.
		log	.	.	.	.	.	.	.	.
Tran	Herl		.	.	.	.	.	.	.	.
		log	.	.	.	.	.	.	.	.
	Hvi1		0.4025	0.0034	0.6248	0.5048	0.0009	0.1437	0.0001	0.2534
		log	0.2186	0.3638	0.7041	0.1808	0.1346	0.2932	0.0035	0.0118
	Hvi2		0.0055	0.0001	0.9814	0.4028	0.2980	0.0172	0.0171	0.1111
		log	0.0016	0.0027	0.9077	0.2190	0.7921	0.0044	0.0457	0.0224
	Hvi3		0.9064	0.0011	0.1617	0.6010	0.0008	0.7202	0.0001	0.6171
		log	0.6664	0.0407	0.3772	0.7229	0.0701	0.6498	0.0185	0.1502

Table 8: Univariate tests for observations complete in all variables

		Herlev				Hvidovre 1			
			mard-	mard-	henze-		mard-	mard-	henze-
		nobs	skew	kurt	zir	nobs	skew	kurt	zir
Malig-		63	0.0001	0.0001	0.0001	43	0.0001	0.0001	0.0002
nancy	log	63	0.0001	0.0001	0.0001	43	0.0001	0.0015	0.8477
PHPT		18	0.3157	0.1616	0.4026	21	0.0051	0.6906	0.1980
	log	18	0.3650	0.1446	0.4366	21	0.1008	0.4206	0.5198
Miscel-		10	0.2137	0.1225	0.1320	12	0.3476	0.1055	0.2514
laneous	log	10	0.2464	0.1013	0.0167	12	0.4839	0.0862	0.2473
Tran-		.	.	.	.	25	0.0017	0.7450	0.0001
sient	log	.	.	.	.	25	0.0039	0.9274	0.0026
		Hvidovre 2				Hvidovre 3			
			mard-	mard-	henze-		mard-	mard-	henze-
		nobs	skew	kurt	zir	nobs	skew	kurt	zir
Malig-		28	0.0001	0.1838	0.0180	16	0.0027	0.5529	0.0158
nancy	log	28	0.5993	0.1901	0.3448	16	0.0741	0.2440	0.0595
PHPT		27	0.0001	0.1831	0.0152	26	0.0001	0.3591	0.0277
	log	27	0.0001	0.3378	0.0550	26	0.0284	0.8265	0.7170
Miscel-		11	0.5306	0.0721	0.1726	.	.	.	.
laneous	log	11	0.4195	0.1044	0.2776	.	.	.	.
Tran-		12	0.2001	0.2025	0.5371	15	0.0493	0.3494	0.0800
sient	log	12	0.2566	0.1540	0.5717	15	0.2633	0.1837	0.1823

Table 9: 8 variables, test for normality, for transformed and log transformed variables

			mard-	mard-	henze-
		nobs	skew	kurt	zir
Malignancy		87	0.0001	0.0001	0.0001
	log	87	0.0002	0.4046	0.5549
PHPT		74	0.0001	0.0001	0.0001
	log	74	0.0022	0.2639	0.0089
Miscellaneous		30	0.0001	0.0066	0.0009
	log	30	0.2186	0.7123	0.5097
Transient		52	0.0001	0.0009	0.0001
	log	52	0.0555	0.6964	0.0001

Table 10: 8 variables only Hvidovre

			mard-	mard-	henze-
		nobs	skew	kurt	zir
Malignancy	All	150	0.0001	0.0001	0.0001
	Hvi	87	0.0002	0.4046	0.5549
PHPT	All	92	0.0001	0.0072	0.0034
	Hvi	74	0.0022	0.2639	0.0089
Miscellaneous	All	40	0.0002	0.0809	0.0003
	Hvi	30	0.2186	0.7123	0.5097
Transient	All	57	0.0023	0.2136	0.0001
	Hvi	52	0.0555	0.6964	0.0001

Table 11: Hvidovre population compared to all, only logged variables. Malignancy (Mali), miscellaneous (Misc), transient (Tran), Hvidovre (Hvid), Herlev (Herl).

	PHPT	Mali	Misc	Tran
lnpho lnbas lnalb	0.3907	0.6279	0.3293	0.0013
+lnhgb	0.5764	0.8547	0.1494	0.0009
+lncat	0.0100	0.0370	0.0240	0.0001
+lncar	0.4449	0.4391	0.0057	0.0070
+lnpro	0.1255	0.0001	0.1061	0.0001
+lnld	0.0037	0.1881	0.2403	0.0001
+lnage	0.0324	0.5764	0.3985	0.0009
+lnhgb lncar	0.6138	0.7392	0.0058	0.0055
+lncat lnhgb	0.1067	0.1583	0.0005	0.0001
+lncat lncar	0.0475	0.0133	0.0007	0.0001
+lncat lnpro	0.0106	0.0001	0.0022	0.0001
+lncat lnld	0.0001	0.0073	0.0411	0.0001
+lncat lnage	0.0001	0.0751	0.0332	0.0001
+lnhgb lnpro	0.1952	0.0006	0.0139	0.0001
+lnhgb lnld	0.0037	0.2540	0.0327	0.0001
+lnhgb lnage	0.1862	0.6671	0.0154	0.0030
+lnpro lncar	0.2863	0.0001	0.0005	0.0006
+lnpro lnld	0.0031	0.0001	0.0592	0.0001
+lnpro lnage	0.0072	0.0001	0.2502	0.0001
+lncar lnld	0.0128	0.1134	0.0254	0.0001
+lncar lnage	0.0526	0.2493	0.0115	0.0002
+lnld lnage	0.0014	0.1545	0.4052	0.0001

Table 12: Test for equality of covariance matrices, subsets with 3-5 variables. Malignancy (Mali), miscellaneous (Misc), transient (Tran).

