Hamburg University of Applied Sciences
Faculty Life Sciences

Influence of a single nucleotide polymorphism (rs12785878) on vitamin D deficiency in pregnant women near delivery: a logistic regression analysis

Bachelor Thesis

in Health Sciences

Submitted by

Wiebke Marie Feddern

Matriculation Number: [redacted]

Hamburg,

the 1st of October 2024

First reviewer:	Prof. Dr. Zita Schillmöller
	(Hamburg University of Applied Sciences)
Second reviewer:	Prof. Dr. Sabine Rohrmann
	(University of Zurich)

This thesis was written in collaboration with the University of Zurich.

Abstract

Background: Vitamin D deficiency is common among pregnant women in Switzerland. A deficiency has serious consequences for both the mother and the foetus. Genetic variants, such as single nucleotide polymorphisms (SNP), can influence vitamin D levels in pregnant women. This thesis investigates the effect of the SNP rs12785878 in gene 7-dehydrocholesterol reductase (DHCR7) on vitamin D deficiency in Swiss pregnant women in the third trimester.

Methods: This thesis uses data from a Swiss cross-sectional study that took place between 2014 and 2016. The target group included pregnant women shortly before delivery at the Zurich University Hospital (n=138). Socio-demographic and vitamin D related information were collected by a questionnaire and blood samples. A multivariate logistic regression is used to answer the research question.

Results: 52.17% of the sample is vitamin D deficient. The logistic model is adjusted for current BMI, supplementation of vitamin D, country of origin and skin type. The probability of vitamin D deficiency decreases with the presence of at least one T allele in the SNP rs12785878 [OR: 0.39, 95%-CI: 0.14 - 1.02].

Conclusion: The findings of this thesis indicate a link between a T allele in rs12785878 and a lower risk of vitamin D deficiency in pregnant women in Switzerland. Although not statistically significant, rs12785878 may play a role in vitamin D status. Future studies should focus on larger cohorts, regional differences, and additional predictors such as cholesterol metabolism. Standardised thresholds for defining vitamin D deficiency would improve comparability across studies and lead to more robust findings.

Content

List of Figures

List of Tables

List of Abbreviations

$1,25(OH)_2D_3$	1,25-Dihydroxyvitamin D
$25(OH)D_3$	25-Hydroxycholecalciferol
A	Adenine
BMI	Body mass index
C	Cytosine
C-section	Caesarean section
CYP24A1	Cytochrom P450 24A1
CYP27A1	Sterol 27-hydroxylase
CYP2R1	Cytochrom P450 2R1
DHCR7	7-Dehydrocholesterol reductase
DNA	Deoxyribonucleic acid
G	Guanine
HIV	Human immunodeficiency virus
NADSYN1	NAD synthetase 1
SNP	Single nucleotide polymorphism(s)
T	Thymine
UVB	Ultraviolet B radiation

List of Statistical Abbreviations

95%-CI	95% Confidence interval
eta^2	Eta squared
OR	Odds Ratio
r_{bs}	Effect size *r* of biserial correlation
rho	Effect size *rho* of Spearman's R
r_{ps}	Effect size *r* of point biserial correlation
SD	Standard deviation
V	Cramer's V
p-value	Probability value
β	Coefficient of a linear regression

Glossary

Term	Meaning
1,25-dihydroxyvitamin D ($1,25(OH)_2D_3$)	Active form of vitamin D, which can connect with receptors in basically every structure of the human body (Zittermann, 2022, p. 18)
25-Hydroxycholecalciferol ($25(OH)D_3$)	Biomarker for measuring vitamin D levels as it has the longest half-life in the body and thus represents the vitamin D situation most validly (Zittermann, 2022, p. 17)
7-Dehydrocholesterol reductase (DHCR7)	Gene in the 11th chromosome, influences the availability of 7-dehydrocholesterol in the skin since it regulates the 7-dehydrocholesterol reductase, which converts 7-dehydrocholesterol to cholesterol (Zittermann, 2022, p. 6)
Adenine (A)	Base of a nucleotide, the structure of the DNA; in combination with other bases the genetic code is formed (Schaaf and Zschocke, 2013, pp. 7–8)
BMI (Body mass index)	Measure for the classification of body weight; *body weight (in kilograms) divided by height (in meters) squared*
Chromosome	Genetic structure which contains the DNA; Humans have 46 chromosome (23 from the biological mother, 23 from the biological father) (Schaaf and Zschocke, 2013, p. 21)
CYP24A1	Gene which encodes the enzyme 1α-Hydroxylase, which converts to $1,25(OH)_2D_3$ (Zittermann, 2022, p. 18)
CYP27A1	Gene which encode the enzyme 25-hydroxylase (Zittermann, 2022, p. 18)

Term	Meaning
CYP2R1	Gene which encode the enzyme 25-hydroxylase, which converts vitamin D_3 to 25(OH)D_3 (Zittermann, 2022, p. 18)
Cytosine (C)	Base of a nucleotide, the structure of the DNA; in combination with other bases the genetic code is formed (Schaaf and Zschocke, 2013, pp. 7–8)
Deoxyribonucleic Acid (DNA)	Carrier of genetic information (Schaaf and Zschocke, 2013, pp. 7–8)
Gestational hypertension	*"Condition identified as hyperglycemia during the second or third trimester of pregnancy. If a woman appears positive for diabetes mellitus in the first trimester, then she must have pre--existing diabetes, either type 1 or type 2"* (Arshad et al., 2022)
Guanine (G)	Base of a nucleotide, the structure of the DNA; in combination with other bases the genetic code is formed (Schaaf and Zschocke, 2013, pp. 7–8)
HELLP-syndrome	H = Haemolysis EL = Elevated liver values LP = Low platelet count (decreasing platelet count ≤ 100 000/µl) (Föhl-Kuse, 2019)
Heterozygous	Due to chromosome pairing while fertilization, humans have two versions of each gene, therefore also of each gene localisation. If the bases at a specific gen location differ, it is called heterozygous individual with respect to the certain SNP

