Research Re-Exam Semester 1 – 2018-2019

May 1, 2019

During the exam, you have access on a computer to these books:

- Baynes & Dominiczak: Medical Biochemistry
- Campbell: Statistics at square one
- Donders: Literature Measurement errors
- Fletcher: Clinical Epidemiology
- van Oosterom en Oostendorp: Medische Fysica
- Petrie and Sabin: Medical Statistics at a Glance
- Turnpenny: Emery's Elements of Medical Genetics
- Form with statistical formula's

You are allowed to use a calculator of the type Casio FX-82MS.

The questions must be answered in English. If you cannot remember a specific English term, you may use the Dutch term.

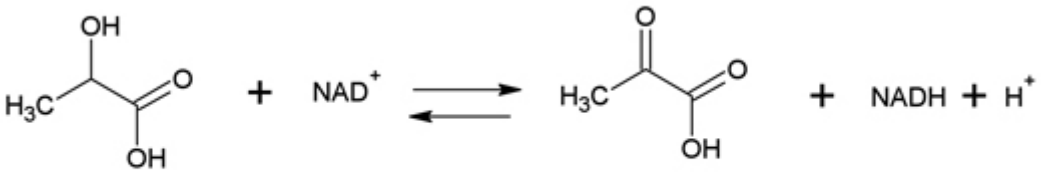
Write your name and student number on the first page of each question

Name:

Student Number:

Question 1

Wet lab research: Dr. G. Bosman
(10 points)

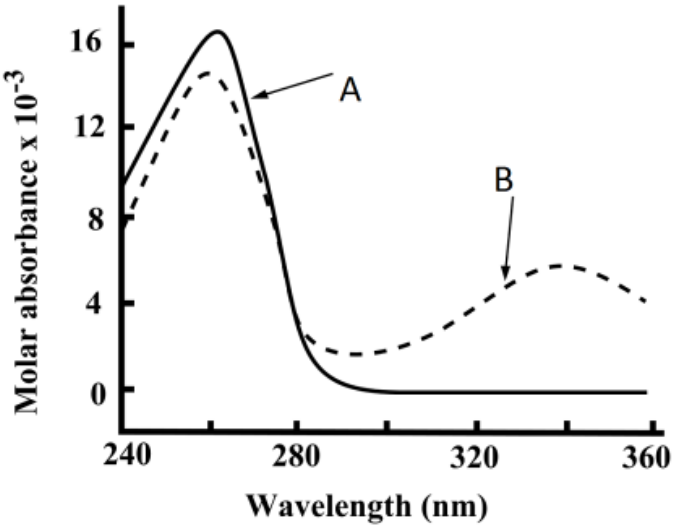


Lactic Acid Pyruvic Acid

The enzyme lactate dehydrogenase (LDH) catalyzes the conversion of lactic acid into pyruvic acid and vice versa.

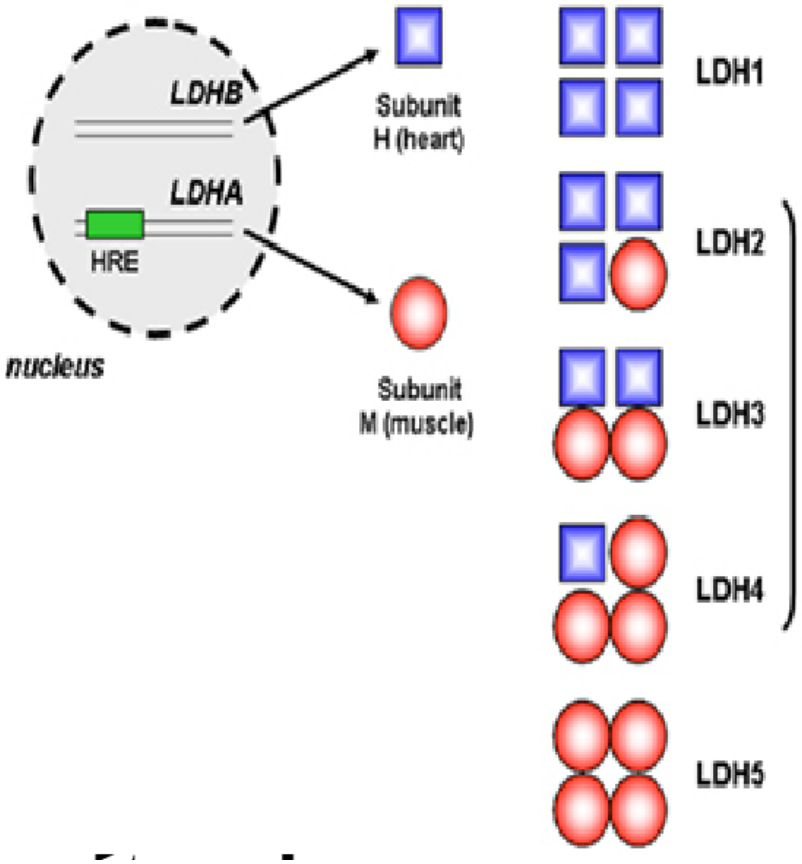
a. The ΔG of this reaction is +25.1 kJ/mol. Which component will, in equilibrium conditions, be present in the highest concentration? Explain your answer. (2 pt)