	Two-way ANOVA			One-way ANOVA	
	kn_kn interaction	kn_n population	kn_k gender	n_n population	k_k gender
lnpho lnalp lnalb	0.2765	0.0017	0.1730	0.0022	0.2466
+lnhgb	0.8196	0.0129	0.0448	0.0148	0.0529
+lncat	0.3828	0.0001	0.1992	0.0001	0.3715
+lncar	0.4518	0.0117	0.3564	0.0139	0.4375
+lnpro	0.3893	0.0006	0.2762	0.0005	0.2670
+lnld	0.5264	0.0025	0.3581	0.0032	0.4737
+lnage	0.6733	0.0090	0.7175	0.0084	0.7053
+lnhgb lncar	0.9095	0.0558	0.1559	0.0544	0.1508
+lncat lnhgb	0.8334	0.0002	0.0324	0.0005	0.0731
+lncat lncar	0.4857	0.0007	0.4239	0.0010	0.5825
+lncat lnpro	0.4704	0.0001	0.2742	0.0001	0.3622
+lncat lnld	0.5671	0.0001	0.3430	0.0001	0.5911
+lncat lnage	0.2353	0.0001	0.2769	0.0001	0.5039
+lnhgb lnpro	0.9075	0.0049	0.0789	0.0045	0.0716
+lnhgb lnld	0.9361	0.0049	0.0804	0.0052	0.0876
+lnhgb lnage	0.4311	0.0090	0.0705	0.0111	0.0918
+lnpro lncar	0.5437	0.0050	0.4339	0.0043	0.3895
+lnpro lnld	0.6095	0.0009	0.5155	0.0008	0.5120
+lnpro lnage	0.2684	0.0004	0.3623	0.0004	0.3811
+lncar lnld	0.7697	0.0145	0.5715	0.0166	0.6552
+lncar lnage	0.2811	0.0131	0.4540	0.0164	0.5713
+lnld lnage	0.3574	0.0027	0.4216	0.0038	0.5862
+lnhgb lncar lnpro	0.9548	0.0227	0.2203	0.0185	0.1766
+lncat lnhgb lncar	0.8939	0.0037	0.1570	0.0048	0.2157
+lncat lnhgb lnpro	0.9029	0.0001	0.0512	0.0002	0.0846
+lncat lnhgb lnld	0.9227	0.0001	0.0522	0.0001	0.1033
+lncat lnhgb lnage	0.4774	0.0003	0.0517	0.0006	0.1188
+lncat lnpro lncar	0.5413	0.0004	0.4733	0.0004	0.5099
+lncat lnpro lnld	0.6207	0.0001	0.4613	0.0001	0.6005
+lncat lnpro lnage	0.3351	0.0001	0.3538	0.0001	0.4777
+lncat lncar lnld	0.7342	0.0006	0.6284	0.0007	0.7718
+lncat lncar lnage	0.3096	0.0011	0.5176	0.0016	0.6976
+lncat lnld lnage	0.3938	0.0001	0.4070	0.0001	0.6894
+lnhgb lncar lnld	0.9748	0.0212	0.2536	0.0193	0.2319
+lnhgb lncar lnage	0.5876	0.0563	0.2117	0.0582	0.2220
+lnhgb lnpro lnld	0.9620	0.0019	0.1349	0.0016	0.1143
+lnhgb lnpro lnage	0.6078	0.0037	0.1128	0.0037	0.1129
+lnhgb lnld lnage	0.6662	0.0042	0.1057	0.0049	0.1294
+lncar lnpro lnld	0.8167	0.0067	0.6756	0.0057	0.6256
+lncar lnpro lnage	0.3945	0.0058	0.5176	0.0054	0.5015
+lncar lnld lnage	0.5726	0.0183	0.6290	0.0223	0.7440
+lnpro lnld lnage	0.4705	0.0010	0.5703	0.0010	0.6095

Table 13: Multivariate test for equality of mean vectors (Wilks' test) for malignancy. The first column labeled kn_kn is test for interaction, the second column labeled kn_n is test for population effect in an additive model, the third column (kn_k) tests for gender effect in the additive model. The fourth and fifth model test for population and gender effects when no other effects are present.

	Two-way ANOVA			One-way ANOVA	
	kn_kn inter-action	kn_n popu-lation	kn_k gen-der	n_n popu-lation	k_k gen-der
lnpho lnalp lnalb	0.8274	0.5875	0.7503	0.6205	0.8359
+lnhgb	0.8004	0.6286	0.0134	0.6482	0.0130
+lncat	0.7199	0.0593	0.8162	0.0625	0.8737
+lncar	0.7442	0.6752	0.6131	0.6892	0.6443
+lnpro	0.9217	0.0181	0.6875	0.0219	0.8305
+lnld	0.7764	0.3115	0.4810	0.3116	0.4862
+lnage	0.5126	0.0080	0.0167	0.0113	0.0255
+lnhgb lncar	0.7666	0.7224	0.0131	0.7141	0.0109
+lncat lnhgb	0.6711	0.1212	0.0263	0.1234	0.0253
+lncat lncar	0.5345	0.1107	0.7113	0.1105	0.7292
+lncat lnpro	0.8646	0.0011	0.6676	0.0014	0.7893
+lncat lnld	0.7620	0.01810	0.5955	0.0176	0.6061
+lncat lnage	0.2475	0.1224	0.0349	0.1295	0.0364
+lnhgb lnpro	0.9079	0.0320	0.0209	0.0297	0.0176
+lnhgb lnld	0.7921	0.4486	0.0147	0.4687	0.0145
+lnhgb lnage	0.3616	0.7330	0.0001	0.7796	0.0001
+lnpro lncar	0.8349	0.0299	0.4487	0.0342	0.5292
+lnpro lnld	0.9082	0.0083	0.3878	0.0100	0.4840
+lnpro lnage	0.3358	0.0451	0.0540	0.0490	0.0592
+lncar lnld	0.7496	0.1970	0.2530	0.2141	0.2871
+lncar lnage	0.3314	0.7069	0.0102	0.7320	0.0103
+lnld lnage	0.3861	0.4704	0.0505	0.4606	0.0448
+lnhgb lncar lnpro	0.8342	0.0592	0.0128	0.0497	0.0089
+lncat lnhgb lncar	0.5639	0.2499	0.0243	0.2383	0.0201
+lncat lnhgb lnpro	0.8380	0.00290	0.0338	0.0025	0.0285
+lncat lnhgb lnld	0.7639	0.0643	0.0281	0.0671	0.0282
+lncat lnhgb lnage	0.3930	0.1931	0.0002	0.2157	0.0002
+lncat lnpro lncar	0.6826	0.0027	0.4888	0.0029	0.5506
+lncat lnpro lnld	0.8891	0.0003	0.4279	0.0004	0.5324
+lncat lnpro lnage	0.3503	0.0041	0.0759	0.0045	0.0866
+lncat lncar lnld	0.6368	0.0124	0.3584	0.0132	0.3955
+lncat lncar lnage	0.2621	0.1665	0.0203	0.1748	0.0206
+lncat lnld lnage	0.3939	0.0436	0.0815	0.0420	0.0760
+lnhgb lncar lnld	0.7904	0.3265	0.0084	0.3414	0.0081
+lnhgb lncar lnage	0.4460	0.7373	0.0001	0.7602	0.0001
+lnhgb lnpro lnld	0.9179	0.0255	0.0201	0.0233	0.0166
+lnhgb lnpro lnage	0.4671	0.0612	0.0003	0.0612	0.0003
+lnhgb lnld lnage	0.5457	0.6406	0.0004	0.6802	0.0004
+lncar lnpro lnld	0.8509	0.0033	0.123	0.00510	0.2001
+lncar lnpro lnage	0.3871	0.0523	0.0196	0.0579	0.0216
+lncar lnld lnage	0.4613	0.2835	0.0136	0.3231	0.0164
+lnpro lnld lnage	0.4797	0.0203	0.0983	0.0207	0.1008