Term	Meaning
Homozygous	Due to chromosome pairing while fertilization, humans have two versions of each gene, therefore also of each gene localisation. If the bases at a specific gen location match, it is called homozygous individual with respect to the certain SNP
NADSYN1	Gene which is overlapping with DHCR7
Preeclampsia	*"Hypertension (>140 mmHg/>90 mmHg) diagnosed between 20 and 34 weeks of gestation, along with proteinuria (urine protein ≥0.3 g/24 h twice measured ≥6 h apart) together with other complications like liver dysfunction, hematological disturbance, and renal and neurological problems" (Arshad et al., 2022)*
Preterm delivery	Birth before the 37th pregnancy week (Arshad et al., 2022)
rs12785878	SNP in DHCR7; Major allele = G, minor allele = T
Single nucleotide polymorphism(s) (SNP)	Genetic variations in which one allele pair is replaced by another
Small-for-gestational-age infants	Foetuses or newborns who are smaller than average for their gestational age, typically defined as having a weight below the 10th percentile for their specific gestational age (Schlaudecker et al., 2017)
Thymine (T)	Base of a nucleotide, the structure of the DNA; in combination with other bases the genetic code is formed (Schaaf and Zschocke, 2013, pp. 7–8)

1 Introduction

A deficiency of vitamin D, as defined in this thesis as 25-hydroxyvitamin D ($25(OH)D_3$) blood levels less than 50 nmol/L, is a common occurrence among pregnant women[1]: In European countries, there are indications that between 20% and 80% of pregnant and breastfeeding women are insufficiently supplied with vitamin D (Palacios and Gonzalez, 2014). In Switzerland, a study showed that 53% of 305 Swiss pregnant women in the third trimester have a vitamin D deficiency (Krieger et al., 2018). In the first trimester, two studies implicate 73% (Christoph et al., 2020) and 63% (Cabaset et al., 2019) of Swiss pregnant women as being deficient. If these initial results are confirmed in future representative studies, 44,000 to 60,600 women of the average 83,000 Swiss pregnant women per year[2] would be affected.

The role of vitamin D in the human body becomes apparent upon closer investigation of its functions: It plays a primary role in the maintenance of bones and teeth, at the same time, a sufficient levels positively influences muscle function, cardiovascular health and the function of the immune system, including autoimmune diseases. Possible consequences of a persistent vitamin D deficiency include bone substance loss, increased blood pressure and coronary heart disease. Links with the occurrence of depression, multiple sclerosis, type 2 diabetes mellitus, dementia, alzheimer's disease and increased mortality have been discussed currently (Rebelos et al., 2023; Zittermann, 2022, pp. 30–35).

The recommended daily vitamin D intake for people between the ages of three and 60 is 15 µg. The same applies to pregnant women, despite being characterised by a unique dependency on vitamin D (Schweizerische Gesellschaft für Ernährung, 2019). Both, the placenta and foetal tissue have vitamin D receptors, leading to the fact that vitamin D influences for example the implantation of the fertilised egg in the uterus, the development of the foetal skeleton and the maturation of the immune system (Demay et al., 2024; Ströhle and Hahn, 2018).

The vitamin D metabolism is regulated by specific genes, which is why genetic variations, such as single nucleotide polymorphisms (SNP), can impact vitamin D status (Dastani et al., 2013). SNP rs12785878 in the gene 7-dehydrocholesterol reductase (DHCR7) has shown influence on $25(OH)D_3$ levels in the general population as well

[1] In this thesis, the term „(pregnant) women" includes biological women who are pregnant.
[2] Calculation based on the documentation for the last 3 years, Bundesamt für Statistik (2024).

as in pregnant women (Abu El Maaty et al., 2013; Moon et al., 2017; Wang et al., 2010). Data for the influence of this SNP in pregnant women in Switzerland is missing.

In light of this background, section 2 deals with the topic of human genetics and the significance of single nucleotide polymorphisms in chapter 2.1. Chapter 2.2 describes the development and effects of vitamin D, with chapter 2.2.1 explaining various factors that influence vitamin D status, while chapter 2.2.2 focusses on vitamin D deficiency during pregnancy. Section 3 provides a summary of the background information necessary for formulating the research question of this thesis. The methodology chosen to address this question is detailed in section 4, which includes an explanation of the quantitative data set used (chapter 4.1) and the statistical analyses performed (chapter 4.2). Section 5 presents the results of this thesis, followed by a discussion of these findings and the limitations of the work in section 6. The thesis ends with a comprehensive conclusion in section 7.

2 Theoretical Background

The theoretical background of this bachelor thesis forms the basis for the analyses carried out in the context of this work. It offers a comprehensive overview of relevant basic knowledge as well as research results, which present the latest state of the art on the subject to the author's best knowledge. Chapter 2.1 provides the necessary genetic background information. Fundamental information, such as the influence on vitamin D status and the situation of vitamin D deficiency in pregnancy, is reported in chapter 2.2.

2.1 Human genetics and the role of polymorphism

In the following, the basic genetic structures of humans are explained. Afterwards, the event of genetic variations is discussed in more detail.

The human body is composed of cells, the majority of which contain a cell nucleus. These nuclei contain 23 pairs of chromosomes, resulting in a total of 46 chromosomes. The chromosome pairs are formed when the chromosomes of the egg and the sperm - 23 chromosomes each - are brought together during fertilisation. This results in one set of chromosomes from the biological mother and one from the biological father (Schaaf and Zschocke, 2013, p. 21). Each chromosome contains Deoxyribonucleic Acid (DNA), presented in Figure 1, the carrier of the humans' genetic information. The DNA consists of two strands, each built from a sequence of many nucleotides. A nucleotide is constituted of a sugar molecule and a phosphate group, both forming the structure of the nucleotide. Additionally, a nucleotide consists of a base, which carries the genetic information. There are four bases, also referred to as 'allele': Adenine (A), Guanine (G), Thymine (T) and Cytosine (C), included in Figure 1 as yellow, blue, green or red structures, respectively. The sequence of the nucleotides forms the genetic code. The two strands of DNA run antiparallel, which means they run in opposite directions, with the end of one strand pointing towards the beginning of the other. In Figure 1, this is demonstrated by both arrows. The strands are wound around each other and connected to each other via complementary base pairings. The bases always pair specifically: A with T and G with C. If, for example, base A is located at one specific location of a strand, base T is automatically found at the same position on the other strand (Schaaf and Zschocke, 2013, pp. 7–8).