Lactic acid; thermodynamic conditions favor the formation of lactic acid as this reaction yields netto energy

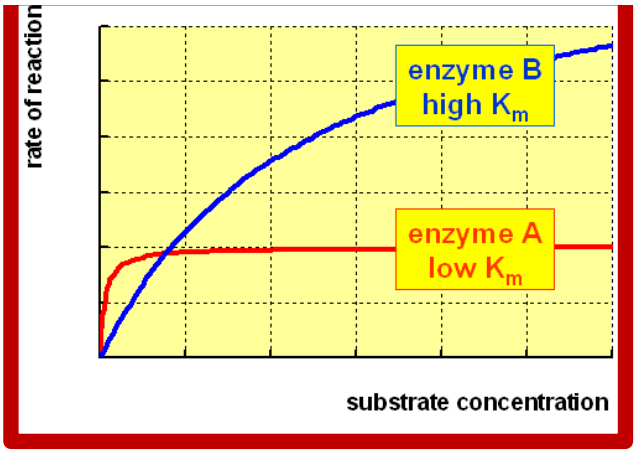


b. The reaction as shown below Question 1a can be used to measure the characteristics of LDH.This graph shows the absorbance spectra of two components involved in such a measurement. Which component is indicated by the letter B? Explain your answer. (2 pts)

NADH; the only reactant that has an absorption at 340 nm (this is commonly used for other oxidoreductases, such as ADH at the enzyme practical)



c. LDH is a multimeric enzyme. As depicted above, multiple isoforms of LDH occur, depending on the combination of the two subunits. LDH5 has a higher Km and a higher Vmax for pyruvate reduction than LDH1. Sketch the Michaelis-Menten graphs of these two isoforms. (6 pt)



Enzyme B = LDH5 (and enzyme A is LDH1) (4 pt); substrate = pyruvate (1 pt); rate of reaction = decrease in absorption at 340 nm per unit time, or 'arbitrary units' (1 pt)

Name:

Student Number:

Question 2
Q1 Modelling – dr. T. Oostendorp
(15 points)

Chlorothiazide is a drug that is used in the treatment of hypertension. The pharmacokinetics of chlorothiazide after intravenous injection is well described by a single compartment model. For a single compartment model with linear kinetics the differential equation that describes the relation between the concentration $c(t)$ and the dose velocity $D(t)$ is

$$V\frac{d}{dt}c(t) = D(t) - k\,c(t)$$

where V is the volume of the compartment.

Figure 2.1 shows the measured plasma concentration of chlorothiazide after intravenous injection of 500 mg chlorothiazide.