Table 14: Multivariate test for equality of mean vectors (Wilks' test) for PHPT. The first column labeled kn_kn is test for interaction, the second column labeled kn_n is test for population effect in an additive model, the third column (kn_k) tests for gender effect in the additive model. The fourth and fifth model test for population and gender effects when no other effects are present. There is no real extra information in these two columns.

	Two-way ANOVA			One-way ANOVA	
	kn_kn inter- action	kn_n popu- lation	kn_k gen- der	n_n popu- lation	k_k gen- der
lnpho lnalp lnalb	0.2726	0.8330	0.6110	0.8528	0.6614
+lnhgb	0.1117	0.2874	0.6240	0.2534	0.5523
+lncat	0.1792	0.5498	0.6540	0.5874	0.7493
+lncar	0.0696	0.3173	0.6010	0.3173	0.6252
+lnpro	0.3782	0.0507	0.0235	0.1258	0.0835
+lnld	0.1816	0.4163	0.6475	0.4385	0.7213
+lnage	0.2752	0.0528	0.0139	0.0828	0.0238
+lnhgb lncar	0.0776	0.0799	0.5546	0.0692	0.5193
+lncat lnhgb	0.0941	0.3523	0.7503	0.3177	0.6973
+lncat lncar	0.0299	0.2077	0.6121	0.2199	0.6872
+lncat lnpro	0.2389	0.0210	0.0475	0.0475	0.1376
+lncat lnld	0.1124	0.3649	0.5890	0.4071	0.7061
+lncat lnage	0.3533	0.4871	0.0534	0.5666	0.0683
+lnhgb lnpro	0.2000	0.0218	0.0345	0.0401	0.0743
+lnhgb lnld	0.0724	0.0907	0.6845	0.0727	0.6059
+lnhgb lnage	0.1607	0.3189	0.2325	0.3007	0.2020
+lnpro lncar	0.1390	0.0280	0.0285	0.0467	0.0537
+lnpro lnld	0.2521	0.0525	0.0421	0.1415	0.1737
+lnpro lnage	0.5718	0.0794	0.0009	0.1179	0.0013
+lncar lnld	0.0962	0.1289	0.6093	0.1276	0.6438
+lncar lnage	0.1537	0.2688	0.1936	0.3132	0.2452
+lnld lnage	0.3806	0.4275	0.0240	0.4643	0.0236
+lnhgb lncar lnpro	0.1590	0.0078	0.0578	0.0100	0.0784
+lncat lnhgb lncar	0.0646	0.1272	0.6820	0.1118	0.6532
+lncat lnhgb lnpro	0.1359	0.0252	0.0628	0.0440	0.1287
+lncat lnhgb lnld	0.0484	0.1622	0.7675	0.1367	0.7104
+lncat lnhgb lnage	0.1307	0.4045	0.3439	0.3816	0.3045
+lncat lnpro lncar	0.0611	0.0138	0.0530	0.0204	0.0852
+lncat lnpro lnld	0.1438	0.0383	0.0700	0.0948	0.2377
+lncat lnpro lnage	0.3937	0.0456	0.0022	0.0650	0.0029
+lncat lncar lnld	0.0372	0.1350	0.5632	0.1417	0.6288
+lncat lncar lnage	0.0731	0.2638	0.2774	0.3093	0.3580
+lncat lnld lnage	0.2528	0.4009	0.0353	0.4591	0.0409
+lnhgb lncar lnld	0.0979	0.0151	0.5028	0.0133	0.5141
+lnhgb lncar lnage	0.1333	0.1141	0.4852	0.1037	0.4652
+lnhgb lnpro lnld	0.1321	0.0186	0.0665	0.0343	0.1450
+lnhgb lnpro lnage	0.2651	0.0413	0.0150	0.0502	0.0168
+lnhgb lnld lnage	0.1476	0.1389	0.2171	0.1180	0.1696
+lncar lnpro lnld	0.1615	0.0259	0.0463	0.0487	0.1048
+lncar lnpro lnage	0.2236	0.0509	0.0102	0.0654	0.0123
+lncar lnld lnage	0.2056	0.1780	0.1740	0.1902	0.1879
+lnpro lnld lnage	0.4464	0.0870	0.0012	0.1599	0.0026

Table 15: Multivariate test for equality of mean vectors (Wilks' test) for miscellaneous. The first column labeled kn_kn is test for interaction, the second column labeled kn_n is test for population effect in an additive model, the third column (kn_k) tests for gender effect in the additive model. The fourth and fifth model test for population and gender effects when no other effects are present. There is no real extra information in these two columns.