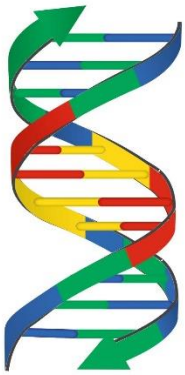

Figure 1: Simplified representation of the DNA. The DNA consists of two antiparallel strands of nucleotides. A sugar, a phosphate group, and one of four bases - adenine (yellow), guanine (blue), thymine (red), or cytosine (green) - compose each nucleotide. Complementary base pairings (A with T, G with C) connect the strands. Figure of Schaaf and Zschocke (2013, p. 8)

Humans have around 20,000 genes (Schaaf and Zschocke, 2013, p. 6). Genes are sections of the DNA, the beginning and end of which are marked by a specific sequence of bases. Between these start and end codes lies a sequence of bases typical for the gene (Schaaf and Zschocke, 2013, p. 12). However, changes can occur in these gene sequences, for example through the exchange of one nucleotide's base pair for another. The presence of such an exchange is called single nucleotide polymorphism (SNP). These genetic exchanges are inherited (Schaaf & Zschocke, 2013, p. 32). The base which is the most frequent in a population at a specific gene location is referred to as the major allele, as the least frequent base is referred to as the minor allele (Harrison et al., 2024a; Kanaka et al., 2023). Within that definition, a population can be the global population as well as a population of a specific region, such as Europe or Africa. Following that, there may be differences in the major allele when looking at different regions (Harrison et al., 2024b). SNP can occur in the section of a gene that is responsible for the production of proteins, which are the building blocks of many processes that take place in the body, for example metabolic processes. On the other hand, they can be located in the non-coding sections of a gene. There, it can also influence the production of a protein, for example by having a regulatory effect (Schaaf and Zschocke, 2013, p. 9).

Due to chromosome pairing during fertilisation, humans have two versions of each gene (one from the biological father, one from the biological mother) and therefore also of each gene localisation. In the event that the bases in question are identical in the respective gene localisation, this thesis refers to a homozygous individual. A distinction is made between a homozygous major allele and a homozygous minor allele. If the copies differ, the individual is considered heterozygous.

In conclusion, SNP influence numerous gene-controlled biological processes in the body. Long-term use of this knowledge enables the implementation of tailored prevention and action measures to avert the adverse effects of SNPs.

2.2 Vitamin D

This chapter starts with a description of vitamin D's function and the factors that influence vitamin D status. Subchapter 2.2.1 examines the genetic influence, with a particular focus on the DHCR7 gene. Subchapter 2.2.2 reviews the prevalence and associated factors of vitamin D deficiency in pregnant women.

Vitamin D is a generic term for several metabolites, which are linked in a common metabolism in the human body. The measurement of vitamin D in people usually takes place through 25(OH)D_3, as it has the longest half-life in the body and thus represents the vitamin D status most reliably (Zittermann, 2022, p. 17).

Humans have an estimated need of 15 µg per day (Schweizerische Gesellschaft für Ernährung, 2019), since vitamin D contributes to

- the maintenance of bone health,
- the maintenance of teeth health,
- a physiological calcium level in the blood,
- a physiological absorption/utilisation of calcium/phosphorus,
- a physiological function of the immune system (physical defence),
- a physiological muscle function as well as
- a function in cell division (Zittermann, 2022, p. 19).

2.2.1 Influence on vitamin D status

In the following, influencing factors on vitamin D status are explained, including environmental circumstances and individual characteristics. The chapter also describes the metabolism of vitamin D. The last part focusses on the influence of the SNP rs12785878 of the gene DHCR7 on vitamin D status.

The synthesis of vitamin D_3, a vitamin D metabolite from which all other vitamin D metabolites are derived from, is characterised by two ways (Zittermann, 2022, p. 18). The first way is the production of vitamin D_3 via cutaneous synthesis by means of the sun's ultraviolet B radiation (UVB) (a detailed description of the vitamin D metabolism follows on page 7). Therefore, the vitamin D production can be influenced by different factors (Zittermann, 2022, pp. 7–10). Regarding cutaneous synthesis, it should be

noted that Switzerland is located between 47- and 54-degrees north latitude, which means that the UVB intensity varies from season to season. The maximum is reached in summer, the minimum in winter, leading to a negligible cutaneous vitamin D synthesis in winter (Zittermann, 2022, p. 9). Other local characteristics affect the vitamin D production as well, for example the time of the day, since the UVB intensity is highest between 12 pm and 3 pm and lowest before 9 am and after 5 pm. Also, the altitude has an influence: The UVB intensity increases by approximately 5% to 20% per 1000 metres of altitude. Snow reflects up to 80% of UVB radiation, which also increases UVB intensity at high altitudes. Another important individual factor that influences the cutaneous vitamin D synthesis is the skin type. In light skinned people, maximum vitamin D production can be achieved after just a few minutes of full sun exposure. In dark skinned people, this process can take up to 120 minutes (Zittermann, 2022, p. 9). Advancing age may also play a role, as the ability to synthesise vitamin D decreases by a factor of three in older people compared to 20-year-olds (Zittermann, 2022, p. 12). Other individual factors such as covering clothing, spent time in the sun and the use of skin/sun creams can also reduce vitamin D production (Zittermann, 2022, p. 8).

The second way by which the body acquires vitamin D_3 is through nutrition or supplementation. Few foods are naturally rich in vitamin D, such as eel, herring and beef liver, making it difficult to get enough vitamin D just from nutrition (Zittermann, 2022, p. 14). Diet is another factor that can promote vitamin D deficiency since obesity and BMI have been repeatedly linked to a vitamin D deficiency (Alzohily et al., 2024; Pereira-Santos et al., 2015).

As explained in chapter 2.1, SNP influence protein synthesis, which in turn influences metabolic processes. Multiple genes are involved in the vitamin D metabolism, so SNP represent a way of affecting vitamin D production by influencing the vitamin D metabolism. The phases of catabolism (breakdown reactions of metabolites) and anabolism (buildup reactions of metabolites) divide vitamin D metabolism. The catabolism of vitamin D primarily takes place in the bone and calcium metabolism, as elaborated on within the consequences of a vitamin D deficiency described in the beginning of chapter 2.2. The process of anabolism is visualised in Figure 2 and described below, with the DHCR7 gene directly and indirectly affecting that process.

