Research Exam Semester 1 – 2017-2018

January 19, 2018

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

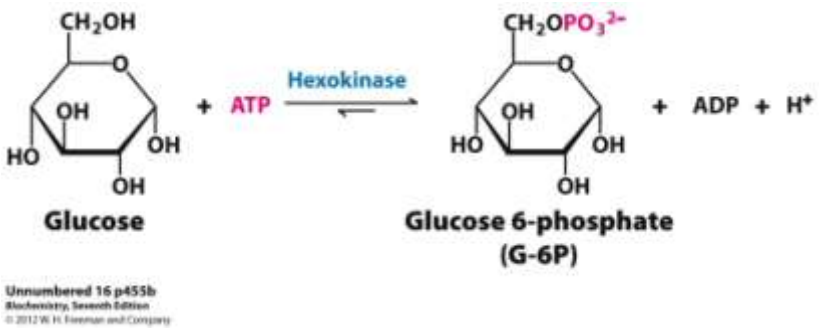
- Baynes & Dominiczak: Medical Biochemistry
- Campbell: Statistics at square one
- Donders: Literature Measurement errors
- Fletcher: Clinical Epidemiology
- van Oosterom en Oostendorp: Medische Fysica
- Petrie and Sabin: Medical Statistics at a Glance
- Turnpenny: Emery's Elements of Medical Genetics

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.

Name: **Studentnummer:** | | | | | | | |

Question 1
Wet lab research: Enzymes in a biomedical context - Dr. G. Bosman
(10 points)



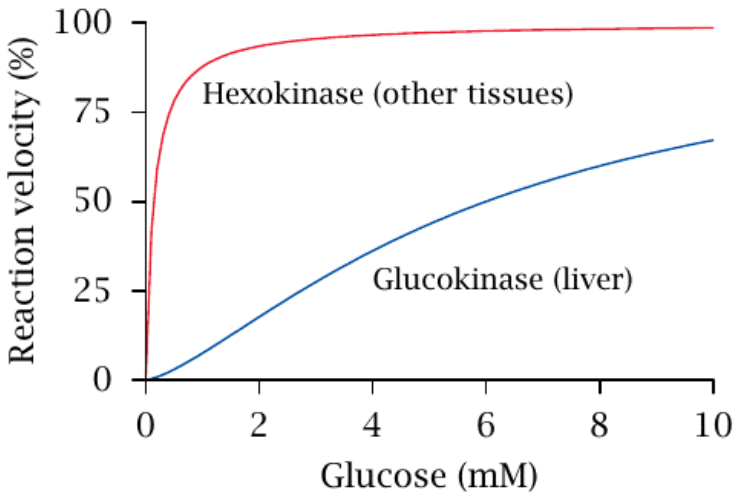
The figure above shows the phosphorylation of glucose as catalyzed by the enzymes glucokinase and hexokinase.

a. To which class of enzymes do these enzymes belong? Explain your answer using the information in the figure.? (2 pts)

Class 2: Transferases: ADP-Pi + glucose → ADP + glucose-P = A-B + C → A + B-C

b. The delta $G^{0'}$ of this reaction is -16.7 kJ/mol. What is the effect of this value on the ratio of the concentrations of glucose 6-phosphate and glucose in equilibrium conditions? Give a short explanation for your answer. (2 pts)

[G-6P] > [G]; equilibrium to the right because the formation of G-6P is energetically favorable (additional clue: hydrolysis of ATP in the formation of glucose-6-P)



c. The graph above shows the Michaelis-Menten kinetics of hexokinase and glucokinase. 100% reaction velocity = 0.30 micromoles Glc-6-P formed $min^{-1}mg^{-1}$.

What are the (approximate) values of the Km en Vmax of these enzymes? (2 pts)

Hexokinase Km = 0.2 mM, Vmax = 0.3 umoles/min/mg; Glucokinase Km = 5 mM, Vmax = 0,2 umoles/min/mg

d. If you need to use one of these enzymes to measure the concentration of glucose in the blood, which enzyme would you choose? Explain your answer (4 pts)

Glucokinase, because the higher Km provides a much larger range of reliable and relevant results, because of the (linear) relationship between the measured reaction velocity and the concentration of glucose in physiological and pathological circumstances.

Name: **Studentnummer:** | | | | | | | |

Test matrix question 1

Objectives:

Q1	**RES - Safety, techniques & Enzymes in practice**
Main Objective	
Objective 1	You are able to use standard wet laboratory equipment in a way that is safe for yourself, your colleagues, and the environment
Objective 2	You have developed basic skills for performing laboratory experiments in a proper way
Objective 3	You are able to report and discuss the results in written form and present them in a group
Objective 4	You are able to recognize sources of error and their impact on your results
Objective 5	You are able to measure the main characteristics of enzymes
Objective 6	You are able to use enzymes for fundamental (research) and clinical purposes
Objective 7	You are able to properly interpret results that are obtained with the use of enzymes, both from the clinician's and from the patient's point of view

Matrix:

Question	a	b	c	d	e	f	g
Objective	5	5	7	6			

Name: **Studentnummer:** | | | | | | | |

Question 2
Modelling: Blood pressure – dr. T. Oostendorp
(15 points)

An important factor in understanding the blood pressure $P(t)$ in the aorta is the fact that the aorta is highly elastic: its volume $V(t)$ depends on $P(t)$ according to $V(t) = C \cdot P(t)$. The constant C (in Pa/ℓ) is called the compliance of the aorta.