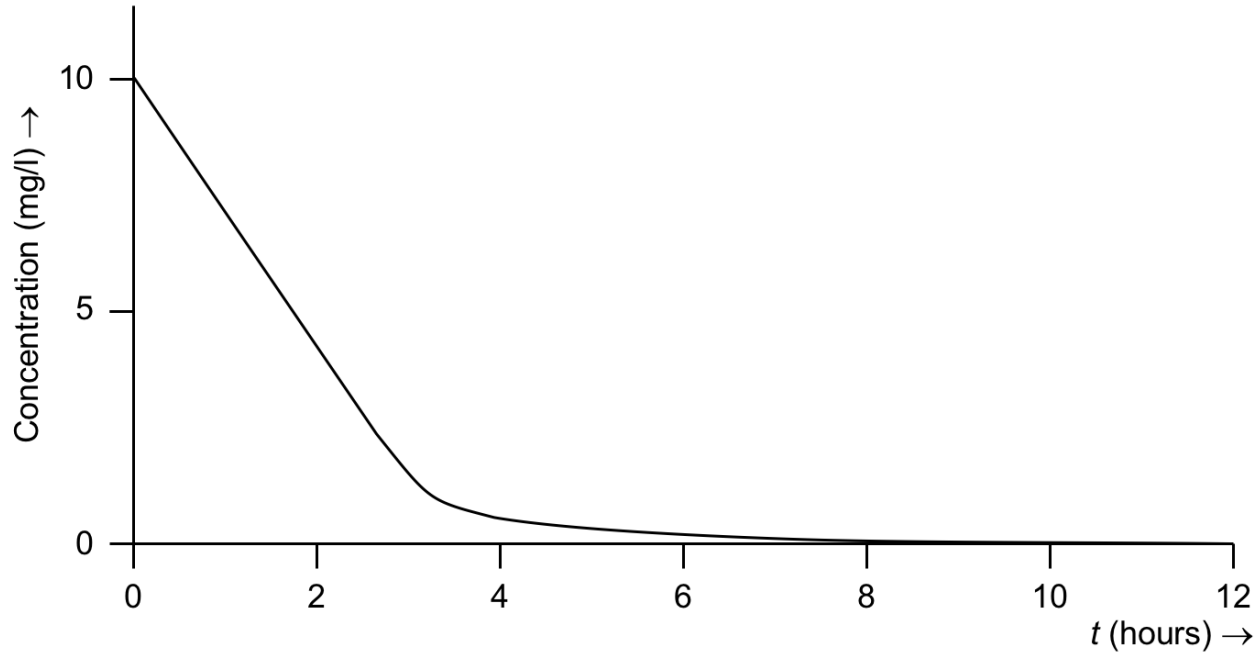


Figure 2.1 Plasma concentration after intravenous injection of 500 mg chlorothiazide.

a. Explain how you can deduct from the plot in figure 2.1 that the elimination of chlorothiazide does not follow linear kinetics. Refer in you answer to the differential equation in the introduction of this question. (4 pts)

In a linear model the elimination is given by $kc(t)$ (see differential equation). This means that, if the dose velocity is zero (which is the case after the injection), the velocity by which c(t) must decrease as c(t) itself decreases. However, figure 2.1 shows that for the first to hours the concentration decreases with almost constant velocity for the first two hours.

b. Use figure 2.1 to calculate the value of V in the single compartment model for chlorothiazide. (2 pt)

The 500 mg dose resulted in a concentration of 10 mg. With concentration = amount / volume we find $V = 50\ \ell$.

The elimination of chlorothiazide is better described by Michaelis Menten (MM) kinetics. in MM-kinetics, the elimination velocity $v(t)$ is given by

$$v(t) = \frac{V_{\max}}{K_{\mathrm{m}} + c(t)}\ c(t)$$

c. Give the differential equation for a single compartment model for chlorothiazide with MM-kinetics elimination. Explain how you arrive at that equation. (4 pt)

Change = in - out

in = dose velocity

out = elimination

change = change in amount or drug = change in (volume times concentration)

$$V\frac{d}{dt}c(t) = D(t) - \frac{V_{\max}}{K_{\mathrm{m}} + c(t)}\ c(t)$$

For patients with a healthy liver, the value of $V_{\max}$ for chlorothiazide is 150 mg/hr. For patients with reduced liver functionality, the value is 100 mg/hr.

d. Explain why a dose of 1000 mg every 8 hours, will result in equilibrium in patients with a healthy liver, whereas in patients with reduced liver functionality no equilibrium is reached. (5 pt)

In patients with a healthy liver the dose every 8 hours does not exceed the maximum elimination in 8 hours ($8 \cdot 150 = 1200$ mg). Hence, the elimination/hour can reach a value equal to the dose velocity. In patients with reduced liver functionality, the maximum elimination in 8 hours is $8 \cdot 100 = 800$ mg. This is less that the dose in 8 hours, so in these patients the concentration will continue to increase.