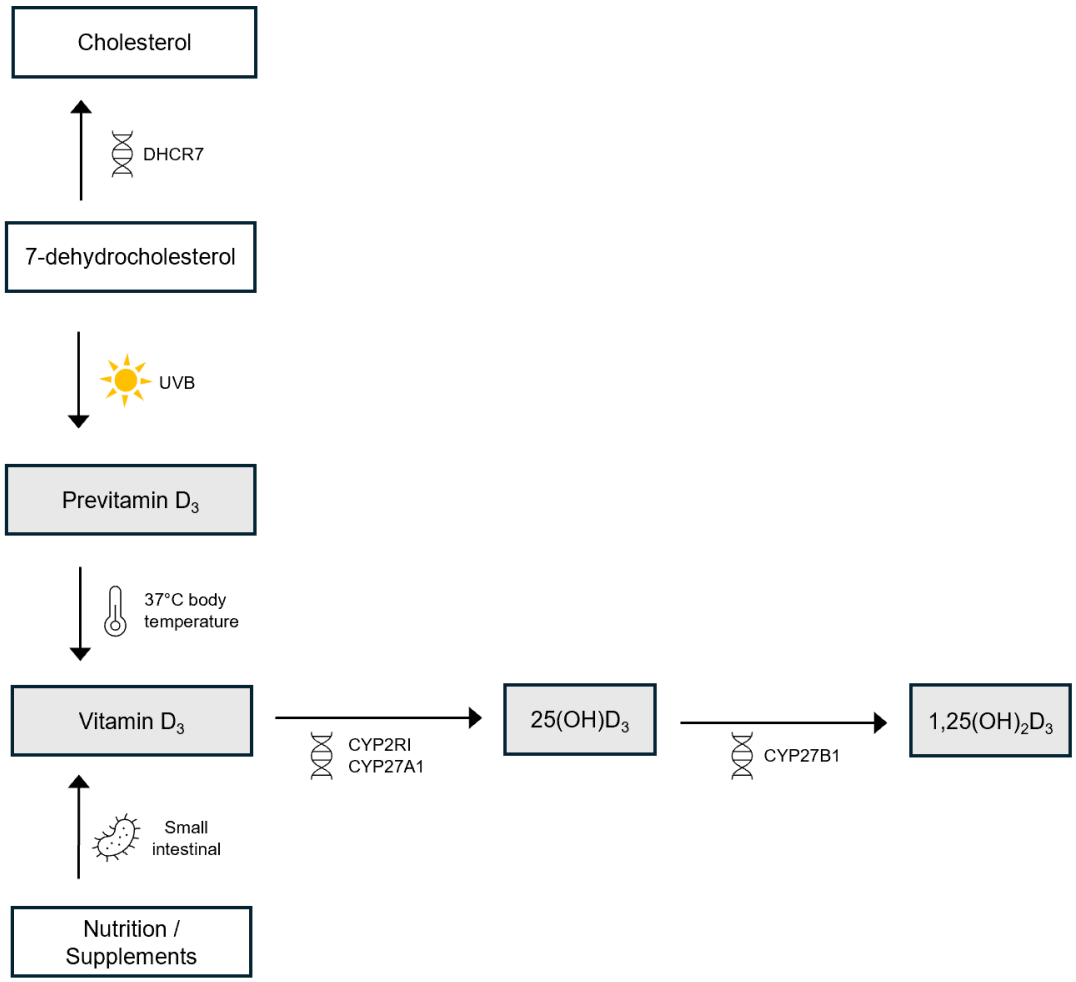

Figure 2: Simplified process of the vitamin D anabolism. Grey boxes demonstrate vitamin D metabolites. Own presentation, adapted from Zittermann (2022, p. 18)

As already explained, Vitamin D_3, from which the other vitamin D metabolites are gradually synthesised through conversion, is available from food or supplement intake or through cutaneous synthesis by means of the sun's radiation. UVB rays cause 7-dehydrocholesterol, located in the epidermis of the skin, to be converted into previtamin D_3. DHCR7 influences the availability of 7-dehydrocholesterol in the skin since it regulates the 7-dehydrocholesterol reductase, which converts 7-dehydrocholesterol to cholesterol. Then, previtamin D_3 is converted into vitamin D_3 when the human body temperature is above 37 °C, which is usually found to occur (Zittermann, 2022, p. 6). Irrespective of the synthesis way, the enzyme 25-hydroxylase, a result of the genes Cytochrom P450 2R1 (CYP2R1) and sterol 27-hydroxylase (CYP27A1), converts Vitamin D_3. The conversion leads to the vitamin D metabolite 25(OH)D_3, the biomarker for measuring vitamin D status. Afterwards, 25(OH)D_3 is converted to 1,25-dihydroxyvitamin D (1,25$(OH)_2D_3$) by means of the kidney and a enzyme synthesised through gene Cytochrom P450 24A1 (CYP24A1). 1,25$(OH)_2D_3$ is the active form of

vitamin D, which can connect with receptors in basically every structure of the human body (Zittermann, 2022, p. 18). But also 25(OH)D_3 receptors are found in multiple structures of the body, for example placenta, thyroid and the uterus (Zittermann, 2022, p. 19).

For all these genes and some of their SNP, an association to 25(OH)D_3 levels has been identified within the general population (Brouwer-Brolsma et al., 2016; Bu et al., 2010; McGrath et al., 2010; Qiu et al., 2020; Zhang et al., 2012). The effect of SNP can be both an increase and a decrease in 25(OH)D_3 levels, depending on the individual effect on protein synthesis (McGrath et al., 2010). The DHCR7 gene with its SNP, named rs12785878, has been identified as an influencing factor on 25(OH)D_3 levels in several studies, including a genomic analysis of 30,000 genes (Abu El Maaty et al., 2013; Ahn et al., 2010; Wang et al., 2010). DHCR7 and thus the SNP rs12785878 are found in the 11th chromosome. Globally, the major allele of the SNP is G, since approximately two third of the genomic information of people worldwide show such base at rs12785878. The minor allele is a T base, with the remaining one third (Harrison et al., 2024b). rs12785878 is located at an intron, which means it is not at a protein-coding section of the gene. However, the SNP can have a regulatory effect on the formation of proteins, as described in more detail in chapter 2.1. The DHCR7 gene partially overlaps with the NAD synthetase 1 (NADSYN1) gene, so that the DHCR7/NADSYN1 locus is often referred to in relation to rs12785878. However, in alignment with relevant literature, the term ‘DHCR7’ is used in this work.

This subchapter demonstrates that environmental factors such as sun exposure and individual factors such as skin type and genetic variations play a significant role in the vitamin D metabolism. Several studies, including a genomic analysis of 30,000 genes, have documented the relevance of the SNP rs12785878 in the DHCR7 gene for vitamin D synthesis.