Figure 2.1 shows the so-called windkessel model of the blood pressure. In this model, the heart ejects blood into the aorta with a flow rate $\Phi(t)$ (in ℓ/min), which will cause the aorta to swell.
The blood flows back from the aorta to the heart through the peripheral blood vessels, with a flow rate $\Psi(t)$ that is proportional to the blood pressure: $\Psi(t) = P(t)/R$. The constant R is called the peripheral resistance.

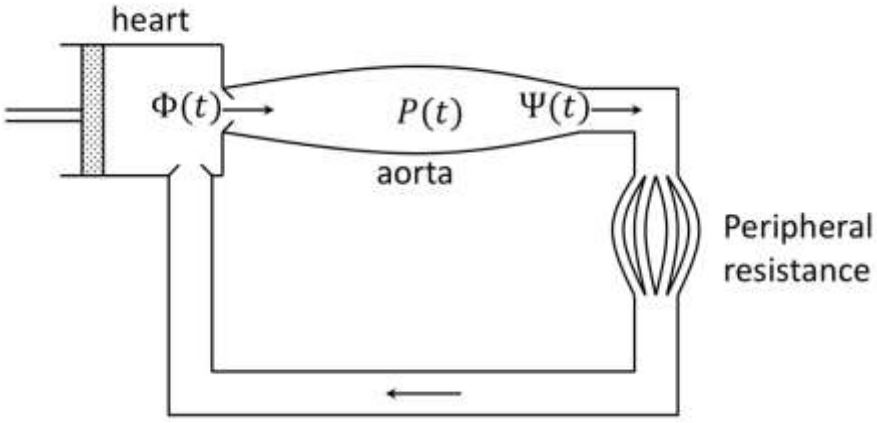


Figure 2.1 Windkessel model of blood pressure.

a. Show that the system equation for this model is (5 pt):

$$C\frac{d}{dt}P(t) = \Phi(t) - \frac{1}{R}P(t) \qquad (1)$$

We start with change=in-out.

For “in” we take the blood flowing into the aorta: $\Phi(t)$

For “out” we take the blood flowing out off the aorta: $\Psi(t) = P(t)/R$

Apparently, we are talking about the change in amount of blood in the aorta, which is equal to the volume of the aorta: $V(t) = C \cdot P(t)$.

“change” is the derivative of the amount of blood: $\frac{d}{dt}V(t) = \frac{d}{dt}V \cdot P(t) = V\frac{d}{dt}P(t)$

For an average adult in rest, the heart pumps 5 ℓ/min into the aorta. The mean blood pressure in the aorta is then 13 kPa.

b. Calculate the peripheral resistance for an average adult in rest. Include the correct unit. (3 pts)

We are talking about the equilibrium situation (rest, mean blood pressure). In that situation, the blood pressure is (on average) constant, hence $\frac{d}{dt}P(t) = 0$. Applying this to the system equation we get

$$\bar{\Phi} - \frac{1}{R}\bar{P} = 0$$

(The bars above Φ and R indicate that these are mean values, but the students don't need to indicate that)

$$R = \frac{\bar{P}}{\bar{\Phi}} = \frac{13}{5} = 2.6 \text{ kPa min/}\ell$$

The flow from the heart into the aorta is not constant; between two heartbeats, the flow $\Phi(t)$ is zero. As a result, the blood pressure drops between two heart beats.

During exercise the peripheral resistance decreases. This influences the velocity by which the blood pressure drops after a heartbeat.

c. Use the system equation (1) to explain how the velocity by which the blood pressure drops after a heartbeat changes if the peripheral resistance decreases. (3 pt)

Between two heartbeats, we have $\Phi(t) = 0$, so $C\frac{d}{dt}P(t) = -\frac{1}{R}P(t)$, and from this $\frac{d}{dt}P(t) = -\frac{1}{RC}P(t)$. If the value of R decreases, the derivative will be more negative, so the blood pressure will drop faster.

d. Show that, between two heartbeats (when there is no flow from the heart to the aorta), the blood pressure is given by (4 pts):

$P(t) = P_0 \mathrm{e}^{-t/RC}$ (2)

Between two heartbeats, we have $\frac{d}{dt}P(t) = -\frac{1}{RC}P(t)$. To check whether $P(t) = P_0 \mathrm{e}^{-t/RC}$ fits we substitute it into the system equation:

$$\frac{d}{dt}P(t) = -\frac{1}{RC}P(t)$$

$$\frac{d}{dt}\left(P_0 \mathrm{e}^{-t/RC}\right) \stackrel{?}{=} -\frac{1}{RC}\left(P_0 \mathrm{e}^{-t/RC}\right)$$

$$-\frac{1}{RC}\left(P_0 \mathrm{e}^{-t/RC}\right) \stackrel{?}{=} -\frac{1}{RC}\left(P_0 \mathrm{e}^{-t/RC}\right)$$

Yes, it fits.