Name:

Student Number:

Question 3
Signals: Electromyography – dr. T. Oostendorp
(10 points)

Figure 3.1 shows the electric activity of a muscle recorded at the body surface (the EMG).

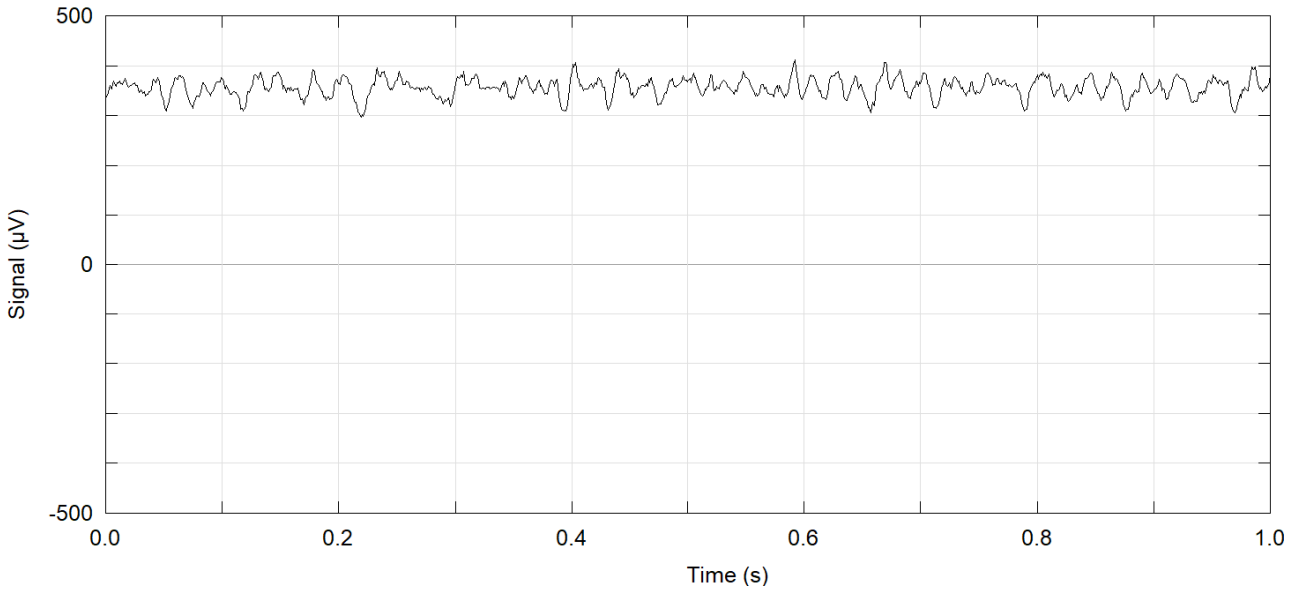


Figure 3.1 Recorded EMG signal.

a. Explain why 20 Hz is not a proper sample rate for this EMG-signal. (3 pt)

A sample rate of 20 Hz will result in 1 sample every 50 ms. Figure 3.1 demonstrates clearly that there is a lot of variation in the signal within 50 ms, so 20 Hz is way too low.

Figure 3.2 shows the spectra of three different signals.