2.2.2 Vitamin D deficiency in pregnancy

This chapter begins with the distinctive attributes of pregnant women in relation to vitamin D, followed by the prevalence of vitamin D deficiency. Afterwards, the consequences of vitamin D deficiency in pregnant women and the factors that predispose to a deficiency are described.

In Switzerland, there is an average of 83,000 pregnant women per year based on the documentation for the last three years (Bundesamt für Statistik, 2024). Pregnant women have particularly high nutrient requirements. Examples of the nutrients are

zinc, iron, iodine and vitamin A. The overall energy intake should also be slightly increased (Ströhle and Hahn, 2018). Regarding vitamin D, the Swiss Society for Nutrition (2019) does not recommend an increased intake for pregnant women when compared to the general population aged 3 to 60. Despite this, the vitamin D metabolism of pregnant women is different compared to non-pregnant women, since both the placenta and foetal tissue have vitamin D receptors. It is shown that the maternal vitamin D metabolism is linked to the foetus' vitamin D situation (Arshad et al., 2022). Subsequently, maternal vitamin D metabolism changes, which can lead to a decrease in maternal serum $25(OH)D_3$ levels in women who do not receive vitamin D supplementation as their pregnancy progresses (Pilz et al., 2018). The cross-sectional study 'Vitamin and mineral status among German women (VitaMinFemin)' with 429 German pregnant women and the same number of control cases showed that women in their third trimester are more likely to be vitamin D deficient compared to women in their first trimester [OR: 2.3, 95%-Confidence interval (95%-CI): 1.3 – 4.2]. They also noted that pregnant women are 3.7 times more likely to be vitamin D deficient than non-pregnant women [OR: 3.7, 95%-CI: 2.5 – 5.4] (Gellert et al., 2017).

In European countries (including Spain, Germany and the Netherlands among others), indications suggest that between 20% and 80% of pregnant and breastfeeding women are vitamin D deficient (Palacios and Gonzalez, 2014). Within a group of German pregnant women, the VitaMinFemin study showed that 78.1% are deficient (Gellert et al., 2017). With the help of the study 'Evaluation of Vitamin-D Status and Its Determinants in Switzerland and of Possibilities to Improve Vitamin-D Status in the Swiss Population', findings on the vitamin D status of pregnant women in Switzerland were investigated exclusively for those in the third trimester for the first time (Schweizerischer Nationalfond, 2014). In this context, 53% of 305 pregnant women in the third trimester were characterised by a vitamin D deficiency (Krieger et al., 2018).

A vitamin D deficiency has far-reaching consequences during pregnancy. Apart from the general consequences of vitamin D deficiency for humans, which are detailed in chapter 2.2, there are several additional aspects to consider during pregnancy. A deficiency in vitamin D can result in different hypertensive pregnancy disorders, such as

- gestational hypertension,
- preeclampsia, as well as
- the HELLP-syndrome (Hemolysis, Elevated Liver enzymes, and Low Platelets) (Demay et al., 2024).

The recently published guideline ‘Vitamin D for the Prevention of Disease: An Endocrine Society Clinical Practice Guideline’ describes the consequences of these conditions as follows:

> *“Hypertensive disorders of pregnancy increase risks for fetal growth restriction, small-for-gestational-age infants, and induced preterm delivery, with potentially serious and lifelong consequences for infant bone and brain development, as well as maternal and offspring long-term cardiometabolic health” (Demay et al., 2024)*

Following that, it is evident that the consequences of a deficiency extend beyond the biological mother to also affect the foetus. Given the unique vitamin D metabolism of pregnant women and the consequences of vitamin D deficiency for both mother and foetus, factors influencing the $25(OH)D_3$ levels in pregnant women should be studied separately. Three cross-sectional studies in Germany and Switzerland analyzed and determined influencing factors, all involving women in their third trimester. Table 1 provides an overview of these studies. The accompanying text reports the odds ratios (OR) related to the influencing factors, if available in the original paper. If this information is not available, the significance or any other statement of the paper in relation to the respective factor is stated. Every other study cited in the accompanying texts involves pregnant women in the third trimester, if not stated differently.

One study was conducted in Switzerland (Krieger et al., 2018), while two were conducted in Germany (Gellert et al., 2017; Wuertz et al., 2013). The studies took place between February 2012 and June 2016. Two of them only looked at women in the third trimester, while Gellert et al. (2017) included women from 2 to 41 weeks gestation. However, 217 women, about half of the sample of Gellert et al., were in their third trimester. Of the entire sample, 78.1% were deficient. Krieger et al. (2018) analysed a sample of 305 individuals and found 53.4% to be vitamin D deficient. In comparison, Wuertz et al. (2013) studied a sample of 261 individuals, with a higher deficiency rate of 77.0%. The studies selected different thresholds to calculate the factors influencing vitamin D status. Krieger et al. (2018) and Wuertz et al. (2013) used the threshold of vitamin D deficiency (< 50 nmol/L) in their calculations, while Gellert et al. (2017) based the calculation on the threshold < 25 nmol/L, which includes severe deficiency.

Table 1: Overview of cross-sectional studies conducted in Switzerland and Germany regarding the vitamin D situation in pregnant women in the third trimester, own presentation

Krieger et al. (2018)	Gellert et al. (2017)	Wuertz et al. (2013)
Country (city)		
Switzerland *(Zurich, Bellinzona, Samedan)*	Germany *(125 study sites)*	Germany *(Gießen)*
Sample size of pregnant women (pregnancy status)		
305 *(3rd trimester)*	429 *(2nd - 41st pregnancy week)*	261 *(3rd trimester – after birth)*
Time window of recruitment		
Aug 2014 - Jun 2016	Apr 2013 - Mar 2015	Dec 2010 – Feb 2012
Prevalence of vitamin D deficiency[1]		
53.4 %	78.1%	77.0 %
Threshold for calculation of influencing factors regarding vitamin D status		
< 50.0 nmol/L	< 25.0 nmol/L	< 50.0 nmol/L
Season		
Blood sampling in winter is more likely to show vitamin D deficiency ●	Blood sampling in spring, fall or winter vs. summer is more likely to show vitamin D deficiency ●	Blood sampling in winter is more likely to show vitamin D deficiency ●
Supplementation		
Supplementation of vitamin D is more likely to show vitamin D deficiency ●	NA	NA
Country of origin		
Another country of origin than Germany or Switzerland is more likely to show vitamin D deficiency ●	NA	Being non-European is more likely to show vitamin D deficiency ●
Skin type		
Dark skin type is less likely to show vitamin D deficiency ○	"Not related"	Medium and dark skin type is more likely to show vitamin D deficiency ○
BMI		
Higher near-term BMI is more likely to show vitamin D deficiency ○	Higher BMI at recruitment is more likely to show vitamin D deficiency ○	Pre-pregnancy BMI over 25 is more likely to show vitamin D deficiency ○
rs12785878		
NA	NA	NA
● significant ○ not significant [1] < 50.0 nmol/L		