	Two-way ANOVA			One-way ANOVA	
	kn_kn inter- action	kn_n popu- lation	kn_k gen- der	n_n popu- lation	k_k gen- der
lnpho lnalp lnalb	0.0013	0.6153	0.3343	0.5094	0.2135
+lnhgb	0.0048	0.0777	0.3933	0.0601	0.2978
+lncat	0.0023	0.1918	0.4299	0.1609	0.3425
+lncar	0.0036	0.7220	0.3119	0.6913	0.2668
+lnpro	0.0026	0.0976	0.3342	0.1036	0.3697
+lnld	0.0097	0.4101	0.2414	0.3800	0.2036
+lnage	0.0254	0.3010	0.5569	0.2724	0.4982
+lnhgb lncar	0.0121	0.1631	0.4137	0.1398	0.3461
+lncat lnhgb	0.0045	0.0164	0.4138	0.0159	0.4308
+lncat lncar	0.0064	0.2837	0.4191	0.2710	0.3952
+lncat lnpro	0.0043	0.0053	0.4674	0.0051	0.4856
+lncat lnld	0.0158	0.1798	0.2813	0.1911	0.3093
+lncat lnage	0.0012	0.1967	0.5282	0.1796	0.4824
+lnhgb lnpro	0.0062	0.0994	0.4390	0.0707	0.3018
+lnhgb lnld	0.0133	0.2710	0.1614	0.2732	0.1579
+lnhgb lnage	0.0018	0.0601	0.4900	0.0511	0.4333
+lnpro lncar	0.0067	0.2092	0.4418	0.2254	0.4982
+lnpro lnld	0.0100	0.2078	0.3454	0.2009	0.3323
+lnpro lnage	0.0024	0.1014	0.4746	0.0982	0.4733
+lncar lnld	0.0248	0.6029	0.2091	0.6104	0.2075
+lncar lnage	0.0025	0.4774	0.3823	0.4863	0.3946
+lnld lnage	0.0040	0.3482	0.3334	0.3382	0.3160
+lnhgb lncar lnpro	0.0187	0.2437	0.5810	0.2010	0.4720
+lncat lnhgb lncar	0.0149	0.0397	0.4883	0.0360	0.4717
+lncat lnhgb lnpro	0.0113	0.0074	0.5471	0.0048	0.4151
+lncat lnhgb lnld	0.0102	0.1039	0.1233	0.1511	0.2116
+lncat lnhgb lnage	0.0019	0.0174	0.5222	0.0176	0.5643
+lncat lnpro lncar	0.0112	0.0163	0.5585	0.0158	0.5812
+lncat lnpro lnld	0.0142	0.0254	0.4709	0.0219	0.4334
+lncat lnpro lnage	0.0046	0.0047	0.6012	0.0040	0.5731
+lncat lncar lnld	0.0410	0.3267	0.2706	0.3539	0.3088
+lncat lncar lnage	0.0044	0.2173	0.4975	0.2233	0.5255
+lncat lnld lnage	0.0067	0.2016	0.3861	0.2162	0.4310
+lnhgb lncar lnld	0.0340	0.4276	0.2031	0.4258	0.1926
+lnhgb lncar lnage	0.0051	0.0943	0.5265	0.0807	0.4676
+lnhgb lnpro lnld	0.0069	0.2893	0.2843	0.2625	0.2390
+lnhgb lnpro lnage	0.0031	0.0876	0.5704	0.0543	0.3608
+lnhgb lnld lnage	0.0064	0.2332	0.2236	0.2487	0.2452
+lncar lnpro lnld	0.0227	0.3725	0.4349	0.3705	0.4339
+lncar lnpro lnage	0.0079	0.1462	0.5744	0.1444	0.5861
+lncar lnld lnage	0.0131	0.4372	0.2882	0.4556	0.3084
+lnpro lnld lnage	0.0092	0.2213	0.4718	0.1960	0.4100

Table 16: Multivariate test for equality of mean vectors (Wilks' test) for transient. The first column labeled kn_kn is test for interaction, the second column labeled kn_n is test for population effect in an additive model & the third column (kn_k) tests for gender effect in the additive model. The fourth and fifth model test for population and gender effects when no other effects are present. There is no real extra information in these two columns.

	Hvidovre 1	Herlev	Hvidovre 2	Hvidovre 3
lnpho lnalp lnalb	0.0065	0.0001	0.0001	0.0072
+lnhgb	0.1085	0.0022	0.0016	0.0169
+lncat	0.0001	0.0001	0.0001	0.0001
+lncar	0.0017	0.0005	0.0001	0.0180
+lnpro	0.0002	0.0001	0.0001	0.0001
+lnld	0.0001	0.0001	0.0001	0.0007
+lnage	0.0001	0.0001	0.0011	0.0367
+lnhgb lncar	0.0653	0.0072	0.0008	0.0497
+lncat lnhgb	0.0001	0.0001	0.0001	0.0002
+lncat lncar	0.0001	0.0001	0.0001	0.0001
+lncat lnpro	0.0001	0.0001	0.0001	0.0001
+lncat lnld	0.0001	0.0001	0.0001	0.0001
+lncat lnage	0.0001	0.0001	0.0001	0.0003
+lnhgb lnpro	0.0059	0.0001	0.0001	0.0001
+lnhgb lnld	0.0010	0.0001	0.0001	0.0009
+lnhgb lnage	0.0001	0.0080	0.0034	0.0747
+lnpro lncar	0.0003	0.0001	0.0001	0.0001
+lnpro lnld	0.0001	0.0001	0.0001	0.0001
+lnpro lnage	0.0001	0.0001	0.0001	0.0001
+lncar lnld	0.0001	0.0001	0.0001	0.0022
+lncar lnage	0.0001	0.0001	0.0003	0.0618
+lnld lnage	0.0001	0.0001	0.0001	0.0025

Table 17: Test for homogeneity of covariance matrices between diagnoses within populations.