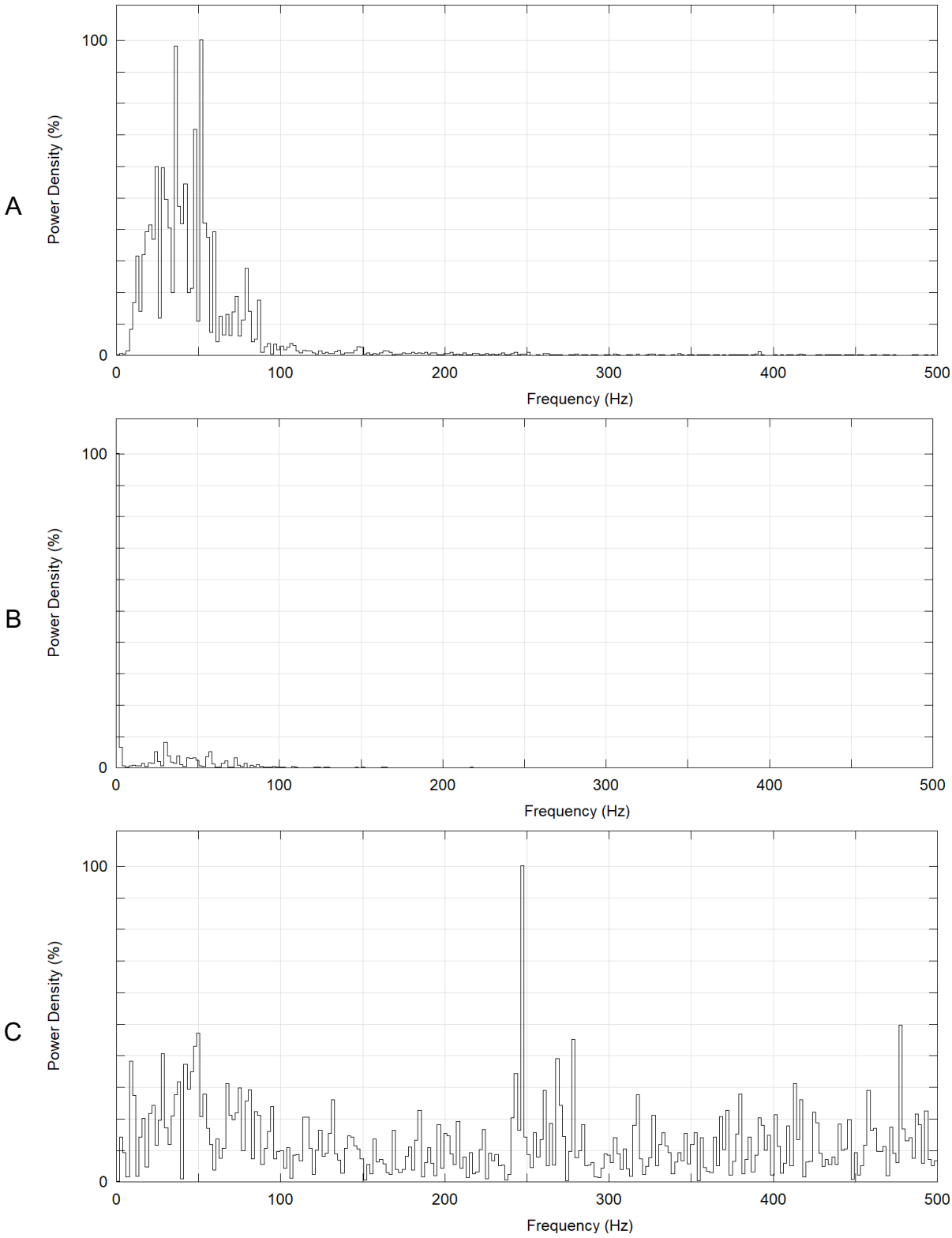


Figure 3.2 Three different spectra.

b. Which of the 3 spectra in figure 3.2 corresponds to the signal of figure 3.1? Explain why the other 2 cannot correspond to that signal. (3 pt)

It must be spectrum B. It cannot be spectrum A, because the signal clearly has a component of 0 Hz (the average is above zero). It can also not be spectrum C, as the signal does not have clear high frequency components.