These three studies contain influencing factors that have already been mentioned in relation to the general population (see chapter 2.2). Firstly, in all studies the **season** is shown to have an influence on vitamin D deficiency. Blood sampling in winter overall results in higher probability of vitamin D deficiency. Krieger et al. (2018) show significant[3] results in the comparison of summer vs. winter [OR: 0.33, 95%-CI: 0.13 - 0.84]

[3] Significance regarding odds ratios says, that the 95%-CI does not include the value '1', which indicates that the direction of the variable's effect is apparent

and autumn vs. winter [OR: 0.40, 95%-CI: 0.19 - 0.86]. Gellert et al. (2017) found that autumn [OR: 2.5, 95%-CI: 1.3 - 4.8], winter [OR: 13.5, 95%-CI: 6.4 - 28.2) and spring [OR: 9.5, 95%-CI: 4.2 - 21.6] showed in comparison to summer higher odds of vitamin D deficiency. Wuertz et al. (2013) even found an OR of 47.07 with an 95%-CI of 10.76 - 205.75 for winter blood sampling vs. summer. This picture is also confirmed at European level by a systematic review carried out in the Mediterranean region (including studies from Greece, Turkey, Spain, Italy and Tunisia) (Karras et al., 2016), as well as in a study from Slovenia (Dovnik et al., 2017). It has also been described internationally in China with the help of a systematic review (Chen et al., 2024).

Further, Krieger et al. (2018) identified **vitamin D supplementation** as an influencing factor that led to lower odds of vitamin D deficiency [OR: 0.42, 95%-CI: 0.22 - 0.80]. In addition, such an influence was also shown in Swiss women in the first trimester [OR: 0.33, 95%-CI 0.16 - 0.65] (Cabaset et al., 2019). Karras et al. (2016) also confirm this for the Mediterranean region. Internationally, a study from Taiwan found that the intake of foods containing vitamin D as well as supplementation was associated with higher 25(OH)D_3 levels in pregnant women (Huang et al., 2023).

In 2018, Krieger et al. found with respect to the **country of origin** that women from countries other than Germany or Switzerland are more likely to be vitamin D deficient. They compared these women to women from (1) ‘Northern America, Northern Europe, Caucasus, Central Asia, and New Zealand’ [OR: 2.4, 95% CI: 1.21 - 4.75], (2) ‘Southern Europe, Australia, Latin America, and the Caribbean’ [OR: 2.21, 95% CI: 1.08 - 4.52], (3) "South- and East Asia and Pacific" [OR: 7.81, 95% CI: 2.16 - 28.23], and (4) ‘Africa and the Middle East’ [OR: 4.74, 95% CI: 1.43 - 15.77]. In line with this, Wuertz et al. (2013) show that non-European women are more likely to suffer from a vitamin D deficiency [OR: 3.09, 95%-CI: 1.22 - 7.73].

Skin type is also a factor influencing vitamin D deficiency. For Wuertz et al. (2013) a medium or dark skin type compared to a light skin type results in a non-significant odds ratio of 1.70 [95%-CI: 0.85 – 3.35]. The relation between lighter skin types and lower 25(OH)D_3 levels has also been reported several times in Mediterranean regions (Karras et al., 2016). Also internationally, this has been shown within an Australian study [OR: 2.7, 95%-CI: 1.6 - 4.5] (Bowyer et al., 2009). Krieger et al. (2017) found contradictory data on this, as they found a dark skin type associated with a lower probability of vitamin D deficiency [OR: 0.41, 95%-CI: 0.15 - 1.12]. This result is not discussed further in the work of Krieger’s et al..

As another influencing factor of vitamin D deficiency, the **BMI** should be mentioned. Krieger et al. (2018) show that in their sample women with a higher BMI are slightly more likely to have a vitamin D deficiency [OR: 1.05, 95%-CI: 0.99 - 1.11]. Gellert et al. (2017) report similar values for their sample, with showing an OR of 1.10 [95%-CI: 1.0 - 1.1]. Wuertz et al. (2013) looked at the pre-pregnancy BMI, which led to an odds ratio of 1.72 [95%-CI: 0.79 - 3.74], indicating no significance, but the hypothesis of a higher BMI (> 25 kg/m2) resulting in higher odds of vitamin D deficiency. These trends are also shown in Canada within pregnant women in their first trimester. The significant results show that the odds ratio increases with higher BMI values. Compared to a BMI below 25, the odds ratio for a BMI between 30 and 35 is 2.00 [95%-CI: 1.4 - 3.1]. With a BMI over 35, the odds ratio rises to 2.8 [95%-CI: 1.7 - 4.5] (Woolcott et al., 2016).

Looking at the genetic influence, none of the studies from Germany and Switzerland have investigated the influence of **rs12785878** in the DHCR7 gene on vitamin D levels in pregnant women during the third trimester. Additionally, no evidence has been found for women in the first or second trimesters in these countries. DHCR7 and its SNP may be important because cholesterol metabolism undergoes significant changes during pregnancy, with levels increasing after the first trimester (Ghio et al., 2011; Wiznitzer et al., 2009). This shift may influence vitamin D metabolism, since DHCR7 converts 7-dehydrocholesterol to cholesterol and thus influences the availability for the conversion into previtamin D_3 (see chapter 2.2.1). The investigation of the potential interactions between vitamin D and cholesterol metabolism regarding DHCR7 is still part of recent research (Prabhu et al., 2016). Nevertheless, international studies have already shown associations of rs12785878 and 25(OH)D_3 levels in pregnant women in their third trimester. Moon et al. (2017), a study from the UK, found an association of rs12785878 and 25(OH)D_3 levels. The T allele is significantly[4] associated with 25(OH)D_3 levels [β = 3.1 nmol/L; 95%-CI: 1.0 - 5.2 nmol/L; p-value = 0.004]. So, if at least one T allele is found in the rs12785878 locus, the 25(OH)D_3 level appears to be higher. In another study from Indonesia a minimum of one T allele is associated with a lower probability of vitamin D deficiency [OR: 0.61, 95%-CI: 0.32 – 1.15], but not significantly (Aji et al., 2020). The same appears in women in Southeast China of the first trimester without supplementation, where the T allele is in a positive association with vitamin D (Shao et al., 2018).