	Have disease	Healthy
Test positive	a	b
Test negative	c	d

Sensitivity	$\frac{a}{a+c}$
Specificity	$\frac{d}{d+b}$
NPV	$\frac{d}{d+c}$
PPV	$\frac{a}{a+b}$

Table 18: Illustration and definition of the four concepts: Sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV)

	Hvidovre						Herlev			
	Women		Men		Both		Women		Men	
	step	prob	step	prob	step	prob	step	prob	step	prob
lnalb	1	0.0001	1	0.0001	1	0.0001	5	0.0635		
lnpho	2	0.0001			3	0.0001			5	0.0237
lnpro	3	0.0002	2	0.0033	2	0.0001	6	0.0668		
lnhgb	4	0.0015			5	0.0243	2	0.0003		
lnld	5	0.0049			4	0.0011	3	0.0051		
lnage	6	0.0431	4	0.0398			4	0.0119	2	0.0091
lnalp			3	0.0109	6	0.0332	1	0.0001	4	0.0212
lniro									1	0.0007
lnast									3	0.0485
lnura									6	0.0226
lnleu									7	0.0846

Table 19: Entry step and significance level at entry for variables selected by stepwise procedure.

WOMEN	Malignancy	PHPT	Transient
Intercept	-2354	-2412	-2467
alb	268.24	279.11	276.60
pho	-16.87	-19.99	-15.29
pro	397.63	393.75	406.32
hgb	126.43	131.93	136.98
ld	73.19	69.88	71.34
age	171.27	174.94	174.52
MEN	Malignancy	PHPT	Transient
Intercept	-1434	-1477	-1495
alb	223.31	237.45	232.80
pro	227.67	221.31	237.93
alp	0.56	-1.43	-2.31
age	131.23	129.65	125.30

Table 20: The table shows the diagnostic hypercalcemia discriminant functions for each gender in the Hvidovre population. The variables were logarithmically transformed before the values of the discriminant functions were calculated.

	Hvidovre 1 Women		
	Malignancy %	PHPT %	Transient %
Malignancy	76 (36/47)	13 (6/47)	19 (5/47)
PHPT	9 (3/32)	72 (23/32)	10 (6/32)
Miscellaneous	57 (8/14)	28 (4/14)	14 (2/14)
Transient	10 (5/47)	17 (8/47)	72 (34/47)
	Hvidovre 1 Men		
Malignancy	77 (24/31)	16 (5/31)	6 (2/31)
PHPT	0 (0/11)	72 (8/11)	27 (3/11)
Miscellaneous	30 (3/10)	20 (2/10)	50 (5/10)
Transient	19 (3/16)	19 (3/16)	62 (10/16)

Table 21: The table shows the results from discriminant functions in Hvidovre 1.

	Hvidovre 1		Herlev
Different serum calcium (cor) cut-off-levels	2.58 Mean + 3 SD	2.68 Corrected cut-off level	2.68 Mean + 3 SD
Medical diagnoses	Number per 10,000 patients		
Malignancy	78	25	95
PHPT	43	12	32
Miscellaneous	24	3	15
Transient	64	2	8
Total number	257	42	150

Table 22: The table shows distribution of hypercalcemia-associated diagnosis in the Hvidovre population at various cut-off levels for calcium (corr) concentrations. Distribution of the Herlev population is shown in the right column. The numbers are calculated per 10,000 patients. Number of patients for different hypercalcaemia cut off levels in Hvidovre, compared to the cut-off level in Herlev. The level 2.58 is the actual limit in Hvidovre, 2.68 is the Herlev limit.

Women	Malignancy	PHPT	Miscellaneous
Hemoglobin	3.10	9.31	4.58
LD	4.61	2.76	4.37
Albumin	108.42	114.17	114.75
Protein	219.57	214.09	209.91
Age	96.78	98.85	93.11
ALP	24.03	22.67	21.94
Intercept	-1085	-1100	-1059
Men	Malignancy	PHPT	Miscellaneous
ALP	25.48	23.68	19.85
Age	71.13	73.08	57.13
Iron	8.26	11.17	13.10
ASAT	-12.86	-13.38	-8.91
Phosphate	-21.40	-24.09	-4.74
Urea	-0.05	3.74	-6.36
Leucocyte	4.89	4.70	9.30
Intercept	-218.40	-218.11	-175.11

Table 23: The table shows the diagnostic hypercalcemia discriminant functions for each gender in the Herlev population. The variables were logarithmically transformed before the values of the discriminant functions were calculated.

	Training population: Hvidovre 1									
	Women					Men				
	Diagnostic accuracy	Sensitivity	Specificity	PPV PPV	NPV NPV	Diagnostic accuracy	Sensitivity	Specificity	PPV	NPV
Hvidovre 1	83	87	74	87	74	75	77	69	89	48
Hvidovre 2	79	83	40	93	20	64	60	75	86	43
Hvidovre 3	67	69	62	89	31	52	55	43	71	27
Herlev	79	81	33	97	5	70	80	0	85	0
	Training population: Herlev 1									
	Women					Men				
Herlev	82	77	92	94	71	77	87	65	77	78
Hvidovre 1	76	57	86	68	80	62	59	64	55	67
Hvidovre 2	89	81	92	81	92					
Hvidovre 3	91	88	91	70	97					

Table 24: The table shows the diagnostic performance of the Hvidovre 1 discriminant functions discriminating between significant and transient hypercalcemia, and of the Herlev discriminant functions, discriminating between malignancy-associated and non-malignancy-associated hypercalcemia in the different hypercalcemia populations. The diagnostic performance was assessed from the diagnostic accuracy, sensitivity, and specificity, PPV, NPV. The Herlev function for men could not be applied in the Hvidovre 2 and Hvidovre 3 populations as ASAT was not measured in the Hvidovre 2 study and iron was not measured in the Hvidovre 3 study.

	DA	Sensitivity	Specificity	PPV	NPV
Hvidovre 1	78.37	78.62	84.50	89.06	61.25
Hvidovre 2	74.36	74.24	75.00	94.23	34.62
Hvidovre 3	58.21	62.26	42.86	80.49	23.08
Herlev	81.75	84.50	37.50	95.61	13.04

Table 25: Diagnostic performance for function trained on both genders in Hvidovre 1. The function discriminates between clinical significant and non-significant hypercalcaemia. DA (diagnostic accuracy).