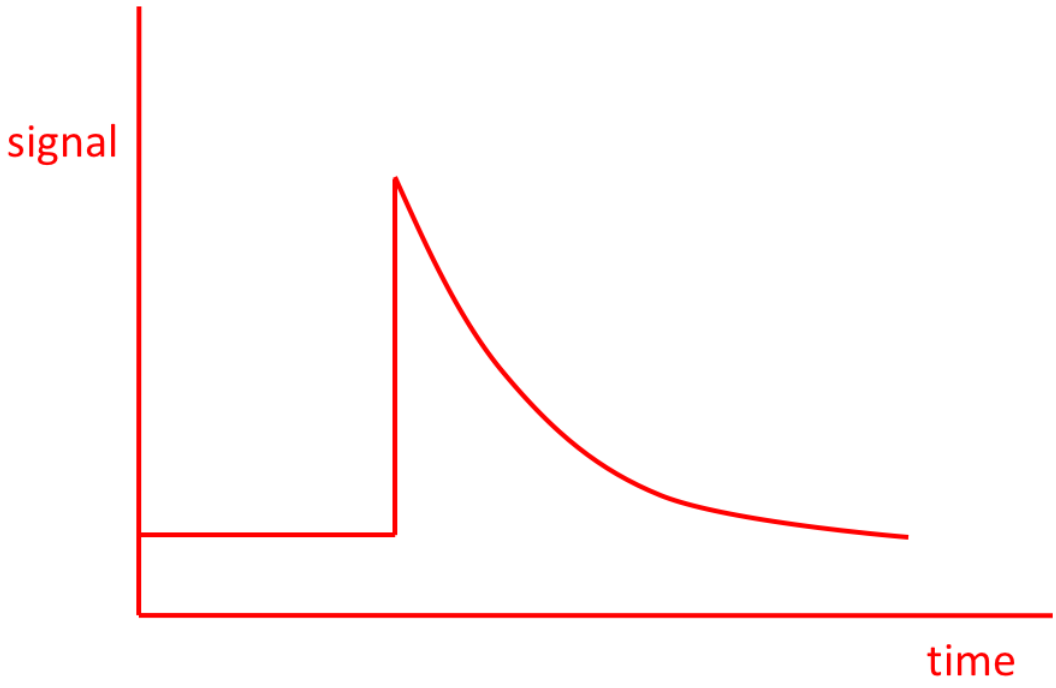
The researcher who recorded the signal shown in figure 3.1 decides to apply a high-pass filter to the signal.

c. What is indicated along the x- and y-axis of the step response? (2 pt)

x-axis: time

y-axis: amplitude (or “signal strength” or something similar)

d. Sketch the step response of a high pass filter (2 pts)



Name:

Student Number: | | | | | | | |

Question 4
Methods, measures and study designs in epidemiological studies – dr. F. de Vegt (20 points)

Use '*Reeves et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study'- abstract and table*' for question 4 and 5.

a. In the study of Reeves et al., information concerning physical activity was obtained by using a questionnaire. Give two advantages and two disadvantages for this method for the assessment of physical activity. (4 pnt)

+: easy to obtain in large groups of people, takes not much effort for participants, not invasive/painful

-: questionnaires not very accurate, people may give desirable answers, not easy to fill out questions concerning activities

b. Which are the determinant(s), the outcome(s) and the study population in this study? (3 pnt)

Determinant: body mass index (the other variables are confounders)
Outcomes: all cancers, and 17 specific types of cancer
Study population: 1.2 million UK women recruited into the Million Women Study, aged 50-64 during 1996-2001

c. The study design was a prospective cohort study. Explain why a case-control design is not appropriate to study this research question. (3 pnt)

In a case-control designs you start with the selection of cases (people with the outcome of interest) and controls (people without the outcome of interest) and you collect information concerning exposures in the past. In this research question you are interested in multiple outcomes (many types of cancer) and also many exposures (main determinant and many confounders). Then a prospective cohort study is the design of choice.

d. See Table 1, third row. Give an interpretation for the result 27.0 (25.9-28.2) (3 pnt).

In the women with a BMI of 25-29 kg/m2 the median (middle value) of the BMI was 27 kg/m2. The interquartile range is from 25.9 – 28.4, meaning that the middle 50% of observations were between these values.

e. See Table 1, the results concerning follow-up. Calculate the cumulative incidence (CI) of all cancers in the study population (2 pnt).

45 037 incident cancers occurred over the follow-up period in the study population of 1222630 women. So the CI is 3.7%.

f. See Table 1, the results concerning follow-up. Calculate the incidence density (ID) of all cancers in the study population (2 pnt)

45037 incident cancers occurred over the follow-up period of 6 419 000 person years. So the ID is 7.0 per 1000 person years.

g. In which situations is it better to calculate an ID instead of an CI? Explain (3 pnt)

In dynamic populations, because participants are not in the study for an equal time period, so you sum the follow-up time every participants add.

In populations with a high number of lost to follow-up.

Name:

Student Number: | | | | | | | |

Question 5
Statistics – dr. R. Donders
(15 points)

Use '*Reeves et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study'- abstract and table*' for question 4 and 5.