Test matrix question 2

Objectives:

Q1	**RES - Introduction to the modelling of physiological systems**
Main Objective	***You apply basic models in different fields of biomedical science***
Objective 1	You can deduct the system equation of a model by reasoning "change=input-output"
Objective 2	For simple cases, you can show whether a proposed solution fits the system equation
Objective 3	You can deduct basis properties of the model from the system equation, such as the stationary solution or the nature of the null response

Matrix:

Question	a	b	c	d
Objective	1	3	3	2

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

The registration of the electric activity of the brain using electrodes on the head is called ElectroEnchephalography (EEG). Figure 3.1 shows an example of EEG for different brain activities. The bars at the right all indicate 50 μV. Figure 2.1 shows the spectrum of one the recordings of figure 3.1.

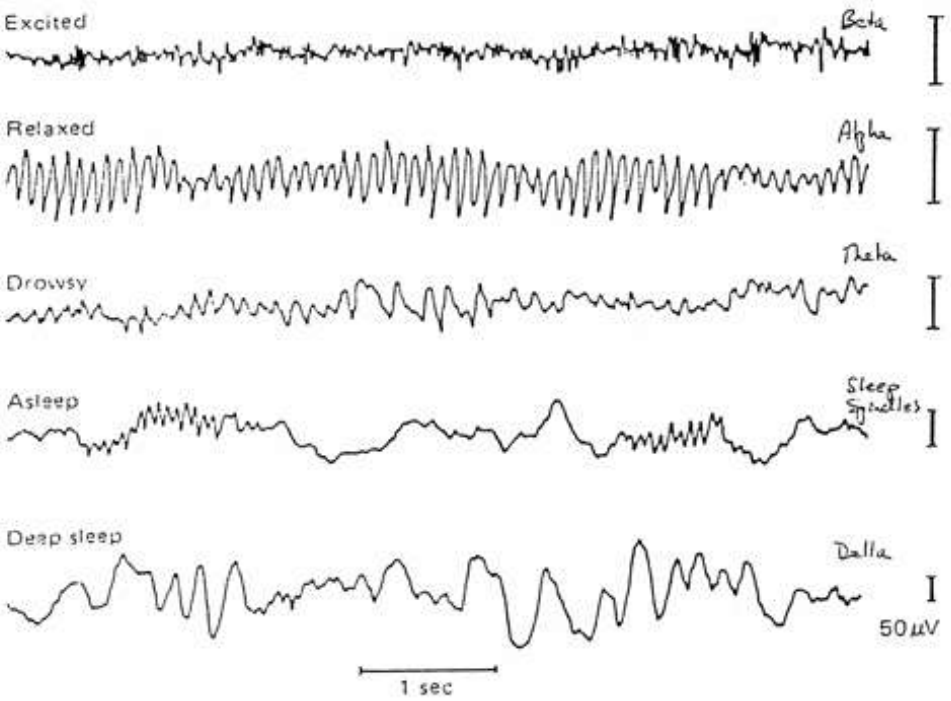


Figure 3.1 EEG recordings for different brain activities.

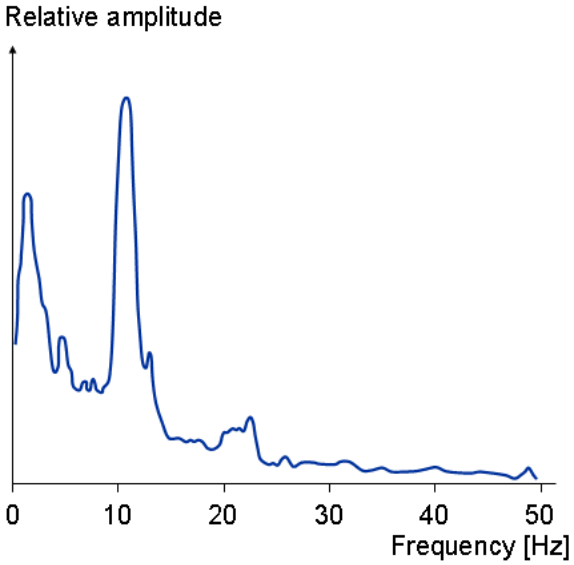


Figure 3.2 Spectrum of an EEG recording.