[4] Significance within a linear regression is usually shown with a p-value below 0.05.

In conclusion, the reviewed studies highlight the unique challenges pregnant women face in maintaining adequate 25(OH)D_3 levels, particularly in the third trimester. Pregnant women seem to be more prone to vitamin D deficiency compared to non-pregnant women, as demonstrated by the findings from Germany (Gellert et al., 2017). Several key factors contribute to this deficiency, including season, BMI, country of origin, and vitamin D supplementation. Importantly, while significant efforts have been made to understand these factors, there remains a gap in research, especially regarding the genetic influence of rs12785878 on vitamin D deficiency in pregnant women in the third trimester.

3 Research Question

Vitamin D deficiency in pregnant women should be avoided due to its far-reaching consequences. It is therefore crucial to identify all influencing factors in order to effectively prevent a deficiency. This thesis focuses on the third trimester, as the risk of deficiency may increase as pregnancy progresses (Pilz et al., 2018). Several factors influencing vitamin D deficiency in pregnant women have already been identified, including the SNP rs12785878 in the DHCR7 gene. In Swiss pregnant women, factors such as season and country of origin have been recognised as significant. However, the role of the SNP rs12785878 in this context remains unclear. This SNP is of particular interest because DHCR7 not only affects vitamin D synthesis but also plays a role in cholesterol metabolism, which undergoes significant changes during pregnancy (Ghio et al., 2011; Wiznitzer et al., 2009). To determine whether pregnant women in Switzerland may be at increased risk of vitamin D deficiency due to genetic factors, this thesis addresses the following research question:

How does the SNP rs12785878 in the gene DHCR7 influence vitamin D deficiency in Swiss pregnant women shortly before delivery?

4 Methods

This chapter presents the methodology employed in this thesis. For a structured recording of the methodology and the results, this work is based on the STROBE statement as seen in the Appendix, which is assumed to cover all important aspects of the reporting of cross-sectional studies (Elm et al., 2008). Chapter 4.1 begins with a description of the study from which the data used originates, including a definition of the sample. The survey methods used in the thesis and the resulting information identified are then presented. Chapter 4.2 addresses the methodology used in this thesis. This includes the definition and determination of the predictors, as well as the data preparation (subchapter 4.2.1). The methodology of univariate and bivariate analysis (subchapter 4.2.2) and logistic regression (subchapter 4.2.3) is explained afterwards.

4.1 Dataset

The study 'Evaluation of Vitamin D Status and Its Determinants in Switzerland and of Possibilities to Improve Vitamin D Status in the Swiss Population', under the direction of Sabine Rohrmann and Katharina Christine Quack Lötscher and funded by the 'Schweizerischer Nationalfond' (project number: 145194), provided the data set for this work. The study provided findings on the vitamin D situation of pregnant women in the third trimester in Switzerland for the first time (Swiss National Science Foundation, 2024). The Zurich Ethics Committee accepted the study, which complied with the Declaration of Helsinki's 'Ethical Principles for Medical Research Involving Human Subjects' guidelines. Written informed consent was obtained from all participants. The following presents the aim, structure, and sample of the study.

The entire study was designed to determine the prevalence of vitamin D deficiency in mothers and newborns of different skin types as well as genetic and epigenetic determinants of $25(OH)D_3$ and $1,25(OH)_2D_3$ concentrations at three locations (Zurich, Bellinzona and Samedan). However, the University Hospital Zurich alone served as the collection site for the sub-project focusing on the genetic influences from which the data analyzed in this thesis originated.

In Zurich, women who had planned a caesarean section (c-section) at the University Hospital Zurich were recruited between August 2014 and June 2016. Table 2 displays the reasons for the inclusion as well as exclusion. A total of 138 women were finally included.

Table 2: Inclusion and exclusion criteria of the study underlying the analysed dataset, own presentation

Inclusion	Exclusion
• C-section • ≥ 37 weeks of gestation	• Twin pregnancy • Human Immunodeficiency Virus (HIV) • History of parathyroid, renal or liver disease • Chronic malabsorption syndromes • Granuloma forming disorders • Age below 18 years • Known/suspected drug or alcohol abuse • No genetic information

Information within the original study was collected by a questionnaire and a blood sample. The aim of the blood sample was to obtain information on $25(OH)D_3$ levels as well as other vitamin D-related blood parameters, for example $1,25(OH)_2D_3$, and genetic variations. The focus on genetic variations referred to genetic polymorphisms in candidate genes (i.e. DHCR7, CYP2R1, CYP24A1, GC, VDR) that have shown an association with $25(OH)D_3$ levels (Ahn et al., 2010; Bu et al., 2010; Wang et al., 2010). The blood sample of 10 ml was taken from the pregnant women during the c-section. The participant and a doctor completed the questionnaire during the final medical examination prior to the c-section, typically a few days before the birth. It contained different questions about working and free time habits in relation to sun exposure, clothing habits, consumption of vitamin-D-rich foods, use of pregnancy supplements, smoking, alcohol consumption, nationality, and education. The variables relevant to this thesis are presented in more detail in subchapter 4.2.1.

4.2 Statistical analysis

The quantitative methodology of this work is presented below, including the individual steps of the univariate, bivariate and multivariate analyses. All analyses are based on the programming language R 4.4.0 (R Core Team, 2024) using the software R-Studio Version 2024.04.1 for Windows (Posit team, 2024). The R packages used in the following analyses can be found in the Appendix. A significance level of 0.05 is established.