In the paper of Reeves et al the results of a very large cohort study are presented. In total 1,222,630 women participated in this study. In Table 1 of the paper data about the alcohol intake of the women are presented.

a. Is alcohol intake approximately normally distributed? Give a proper motivation based on the data in Table 1 of the paper.(3 pts)

Alcohol intake cannot be smaller than 0. The average intake is 5.0 g/day and the standard deviation is 6.1 g/day. If the distribution would be approximately normal, approximately 42.5% would be between the mean and the mean + 2×SD and 42.5% would be between the mean and the mean – 2×SD. However, that would equal -7.2 and that is an impossible value.

b. Assume that the population standard deviation is equal to the sample estimate: 6.1 g/day. What is the standard error of the sample mean? (3 pts)

The formula for the standard error of a sample mean is , with σ equal to the population standard deviation (6.1 g/day) and n equal to the sample size (1,222,680). So the standard error equals: $6.1/\sqrt{1,222,680} \approx 0.0055$ g/day.

c. The alcohol intake needs to be derived from information provided by the participants. Often they are asked to fill in a diary in which they record which alcoholic beverages they consumed and how many. As with any measurement, this measurement will contain random measurement error. Suppose that the measurement of alcohol intake would have been done without any measurement error, the standard deviation of this variable would become smaller. Explain why this is? (3 pts)

The random measurement error is extra variation that is added to the already existing variation between the true values of the alcohol intake. The present standard deviation (6.1 g/day) is thus both caused by differences in the true scores and random measurement error. If the random measurement error would be removed, the standard deviation would decrease. Note that the presence of a (fixed) systematic measurement error has no influence on the standard deviation.

d. The standard errors for all variables are relatively small due to the very large sample size. So even with random measurement error, a large sample suffices to make very accurate estimates. Can the problem of a systematic measurement error be overcome by a large sample size? Provide a proper motivation. (3 pts)

Suppose that every woman underestimates her own alcohol intake by 1 g/day. In that case the sample average of this huge sample would underestimate the population mean by 1 g/day. Large sample sizes are no fix for systematic measurement errors.

e. Suppose that all women in this sample have been observed for at least one year and every woman would be cancer free at the start of the study. Then these data can be used to estimate the probability of obtaining a certain cancer within the first year of observation. One could also use these data to estimate the probability of dying of this cancer within the first year. For which of these probabilities (proportions) will the standard error be the smaller? Provide a proper motivation. (3 pts)

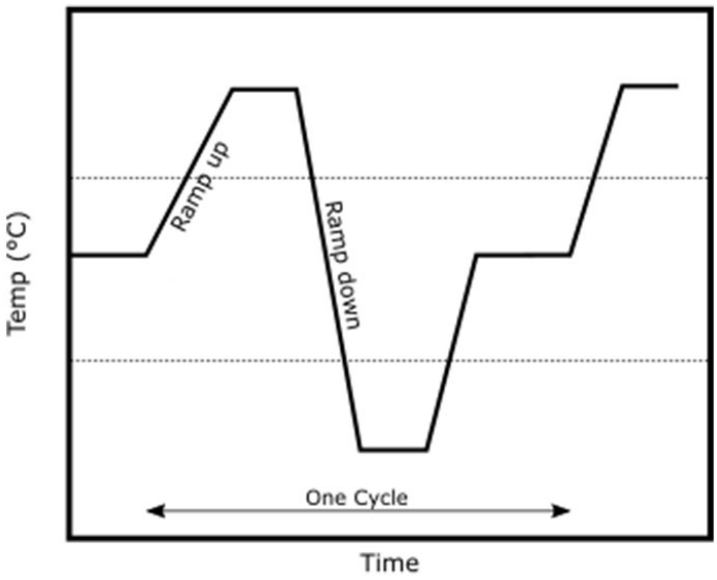
The standard error of the sample proportion is largest for a population proportion of 0.5. It seems reasonable that there is no cancer for which any proportion will be larger than 0.5 (as can be confirmed by these study results). Of course the probability of dying will be smaller than the probability of obtaining that cancer. Thus, the probability of obtaining the cancer will be closer to 0.5 compared to the probability of dying from that cancer. And the standard error for the sample proportion of women obtaining that cancer will be larger than the standard error for the sample proportion of women dying from that cancer.