The notes of the researcher who recorded the data are unclear: he is not sure whether the spectrum of figure 3.2 is that of the second trace in figure 3.1 (marked “Relaxed”) or of the last one (marked “Deep sleep”).

a. To which of the two traces does the spectrum correspond? Explain your answer. (3 pt)

The spectrum shows a clear peak at 10 Hz. The second trace display clear sine waves of about 10 Hz (about 10 cycles fit into the 1 s bar); the last trace does not have a clearly dominating frequency at 10 Hz. So it must be the second trace (“Relaxed”).

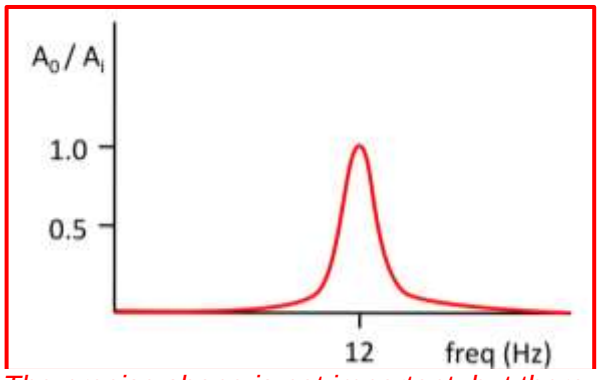
The researcher wants to record EEGs for different brain states. He wants to minimize the sample rate. He also needs to be able to determine whether the signal is stronger at 12 or at 25 Hz.

b. What sample rate should he choose? Explain your answer. (3 pts)

The highest frequency component he needs to be able to observe is 25 Hz. In order to do that, the sample rate should be at least 50 Hz. *Answers between 25 and 50 Hz are acceptable (it is actually wise to go a bit higher to be on the save side), but in order to detect the frequency it is definitely not necessary to have 10 sample per cycle (250 Hz).*

For another research, the researcher wants to record EEGs the spectrum of which he expects to look like that of figure two. He intends to filter the signal, so that only the component corresponding to the peak of about 12 Hz in the spectrum remains.

c. Sketch the amplitude response of the filter he should use. Indicate what is plotted along the axes. (4 pts)



The precise shape is not important, but there must be a peak at 12 Hz and zero elsewhere

Test matrix question 3

Objectives:

Q2	**RES - Recording and analyzing physiological signals**
Main Objective	***You can record and analyze physiological signals***
Objective 1	You can record physiological signals into the computer for further analysis using proper settings
Objective 2	You understand what the spectrum of a physiologic signals is, and can use a computer to determine the spectrum of a recorded signal
Objective 3	You know how to use filtering to improve the signal-to-noise ratio of physiological signals

Matrix:

Question	a	b	c
Objective	2	1	3

The following abstract and Tables belong to question 4 and 5.

Neck circumference and future cardiovascular events in a high-risk population--A prospective cohort study.

Yingnan Dai, Xiaojing Wan, Xin Li, Enze Jin and Xueqi Li. Lipids in Health and Disease 2016;**15**:46. https://doi.org/10.1186/s12944-016-0218-3

Abstract

Background - The distribution of adipose tissue has been evaluated in relation to cardiovascular risk factors and biochemical components of the metabolic syndrome. Neck circumference (NC) has been shown to have a strong relationship with cardiovascular disease (CVD) and may be a novel indicator of CVD. The aim of this study was to compare the incidence of CVD events in cohorts with different NC distributions, and to correlate NC with future CVD events and relative mortality.

Methods - A prospective cohort study was performed on 12,151 high-risk cardiology outpatients from 2004 until 2014. Anthropometric parameters like body mass index, NC, waist circumference, and hip circumference were measured at baseline and follow-up and compared in different cohorts with high, medium, and low NC. Fatal and non-fatal CVD events were compared in the follow-up study, and survival analysis was conducted. Independent Chi-square tests were performed to compare the incidence of CVD events and mortality among the cohorts and analyze the interactions.

Results - The subjects comprised of 6696 women and 5819 men who completed a mean 8.8-year follow-up. All of the participants had two or more CVD risk factors at baseline. At the end of the study, 4049 CVD events had occurred in 2304 participants. The incidence of non-fatal CVD events was 14.08, 16,65, and 25.21 % in the low-NC, medium-NC, and high-NC cohorts, respectively ($p < 0.001$). The all-cause mortality was 9.77, 11.93, and 19.31 %, and CVD mortality, 4.00, 6.29, and 8.01 %, respectively ($p < 0.001$). Compared with baseline, the number of CVD risk factors in participants had increased from 2.6, 3.0, and 3.4 to 3.5, 4.1, and 4.7 in the low-, medium-, and high-NC cohorts (34, 36, and 38 %), respectively. The event-free survival rate was 95.32, 80.15, and 75.47 %, respectively.

Conclusions - A higher NC indicated a higher incidence of future fatal and non-fatal CVD events and all-cause mortality in both male and female high-risk participants. CVD risk factors increased more in the higher NC group. NC as a novel indicator of CVD showed good predictive ability for CVD events and mortality in a high-risk population.