	Imputation number										Joint
Population	1	2	3	4	5	6	7	8	9	10	
Hvidovre 1	alb pho LD pro	alb LD pro pho	alb pho ure pro	alb pho pro ure bic alp AST	alb LD pro	alb pho LD pro	alb pho LD pro	alb pho ure pro bic alp	alb LD pro	alb pho ure pro	alb pho LD pro
Hvidovre 2	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD	alb pho LD
Hvidovre 3	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil	alb bil
Hvidovre all	alb pho LD	alb LD pho	alb pho LD	alb pho LD	alb LD pho	alb pho LD	alb pho LD	alb pho LD	alb LD pho	alb pho LD	alb pho LD
Herlev	hgb LD alb alp glo	hgb alp alb LD glo	hgb alp LD alb	alb alp hgb LD	hgb alp LD alb glo	hgb alp LD alb	hgb alp LD alb glo	hgb LD alb alp glo ure	alb alp hgb LD	alb alp hgb LD	hgb alp LD alb
All	alb LD pho hgb alp age	alb LD pho alp hgb	alb LD hgb pho alp	alb LD pho alp hgb	alb LD pho alp hgb	alb LD pho alp hgb age	alb LD pho alp hgb	alb LD pho alp hgb age	alb LD hgb alp pho	alb LD pho alp hgb	alb LD pho alp hgb

Table 26: Discriminant variables in discriminant functions to differentiate between malignancy and PHPT for 10 imputations of variables. The last column is the set chosen for the discriminant functions (joint).

Herlev	All	Malignancy		PHPT	
	DA	Sensitivity	PPV	Sensitivity	PPV
Herlev all patients	84.78	83.26	95.85	89.31	64.33
Herlev complete cases	85.70	85.34	96.00	87.00	62.00
Hvidovre 1 all patients	78.10	75.51	88.84	82.79	65.13
Hvidovre 1 complete cases	75.00	70.21	88.92	83.85	60.42
Hvidovre 1	All	Malignancy		PHPT	
	DA	Sensitivity	PPV	Sensitivity	PPV
Herlev all patients	81.83	78.26	96.84	92.41	58.98
Herlev complete cases	83.02	78.73	98.12	95.56	60.66
Hvidovre 1 all patients	85.70	81.67	95.50	93.02	73.68
Hvidovre 1 complete cases	86.33	80.52	97.90	96.88	73.30
Hvidovre all	All	Malignancy		PHPT	
	DA	Sensitivity	PPV	Sensitivity	PPV
Herlev all patients	78.17	74.30	95.51	89.66	54.07
Herlev complete cases	78.52	73.50	96.71	92.86	55.09
Hvidovre 1 all patients	83.80	81.28	92.69	88.37	72.27
Hvidovre 1 complete cases	83.44	79.19	94.24	91.18	70.62
All	All	Malignancy		PHPT	
	DA	Sensitivity	PPV	Sensitivity	PPV
Herlev all patients	84.61	82.79	96.09	90.00	63.84
Herlev complete cases	84.14	82.50	96.73	90.00	58.97
Hvidovre 1 all patients	82.89	80.13	92.33	87.91	70.95
Hvidovre 1 complete cases	79.86	75.96	92.25	87.50	65.03

Table 27: Performance on complete and completed cases; diagnostic accuracy (DA).

Herlev		All	Malignancy		PHPT	
Test	Training	DA	Sensitivity	PPV	Sensitivity	PPV
Her-lev	Herlev	84.78	83.26	95.85	89.31	64.33
	Herlev priors	89.91	95.81	91.15	72.41	85.44
	Hvidovre 1	81.83	78.26	96.84	92.41	58.98
	Hvidovre 1priors	85.65	85.47	94.84	86.21	66.70
	Hvidovre all	78.17	74.30	95.51	89.66	54.07
	Hvidovre all priors	80.26	77.09	95.67	89.66	56.90
	all patients	84.61	82.79	96.09	90.00	63.84
	all patients priors	89.04	89.19	95.88	88.62	73.47
Hvid-ovre 1	Herlev	78.10	75.51	88.84	82.79	65.13
	herlelv priors	79.34	91.92	79.34	56.51	79.49
	Hvidovre 1	85.70	81.67	95.50	93.02	73.68
	Hvidovre 1 priors	84.17	84.42	90.40	83.72	74.79
	Hvidovre all	83.80	81.28	92.69	88.37	72.27
	Hvidovre all priors	84.63	83.97	91.49	85.81	74.71
	all patients	82.89	80.13	92.33	87.91	70.95
	all patients priors	83.80	87.56	87.35	76.98	77.36
Hvid-ovre 2	Herlev	87.41	81.67	93.25	93.57	82.65
	Herlev priors	85.00	90.33	82.40	79.29	88.46
	Hvidovre 1	92.07	85.00	99.63	99.64	86.13
	Hvidovre 1 priors	91.21	89.33	93.39	93.21	89.09
	Hvidovre all	88.10	77.00	100.0	100.0	80.24
	Hvidovre all priors	91.21	83.00	100.0	100.0	84.60
	all patients	91.38	83.33	100.0	100.0	84.85
	all patients priors	89.48	86.67	92.54	92.50	86.62
Hvid-ovre 3	Herlev	87.78	75.56	92.56	95.93	85.50
	Herlev priors	78.89	88.89	68.26	72.22	90.67
	Hvidovre 1	87.56	80.56	87.35	92.22	87.70
	Hvidovre 1 priors	81.78	83.33	74.29	80.74	87.90
	Hvidovre all	89.33	78.89	93.42	96.30	87.26
	Hvidovre all priors	86.67	83.33	83.33	88.89	88.89
	all patients	84.44	72.22	86.65	92.59	83.35
	all patients priors	83.56	87.78	75.25	80.74	90.85
Hvid-ovre all	Herlev	82.46	76.98	90.41	89.49	75.17
	Herlev priors	80.71	91.11	78.23	67.35	85.50
	Hvidovre 1	88.08	82.94	95.26	94.69	81.20
	Hvidovre 1 priors	85.76	85.71	88.61	85.82	82.37
	Hvidovre all	85.62	79.21	94.33	93.88	77.84
	Hvidovre all priors	86.32	83.13	91.77	90.41	80.66
	all patients	85.40	79.76	93.32	92.65	78.08
	all patients priors	85.22	87.38	86.49	82.45	83.57
All pa-tients	Herlev	83.78	80.14	92.95	89.84	73.07
	Herlev priors	83.89	93.11	83.16	68.50	85.64
	Hvidovre 1	85.96	81.04	95.87	94.17	74.86
	Hvidovre 1 priors	85.72	85.61	91.03	85.91	78.16
	Hvidovre all	83.36	77.64	94.82	92.91	71.35
	Hvidovre all priors	84.54	80.99	93.42	90.47	74.04
	all patients	84.57	80.47	93.99	91.42	73.72
	all patients priors	86.31	88.00	89.91	83.50	80.66