Name:

Student Number:

Question 6
Genetic lab practice research: dr. D. de Bruijn
(10 points)

To the right, you can see a typical thermal profile of a PCR reaction. The double headed arrow indicates the duration of one cycle. Each cycle consists of three distinct steps that are executed at different temperatures. This figure will be used for questions a-b.



a. What are the temperatures that fit best with each of these steps? Write them at (or near) each step in the figure. (2 pts)

Step 1: Above 90° C; Step 2: Anywhere between 50° and 60° C; Step 3: Around 72° C.

b. Each step of the PCR reaction serves an important purpose for the amplification of DNA. Give the accurate name for each of these steps and provide a short explanation of what happens during each of these PCR steps. (3 pts)

Step 1: denaturation, strand separation; Step 2: annealing, primers bind to each strand; Step 3: elongation, copying of each strand.

Below, you see a snapshot from the 1000Genomes browser with genotype information of SNP rs297731. This table contains genotype frequencies and numbers of persons per genotype (between brackets) for four different American populations (i.e. CLM=Colombian in Medellin, MXL=Mexicans in L.A., PEL=Peruvians in Lima, PUR=Puerto Ricans).

1000GENOMES phase_3: CLM	A|A: 0,362 (34)	A|G: 0,521 (49)	G|G: 0,117 (11)
1000GENOMES phase_3: MXL	A|A: 0,484 (31)	A|G: 0,406 (26)	G|G: 0,109 (7)
1000GENOMES phase_3: PEL	A|A: 0,459 (39)	A|G: 0,471 (40)	G|G: 0,071 (6)
1000GENOMES phase_3: PUR	A|A: 0,313 (30)	A|G: 0,635 (61)	G|G: 0,052 (5)

c Use the data in the above table to calculate the *allele frequency* of the A and G alleles in the PUR population (box 1) and calculate the expected *genotype frequencies* (AA, AG and GG) for this population under assumption of Hardy-Weinberg Equilibrium (box 2). Write down your calculations and round your results to 3 decimals. (4 pt)

Box1 (3 decimals)	A	G
Allele frequency PUR	*0.630*	*0.370*

***Method 1**: count the number of alleles:*
A: 2x30 + 61 = 121; G: 2x5 + 61 = 71; Total: 121 + 71 = 192 (= 2x96).
Allele frequencies: A allele: 121 / 192 = 0.630; G allele: 71 / 192 = 0.370
***Method 2**: Add up frequencies and divide hz by 2*
A: 0.313 + (0.635 / 2) = 0.630; G: 0.052 + (0.635 / 2) = 0.370

Errors:
1: Wrong method: square root of AA and/or GG frequencies (0 pt)
2: Wrong method = (freq AA)^2 + freq AG (0 pt)
3: No calculation shown, but right answer (1 pt)
4: No calculation shown, wrong answer (0 pt)
5: Possibly right method 2, but not divided by 2 (1 pt)
6: Number of total alleles = number of persons (1 pt)
7: Minor calculation or rounding error (- 0.5 pt)
8: Any other calculation error (- 1 pt)

Box 2 (3 decimals)	AA	AG	GG
Genotype frequency PUR	*0.397*	*0.466*	*0.137*

***Method**: use HW equation to calculate genotype frequencies:*
AA: (0.630)^2 = 0.397
AG: 2 x 0.630 x 0.370 = 0.466
GG: (0.370)^2 = 0.137

Errors:
1: Right method with wrong numbers from box 1 (2 pt)
2: Calculation error (- 1 pt)
3: Partial reasoning (HWE equation) shown, answer consistent with box 1 (- 0.5 pt)
4: No calculation shown, but right answer (1 pt)
5: No calculation shown, wrong answer (0 pt)
6: Small error due to rounding (- 0.5 pt)
No penalty if errors from box 1 lead to total frequency unequal to 1 in this answer

d. Use the genotype frequencies from the previous question (c) to calculate the expected number of people, round to whole numbers (use Box 3) and compare this with the actual numbers of people for each genotype in the PUR population as given in the table of the 1000Genomes browser. What is the best evaluation of the differences between these genotype distributions? (1 pt)

Box 3 (whole numbers)	AA	AG	GG
Actual number of people	*30*	*61*	*5*
Expected number of people	*38*	*45*	*13*

Major differences between the actual and calculated genotype distributions, so not consistent with HWE (1 pt). No mentioning of any differences or no mentioning of HWE (-0.5 pt). Alternative answers: max. 0.5 pt for an adequate evaluation of absolutely false results due to errors in previous questions.

End of the exam!

Did you write your name and student number on the first page of each question?