Table 1. Characteristics of participants at baseline

	Women	**Men**	***p* value**
	***n* = 6696**	***n* = 5819**	
Age (years)	51.05 ± 15.67	44.29 ± 18.1	0.001
BMI (kg/m^2)	26.85 ± 5.25	26.09 ± 3.75	NS
NC (cm)	36.31 ± 7.84	38.19 ± 5.94	NS
WC (cm)	84.51 ± 8.56	87.96 ± 12.56	NS
HC (cm)	99.1 ± 8.66	94.66 ± 12.68	<0.001
WHR	0.89 ± 0.08	0.97 ± 0.06	<0.001
FPG (mmol/L)	6.16 ± 1.39	6.86 ± 1.31	NS
HbA1c (%)	5.92 ± 0.67	6.13 ± 0.98	NS
TC (mmol/L)	5.51 ± 1.19	5.29 ± 1.09	NS
TG (mmol/L)	1.79 ± 0.98	2.01 ± 1.19	<0.001
HDL-C (mmol/L)	1.53 ± 0.35	1.33 ± 0.26	NS
LDL-C (mmol/L)	3.51 ± 1.08	3.48 ± 0.88	NS
SBP (mm Hg)	135.71 ± 15.14	133.41 ± 16.84	NS
DBP (mm Hg)	77.66 ± 8.47	78.96 ± 8.65	NS
Smoking (n, %)	2690 (40.17 %)	3547 (60.96 %)	NS
Alcohol drinking (n, %)	893 (13.34 %)	985 (16.93 %)	NS

Values represent mean ± SD, n (%). A p value ≤ 0.05 was considered statistically significant. *NS* not significant ($p > 0.05$); *BMI* body mass index; *NC* neck circumference; *WC* waist circumference; *HC* hip circumference; *WHR* waist-hip ratio; *FPG* fasting plasma glucose; *HbA1c* hemoglobin A1c; *TC* total cholesterol; *TG* triglyceride; *HDL-C* high-density lipoprotein cholesterol; *LDL-C* low-density lipoprotein cholesterol; *SBP* systolic blood pressure; *DBP* diastolic blood pressure

Table 2. Differences in CVD risk levels among NC groups between baseline and follow-up

	NC < 36 cm		36 cm ≤ NC < 40 cm		NC ≥ 40 cm		*p* values	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Time**	**Group**
n	*n* = 2759	*n* = 2282	*n* = 6059	*n* = 5254	*n* = 3697	*n* = 2760	-	-
Age (year)	46.33 ± 18.6	56.65 ± 19.12	48.64 ± 17.84	57.42 ± 18.81	47.96 ± 16.58	54.56 ± 16.28	-	0.632
BMI (kg/m^2)	24.53 ± 5.47	28.4 ± 5.91	26.48 ± 4.75	30.4 ± 3.86	28.18 ± 6.92	30.34 ± 6.12	<0.001	0.05
NC (cm)	31.91 ± 4.2	32.5 ± 4.45	37.67 ± 1.58	37.23 ± 1.73	41.23 ± 1.21	41.55 ± 1.45	0.051	<0.001
WC (cm)	83.26 ± 10.33	84.34 ± 11.64	86.8 ± 9.72	87.54 ± 10.73	87.21 ± 10.96	89.54 ± 12.75	<0.001	0.001
HC (cm)	92.71 ± 10.43	92.53 ± 12.85	96.94 ± 9.22	97.22 ± 9.6	100.55 ± 10.53	103.65 ± 10.35	0.127	<0.001
WHR	0.88 ± 0.14	0.89 ± 0.16	0.92 ± 0.13	0.93 ± 0.15	0.97 ± 0.13	0.99 ± 0.16	<0.001	<0.001
FPG (mmol/L)	6.28 ± 1.33	6.25 ± 1.34	6.46 ± 1.26	6.59 ± 1.22	6.68 ± 1.29	6.77 ± 1.28	0.032	0.004
HbA1c (%)	5.78 ± 0.84	5.98 ± 1.05	5.98 ± 0.67	6.12 ± 0.73	6.25 ± 1.02	6.42 ± 1.35	0.048	0.137
TG (mmol/L)	1.72 ± 0.94	1.78 ± 1.08	1.9 ± 0.91	1.95 ± 0.96	2 ± 0.94	2.15 ± 1.11	0.027	<0.001
TC (mg/dL)	5.17 ± 1.17	5.32 ± 1.01	5.43 ± 1.15	5.67 ± 1.74	5.54 ± 1.67	5.73 ± 1.83	0.019	0.082
HDL-C (mg/dL)	1.67 ± 0.37	1.65 ± 0.27	1.43 ± 0.34	1.25 ± 0.36	1.27 ± 0.29	1.13 ± 0.36	0.006	<0.001
LDL-C (mg/dL)	3.17 ± 0.88	3.26 ± 0.93	3.57 ± 0.95	3.62 ± 1.02	3.61 ± 0.86	3.96 ± 1.01	0.034	0.116
SBP (mm Hg)	133.62 ± 16.78	136.9 ± 15.48	134.63 ± 14.95	137.21 ± 15.17	135.42 ± 14.96	140.07 ± 16.39	0.048	0.137
DBP (mm Hg)	76.27 ± 8.46	76.2 ± 7.7	77.55 ± 8.15	77.34 ± 7.98	80.92 ± 8.95	81.11 ± 9.35	0.166	0.039