Table 28: Comparison of functions with equal and proportional priors

Herlev		All	Malignancy		PHPT	
Test	Training	DA	Sensitivity	PPV	Sensitivity	PPV
herl	thvid	78.17	74.30	95.51	89.66	54.07
	thviQ	82.09	79.77	95.54	88.97	59.73
	total	84.61	82.79	96.09	90.00	63.84
	totaQ	85.13	83.60	95.99	89.66	64.85
hvid1	thvid	83.80	81.28	92.69	88.37	72.27
	thviQ	81.16	77.18	92.33	88.37	68.11
	total	82.89	80.13	92.33	87.91	70.95
	totaQ	81.74	77.95	92.55	88.60	68.90
hvid2	thvid	88.10	77.00	100.0	100.0	80.24
	thviQ	86.21	76.67	95.83	96.43	79.41
	total	91.38	83.33	100.0	100.0	84.85
	totaQ	88.97	82.33	95.75	96.07	83.56
hvid3	thvid	89.33	78.89	93.42	96.30	87.26
	thviQ	86.00	77.78	85.96	91.48	86.06
	total	84.44	72.22	86.65	92.59	83.35
	totaQ	87.11	78.89	87.65	92.59	86.82
thvid	thvid	85.62	79.21	94.33	93.88	77.84
	thviQ	83.13	77.14	91.53	90.82	75.55
	total	85.40	79.76	93.32	92.65	78.08
	totaQ	84.69	79.13	92.58	91.84	77.39
total	thvid	83.36	77.64	94.82	92.91	71.35
	thviQ	82.98	78.21	93.52	90.94	71.43
	total	84.57	80.47	93.99	91.42	73.72
	totaQ	83.89	80.57	92.73	89.45	73.39

Table 29: Comparison of performance with and without quadratic.

Herlev		All	Malignancy		PHPT	
Test	Training	DA	Sensitivity	PPV	Sensitivity	PPV
herl	thvid	78.17	74.30	95.51	89.66	54.07
	thviQ	82.09	79.77	95.54	88.97	59.73
	thvip	80.26	77.09	95.67	89.66	56.90
	thvpQ	83.04	81.74	94.87	86.90	61.62
	total	84.61	82.79	96.09	90.00	63.84
	totaQ	85.13	83.60	95.99	89.66	64.85
	totap	89.04	89.19	95.88	88.62	73.47
	totpQ	86.78	86.28	95.62	88.28	68.49
hvid1	thvid	83.80	81.28	92.69	88.37	72.27
	thviQ	81.16	77.18	92.33	88.37	68.11
	thvip	84.63	83.97	91.49	85.81	74.71
	thvpQ	84.63	82.56	92.79	88.37	73.66
	total	82.89	80.13	92.33	87.91	70.95
	totaQ	81.74	77.95	92.55	88.60	68.90
	totap	83.80	87.56	87.35	76.98	77.36
	totpQ	83.06	83.08	89.89	83.02	73.05
hvid2	thvid	88.10	77.00	100.0	100.0	80.24
	thviQ	86.21	76.67	95.83	96.43	79.41
	thvip	91.21	83.00	100.0	100.0	84.60
	thvpQ	89.66	83.33	96.15	96.43	84.38
	total	91.38	83.33	100.0	100.0	84.85
	totaQ	88.97	82.33	95.75	96.07	83.56
	totap	89.48	86.67	92.54	92.50	86.62
	totpQ	91.38	87.00	95.98	96.07	87.37
hvid3	thvid	89.33	78.89	93.42	96.30	87.26
	thviQ	86.00	77.78	85.96	91.48	86.06
	thvip	86.67	83.33	83.33	88.89	88.89
	thvpQ	84.44	77.78	82.35	88.89	85.71
	total	84.44	72.22	86.65	92.59	83.35
	totaQ	87.11	78.89	87.65	92.59	86.82
	totap	83.56	87.78	75.25	80.74	90.85
	totpQ	85.11	80.00	82.32	88.52	86.92
thvid	thvid	85.62	79.21	94.33	93.88	77.84
	thviQ	83.13	77.14	91.53	90.82	75.55
	thvip	86.32	83.13	91.77	90.41	80.66
	thvpQ	85.85	81.98	91.99	90.82	79.68
	total	85.40	79.76	93.32	92.65	78.08
	totaQ	84.69	79.13	92.58	91.84	77.39
	totap	85.22	87.38	86.49	82.45	83.57
	totpQ	85.63	83.57	90.16	88.27	80.71
total	thvid	83.36	77.64	94.82	92.91	71.35
	thviQ	82.98	78.21	93.52	90.94	71.43
	thvip	84.54	80.99	93.42	90.47	74.04
	thvpQ	84.93	81.93	93.14	89.92	74.89
	total	84.57	80.47	93.99	91.42	73.72
	totaQ	83.89	80.57	92.73	89.45	73.39
	totap	86.31	88.00	89.91	83.50	80.66
	totpQ	85.29	84.55	91.30	86.54	77.05

Table 30: Comparison of performance with and without quadratic/priors.