Data represent mean ± SD. P interaction is age adjusted. Abbreviations as in Table 1

Question 4
Epidemiology: Methods, measures and study design – dr. F. de Vegt
(20 points)

a. Which two anthropometric measurements are needed to calculate the variable Body Mass Index (BMI) and how is the formula? (2 pt)

Body weight and height. Body weight (kg) / (height (m) * height (m)

b. See Table 2. Give a description and interpretation of the results concerning the BMI at baseline and at follow-up in the participants with a neck circumference < 36 cm. (4 pnt)

Participants with a neck circumference < 36 cm had at baseline a BMI of 24.53 kg/m^2, with a standard deviation of 5.47 kg/m^2. At follow-up the mean BMI was 28.4 kg/$m^{2,}$ in this group, with a standard deviation of 5.91 kg/m^2 . So large variation between persons. So at follow-up the BMI increases with nearly 4 BMI points. Remark that at follow up, the number of participants had decreased by 450!

c. List two other ways to obtain information about obesity from humans, define the measure for obesity that you will obtain with these, and mention one advantage and one disadvantage for both measurement methods. (4 pnt)

1. By measuring waist and hip circumferences
Measure: WH-ratio, how the fat is distributed around hip and waist
+ more objective compared to questionnaire data
- difficult to measure (variation between measurements)
- people are not doing this when asked for in for example a questionnaire, so low response

2. By measuring skinfolds
Measure: the body fat percentage.
+ more objective compared to questionnaire data
- Not easy to obtain, participant needs to come to hospital / research center
- Difficult to measure (variation)
- Need equipment (skin calipers)

3. By BIA measurements
Measure: body fat percentage
+ easy to obtain, a lot of scales do have BIA included to calculate body fat proportion
- Not everyone has such a scale, in that case participant needs to come to hospital / research center

In this study the number of CVD events were counted.

d. Explain why this number is an incidence and not a prevalence (2 points)

Because this is a prospective cohort study, so there is a follow-up of several years. At baseline the participants had no CVD event developed yet (in that case it would have been a prevalence measure). So after several years the new CVD events are counted in a population free of CVD events at baseline, which is an incidence measure.

e. In this study, a prospective cohort study design was used. How would a case-control study been set up to study the same research question? Why is it hardly feasible to address the same research question in a case-control study? Explain your answer (4 points)

In a case control study, you start with the selection of cases and controls. In this case who are the cases? Subjects with CVD. Then you have to select controls (participants without CVD). You then have to know neck circumferences of more than 8 years ago, and also information on confounders….. And this is probably not registered, and participants don't know the value of their neck circumference of 8 years ago…

See the abstract, results section: 'The incidence of non-fatal CVD events was 14.08%, 16.65%, and 25.21% in the low-NC, medium-NC, and high-NC cohorts, respectively.'

f. Calculate the relative risk (RR) and the attributable risk (AR) of a non-fatal CVD for a high neck circumference compared to low neck circumference. (2 points)

RR: 25.21 / 14.08 = 1.79
AR: 25.21 – 14.08 = 11.13%

g. Give an interpretation for the relative risk (RR) you calculated in question g. (2 points)

Participants with a neck circumference > 40 cm had a 1.79 times higher risk for developing a non-fatal CVD event compared to participants with a neck circumference < 36 cm.

Test matrix question 4

Objectives:

Q1	**RES - Methods and measures in human based research**
Main Objective	***You can describe the discipline of epidemiology, and some basic methods and measures involved in human based research***
Objective 1	You can describe various methods for measurements and data collection in humans
Objective 2	You can perform some anthropometric measurements and collect data upon body fat, bodyweight and body circumferences
Objective 3	You can apply knowledge gained in CSI concerning the basic measures to describe data.

Q2	**RES - Etiological research on pregnancy outcomes**
Main Objective	***You can design an etiological study and use and interpret association and impact measures in human based research***
Objective 1	You can design an etiological study, describe the methods of data collection and communicate your findings
Objective 2	You know the differences between observational study designs (cohort study, case control study) and intervention studies
Objective 3	You can describe, use and interpret measures of disease frequency, measures of association and impact measures

Matrix:

Question	a	b	c	d	e	f	g
Objective	Q1-1,2	Q1-3	Q1-1,2	Q2-3	Q2-2	Q2-3	Q2-3

Q2-1 in report

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

Please look back at Table 1 belonging to the abstract '*Neck circumference and future cardiovascular events in a high-risk population--A prospective cohort study*'**.**
(before question 4)

a. Show that the overall proportion of smokers in this sample is approximately 0.5. (2 pt)

In total there are 2690 + 3547 = 6237 smokers. The total sample size is in the abstract given as 12151 but appears to be 6696 + 5819 = 12515. So the proportion smokers is 6237/12515 = 0.498. Indeed approximately 50% of the participants are smokers.

b. Suppose one would choose a sample of 2500 patients from this population and suppose that the true proportion of smokers is equal to 0.5. Show that the standard error of the proportion smokers in the sample is equal to 0.01. (3 pts)

The standard error of a proportion is equal to $\sqrt{(\pi^*(1 - \pi)/n)}$, with π equal to the population proportion (0.5) an n equal to the sample size. Therefore the standard error equals $\sqrt{(0.5*0.5/2500)} = 0.01$.

c. Suppose one would choose a sample of 2500 patients from this population and suppose that the true proportion of smokers is equal to 0.5 so that the standard error equals 0.01. Show that approximately 95% of all sample proportions will be between 0.4804 and 0.5196. (3 pts)

The sample proportions will be approximately normally distributed with mean 0.5 and standard deviation/error 0.01. Approximately 95% of all normally distributed values will be between mean – 1.96*standard deviation and mean + 1.96*standard deviation, where in this case the standard deviation equals the standard error. Filling in gives: 0.5 – 1.96*0.01 = 0.4804 and 0.5 + 1.96*0.01 = 0.5196

d. Again refer back to Table 1. The authors claim there is no significant difference between the proportions smokers amongst males and females. Is this claim valid? Give a proper motivation! (4 pts)

If there would be no difference in proportion, we should assume that both amongst the males and the females the proportion smokers would be equal to approximately 0.5. We have shown that in that case we would expect sample proportions between 0.48 and 5.2 in case we would choose samples of 2500 subjects. Both the sample males and the sample females are much larger, so we would expect sample proportions even closer to 0.5. However the sample proportion smokers amongst the males is approximately 0.6 and the sample proportions smokers amongst the females is approximately 0.4. It is very unlikely to find these sample proportions assuming that the proportion smokers is 0.5 for both the males and the females. Thus the claim of the authors is not valid.

In a separate study, the accuracy of the procedure to measure total cholesterol (TC) was evaluated. A standard solution with a TC of 5 mmol/L was made and five times a measurements was made. The average TC of these 5 measurements was 5.1 mmol/L and the standard deviation was 0.2 mmol/L.

e. How large is the random measurement error of the procedure? (2 pts)

The random measurement error is equal to the standard deviation: 0.2 mmol/L

f. The measurement error is rather large; would that pose a problem in the present study? Explain your answer (1 pts)

No, since the difference between the subjects is much larger than the measurement error. Probably the measurement error is too large for a clinical application where you evaluate individual values, but in a study we look at mean values over groups of subjects and the accuracy of these mean values will decrease due to the measurement error, but the effect of differences between the patients is much larger.

Name: **Studentnummer:** | | | | | | | |

Test matrix question 5

Objectives:

Q1	**RES - Measurement errors**
Main Objective	***You can use the standard deviation to quantify measurement error***
Objective 1	You can differentiate between systematic measurement error and random measurement error
Objective 2	You are able to establish the reliability of the measurements

Q2	**RES - From population to sample**
Main Objective	
Objective 1	You can use probability theory to derive sampling distributions and their properties
Objective 2	You can use SPSS to simulate and investigate the sampling distribution of the sample mean

Matrix:

Question	a	b	c	d	e	f	g
Objective	Q1O1 CSI	Q2O1 (CSI) + objective 3 below	Q2O1 (CSI) + objective 3 below	Q2O1 (CSI) + objective 3 below	Q1O1	Q1O2	

Q1	**RES - Methods and measures in human based research**
Main Objective	***You can describe the discipline of epidemiology, and some basic methods and measures involved in human based research***
Objective 1	You can describe various methods for measurements and data collection in humans
Objective 2	You can perform some anthropometric measurements and collect data upon body fat, bodyweight and body circumferences
Objective 3	**You can apply knowledge gained in CSI concerning the basic measures to describe data.**

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

a. Besides water and genomic DNA, the Polymerase Chain Reaction (PCR) mix contains several other essential components, such as oligonucleotides and ~~di~~deoxynucleotides. Provide a short explanation of what these compounds are and how they contribute to DNA amplification. (2 pt)

Oligonucleotide: A short single stranded DNA sequence (primer) that anneals to the target DNA and functions as the start of the copying process;
~~Di~~deoxynucleotides: These are the building blocks of DNA that are built in during elongation and serve to build the copied strand;
The addition of "di" to this question was an unfortunate error. This will be taken into account in the grading of this exam.
1 point per good answer and explanation

b. Explain why Copy number variations cannot be measured with Sanger sequencing. (2 pt)

Sanger sequencing provides information about the sequence of a piece of DNA, not about the amounts of copies of this target.