Test	Train	DA		Malignancy				PHPT			
				Sensitivity		PPV		Sensitivity		PPV	
		Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
herl	thvid	76.52	79.13	72.09	75.58	95.38	95.59	89.66	89.66	52.00	55.32
	thvip	80.00	80.87	76.74	77.91	95.65	95.71	89.66	89.66	56.52	57.78
	thviQ	80.87	83.48	79.07	81.40	94.44	95.89	86.21	89.66	58.14	61.90
	thvpQ	82.61	84.35	81.40	82.56	94.59	95.95	86.21	89.66	60.98	63.41
	total	82.61	85.22	80.23	83.72	95.83	97.26	89.66	93.10	60.47	65.00
	totap	87.83	89.57	87.21	90.70	95.06	96.25	86.21	89.66	70.27	75.76
	totaQ	84.35	86.09	82.56	84.88	95.95	96.05	89.66	89.66	63.41	66.67
	totpQ	85.22	87.83	84.88	87.21	94.81	96.15	86.21	89.66	65.79	70.27
hvid1	thvid	81.82	85.12	78.21	83.33	92.42	92.86	88.37	88.37	69.09	74.51
	thvip	83.47	86.78	82.05	85.90	90.28	93.06	83.72	88.37	72.55	77.55
	thviQ	80.17	81.82	75.64	78.21	92.19	92.42	88.37	88.37	66.67	69.09
	thvpQ	83.47	85.12	80.77	83.33	92.65	92.86	88.37	88.37	71.70	74.51
	total	81.82	83.47	78.21	82.05	91.43	92.65	86.05	88.37	69.09	72.55
	totap	81.82	86.78	84.62	89.74	85.90	89.74	74.42	81.40	73.33	81.40
	totaQ	80.17	82.64	76.92	79.49	90.91	93.85	86.05	90.70	67.27	70.37
	totpQ	80.99	85.12	79.49	85.90	87.84	91.55	79.07	86.05	69.23	76.60
hvid2	thvid	87.93	89.66	76.67	80.00	100.0	100.0	100.0	100.0	80.00	82.35
	thvip	89.66	91.38	80.00	83.33	100.0	100.0	100.0	100.0	82.35	84.85
	thviQ	86.21	86.21	76.67	76.67	95.83	95.83	96.43	96.43	79.41	79.41
	thvpQ	89.66	89.66	83.33	83.33	96.15	96.15	96.43	96.43	84.38	84.38
	total	91.38	91.38	83.33	83.33	100.0	100.0	100.0	100.0	84.85	84.85
	totap	87.93	89.66	86.67	86.67	89.66	92.86	89.29	92.86	86.21	86.67
	totaQ	87.93	89.66	80.00	83.33	92.59	96.15	92.86	96.43	81.82	84.38
	totpQ	89.66	93.10	83.33	90.00	93.10	96.43	92.86	96.43	84.38	90.00
hvid3	thvid	88.89	91.11	77.78	83.33	93.33	93.75	96.30	96.30	86.67	89.66
	thvip	86.67	86.67	83.33	83.33	83.33	83.33	88.89	88.89	88.89	88.89
	thviQ	84.44	86.67	77.78	77.78	82.35	87.50	88.89	92.59	85.71	86.21
	thvpQ	84.44	84.44	77.78	77.78	82.35	82.35	88.89	88.89	85.71	85.71
	total	82.22	86.67	66.67	77.78	85.71	87.50	92.59	92.59	80.65	86.21
	totap	80.00	84.44	83.33	88.89	71.43	76.19	77.78	81.48	87.50	91.67
	totaQ	86.67	88.89	77.78	83.33	87.50	88.24	92.59	92.59	86.21	89.29
	totpQ	82.22	88.89	77.78	83.33	77.78	88.24	85.19	92.59	85.19	89.29
thvid	thvid	84.82	86.61	77.78	80.95	94.23	94.44	93.88	93.88	76.67	79.31
	thvip	85.71	87.28	81.75	84.92	91.15	93.04	89.80	91.84	79.28	82.57
	thviQ	82.59	83.48	76.19	77.78	91.43	91.59	90.82	90.82	74.79	76.07
	thvpQ	85.27	86.16	80.95	82.54	91.89	92.04	90.82	90.82	78.76	80.18
	total	84.38	85.71	77.78	80.95	92.73	93.52	91.84	92.86	76.47	78.95
	totap	83.48	87.05	84.92	88.89	85.04	88.19	80.61	84.69	80.81	85.57
	totaQ	83.93	85.27	77.78	80.16	91.67	93.46	90.82	92.86	76.47	78.26
	totpQ	83.48	87.05	80.16	84.92	89.08	91.45	86.73	89.80	77.48	82.24
total	thvid	82.89	84.07	76.89	78.77	94.77	94.89	92.91	92.91	70.66	72.39
	thvip	83.78	85.25	80.19	81.60	92.90	94.02	89.76	91.34	73.08	74.84
	thviQ	82.60	83.78	77.36	79.25	92.74	93.85	89.76	91.34	70.73	72.50
	thvpQ	84.37	85.25	81.13	82.55	92.97	93.55	89.76	90.55	74.03	75.50
	total	83.48	85.25	79.25	81.60	93.33	94.51	90.55	92.13	72.33	74.68
	totap	85.25	87.32	86.32	89.15	89.00	91.30	81.89	85.83	78.52	82.58
	totaQ	83.19	84.66	79.72	81.13	91.89	93.48	88.19	90.55	72.44	74.19
	totpQ	84.22	85.99	83.02	85.85	89.80	93.26	84.25	89.76	74.83	78.87

Table 31: Maximum og minimum.

Hvidovre 3		Malignancy		PHPT	
Discriminant functions	Diagnostic % accuracy %	Sensitivity %	PPV%	Sensitivity%	PPV%
PTH included	90 (26/29)	83 (10/12)	91 (10/11)	94 (16/17)	89 (16/18)
Routine variables	89 (40/45)	83 (15/18)	88 (15/17)	92 (25/27)	89 (25/28)

Table 32: The table shows the diagnostic accuracy, sensitivity, and PPV for malignancy and PHPT in a subpopulation of patients in Hvidovre 3 that had serum intact PTH concentration measured.