Below, you see a snapshot from the 1000Genomes browser with genotype information of SNP rs234532. This table contains genotype frequencies and numbers of persons per genotype (between brackets) for four different European populations (i.e. CEU=Northern and Western European ancestry, FIN=Finnish, GBR=British, IBS=Spanish).

1000GENOMES phase_3: **CEU**	**A|A**: 0.040 (4)	**A|G**: 0.384 (38)	**G|G**: 0.576 (57)
1000GENOMES phase_3: **FIN**	**A|A**: 0.000 (0)	**A|G**: 0.182 (18)	**G|G**: 0.818 (81)
1000GENOMES phase_3: **GBR**	**A|A**: 0.044 (4)	**A|G**: 0.275 (25)	**G|G**: 0.681 (62)
1000GENOMES phase_3: **IBS**	**A|A**: 0.449 (48)	**A|G**: 0.458 (49)	**G|G**: 0.093 (10)

c. Use the data in the above table to calculate the *allele frequency* of the A and G alleles in the IBS population (use box 1) and calculate the expected *genotype frequencies* (AA, AG and GG) for the IBS population under assumption of Hardy-Weinberg Equilibrium (use box 2). Show your calculations. (4 pt)

Box1 (3 decimals)	A	G
Allele frequency IBS		

Box1 (3 decimals)	*A*	*G*
Allele frequency IBS	*0.678*	*0.322*

***Method 1**: count the number of alleles:*
A: 2x48 + 49 = 145; G: 2x10 + 49 = 69; Total: 145 + 69 = 214 (= 2x107).
Allele frequencies: A allele: 145 / 214 = 0.678; G allele: 69 / 214 = 0.322
***Method 2**: Add up frequency homozygotes with half frequency heterozygotes*
A: 0.449 + 0.458/2 = 0.678; C: 0.093 + 0.458/2 = 0.322

Points awarded:
1: Right answer and calculations shown (2 pt)
2: Wrong method: square root of AA and/or GG frequencies (0 pt)
3: Wrong method: for example (freq AA)2 + freq AG (0 pt)
4: No calculation shown, but right answer (1 pt)
5: No calculation shown, wrong answer (0 pt)
6: Big calculation error (- 1 pt)
7: Minor calculation error due to rounding (- 0.5 pt)

Box 2 (3 decimals)	AA	AG	GG
Genotype frequency IBS			

Box 2 (3 decimals)	*AA*	*AG*	*GG*
Genotype frequency IBS	*0.460*	*0.437*	*0.104*

***Method**: use HW equation to calculate genotype frequencies:*
AA: (0.678)2 = 0.459
AG: 2 x 0.678 x 0.322 = 0.437
GG: (0.322)2 = 0.104

Points awarded:
1: Right answer and calculations shown (2 pt)
2: Right method shown, but with wrong numbers from Box 1 (2 pt)
3: No calculation shown, but right answer (1 pt)
4: No calculation shown, wrong answer (0 pt)
5: Big calculation error (- 1 pt)
6: Small calculation error due to rounding (- 0.5 pt)

d. Use the genotype frequencies from the previous question to calculate the expected number of people (use Box 3) and compare this with the actual numbers of people for each genotype in the IBS population. What is your evaluation of the differences between these genotype distributions? Show your calculations. (2 pt)

Box 3 (whole numbers)	AA	AG	GG
Actual number of people			
Expected number of people			

Box 3 (whole numbers)	*AA*	*AG*	*GG*
Actual number of people	*48*	*49*	*10*
Expected number of people	*49*	*47*	*11*

Calculation *of expected number of people (1 pt): Multiply genotype frequencies with total number of persons measured (= 107):*
AA: 0.460 x 107 = 49
AG: 0.437 x 107 = 47
GG: 0.104 x 107 = 11

Points awarded:
1: Correct calculation and answer (1 pt)
2: Right method with wrong numbers from question d (1 pt)
3: No calculation shown, right answer (- 0.5 pt), not if 2 points for lacking calculation were taken in previous answers
4: Total number of people in answer unequal to 107 (- 0.5 pt)
5: Calculation/rounding error (- 0.5 pt)

Evaluation *(1 pt): Minor differences between actual and expected (calculated) genotype distributions (0.5 pt). Consistent with HWE (0.5 pt).*

Name: **Studentnummer:** | | | | | | | |

Test matrix question 6

Objectives:

Main Objective	**You are able to perform basic genetic research techniques and interpret the data in terms of association between genetic variants and several specific phenotypes**
Objective 1	You collect data regarding genomic and phenotypic variation in a small group
Objective 2	You demonstrate your knowledge of genetic analysis techniques, the Hardy-Weinberg Equilibrium, the calculation and comparison of genotype and allele frequencies in different groups, and their use in genetic association studies
Objective 3	You use statistical analysis to identify genotype-phenotype correlations in the collected genomic and phenotypic data
Objective 4	You write a report with a clear, concise and accurate description of your findings, focusing on the introduction, statistical analysis methods, results and conclusions

Matrix:

Question	a	b	c	d			
Objective	2	2	2	2			