4.2.1 Predictors and data preparation

The influencing factors presented in subchapters 2.2.1 and 2.2.2 were selected for the predictors due to their relevance for the diagnosis of vitamin D deficiency in pregnant women. The following is a detailed definition of these variables as well as the outcome variable. The derivation of the variables is presented, and any adjustments made to the variables as part of this work are explained. It should be noted that any other adjustments outlined in the following sections, without reference to this work, have already been made in previous analyses conducted by other researchers. The predictors analysed in this thesis include the

1. SNP in DHCR7 (rs12785878),
2. season of birth,
3. current BMI,
4. supplementation of vitamin D,
5. country of origin,
6. as well as the skin type.

rs12785878 is the variable that represents the SNP in DHCR7. The variable was determined by means of a blood sample, after which SNP genotyping was done by an external laboratory. The alleles of the biological father and mother at the respective gene location were documented, resulting in the characteristics of this variable. Thus, this nominal variable has three categories in the original dataset: Major allele homozygous (G:G), minor allele homozygous (T:T) and heterozygous (G:T). For the categorisation, which allele is the major allele and which is the minor allele, this work follows the results of the global analysis of Ensembl[5] (Harrison et al., 2024b). For this thesis, the variable was dichotomised by defining one expression as ‘two major

[5] Ensembl is a website, which annotate genes, predicts regulatory function and collects disease data

alleles' and the other as 'at least one minor allele' in order to be able to represent the effect of a variant of at least one allele.

Season of birth describes a categorisation of the date of birth into seasons, which also represents the season in which the blood sample was taken, as it was taken from the mother at birth. The classification is based on the astronomical view of the seasons. This means that spring is defined from 20st March to 21st June, summer from 21st June to 23rd September, autumn from 23rd September to 21st December, and winter from 21st December to 20st March.

The current BMI is calculated using the following formula: Body weight (in kilograms) divided by height (in meters) squared. Thus, it is a metric variable, which is documented by the doctor at the last appointment before birth.

Supplementation of vitamin D was quantified through the inclusion of a free-text question in the questionnaire regarding the type and frequency of vitamin supplement intake. The questionnaire explicitly enquired about vitamin D supplements and multivitamins. The multivitamins were then evaluated for their vitamin D content as part of previous analyses, and a new variable was constructed to encompass the entirety of vitamin D supplementation. The values are binary (yes/no), indicating whether supplementation occurred.

The **country of origin** was initially collected via the questionnaire, with a format allowing free responses designated for the heading 'country of origin'. The answers were subsequently categorised following the regional groupings of the World Bank Map (The World Bank, 2018). Based on that, this new variable was created. Consequently, the variable 'Switzerland/Germany' represents the first group, 'Northern America, Northern Europe, Caucasus, Central Asia, and New Zealand' the second, 'Southern Europe, Australia, Latin America and the Caribbean' the third, 'South- and East Asia and Pacific' the fourth and 'Africa and Middle East' the fifth.

Skin type was determined by adapting the classification by Fitzpatrick (Fitzpatrick, 1988). The origin scale shows six different skin phenotypes and their reaction to sun exposure. The adapted scale used in the study includes five different skin types (see Figure 3). As part of the study, the participant was asked to judge their skin type based on the image and if uncertain in combination with the question "What happens to the untanned skin if it is exposed in the early summer at noon for 45 to 60 min to the sun?*".* The potential outcomes for these questions are as follows: (I) "painful sunburn after 24 h and no tanning after one week"; (II) "painful sunburn after 24 h and minimal tanning after one week"; (III) "mild sunburn (IV) "no sunburn after 24 h and tanned

after one week" to (V) "skin is deeply pigmented brown/black, no sunburn and tanned after one week". The interviewer then evaluated the participants' skin type. If the interviewer's and the participant's statements diverged, the rounded mean value of the two skin type numbers was used. The variable was dichotomised by defining skin types I-III as 'light skin type' and skin types IV-V as 'dark skin type'.

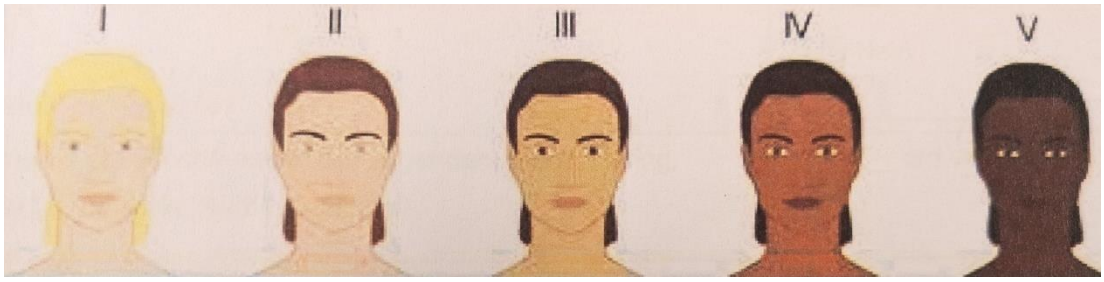

Figure 3: Scale used to determine skin type, adapted from Fitzpatrick (1988). Extract of the questionnaire of the study 'Evaluation of Vitamin-D Status and Its Determinants in Switzerland and of Possibilities to Improve Vitamin-D Status in the Swiss Population' (Schweizerischer Nationalfond, 2014)

The outcome, **vitamin D deficiency,** is derived from an initial measurement of the participants' 25(OH)D_3 levels. 25(OH)D_3 is the main circulating metabolite of vitamin D. For vitamin D deficiency a cut-off value of 50 nmol/L is defined, according to several authorities, including the Endocrine Society, the International Osteoporosis Foundation, and the Canadian Osteoporosis Society) (Dawson-Hughes et al., 2010; Hanley et al., 2010; Holick et al., 2011).

4.2.2 Uni- and bivariate analyses

The following section presents the procedure for the univariate and bivariate analyses. Initially, the handling of the respective levels of measurement is discussed. Subsequently, the statistical tests used with their effect sizes in the bivariate analysis are shown.

The first step in the data analysis is the univariate analysis. For the nominal variables (vitamin D deficiency, rs12785878, vitamin D supplementation, country of origin and skin type), the mode as well as the absolute and relative frequencies are calculated. The same is calculated for the ordinal variable (season of birth), adding the range. For the metric variable (current BMI), the median, mean with corresponding standard deviation and range are described. Table 3 provides an overview of the levels of measurement for the variables. Within the univariate analysis it is also reviewed whether a variable has more than 5% missing values. If that is the case, the variable will be left out in the further analyses (Azur et al., 2011).

Table 3: Levels of measurement of all variables, own presentation

Dichotomous nominal	Nominal	Ordinal	Metric
Vitamin D Deficiency rs12785878 Vitamin D supplementation Skin type	Country of origin	Season of birth	Current BMI

Depending on the level of measurement, the bivariate analysis is carried out using one of the statistical tests listed in Table 4 and their suitable effect measures. Each independent variable is correlated with the outcome in this context. In addition, all independent variables are checked for correlations with each other.

Table 4: Used statistical tests according to the combination of the variables' levels of measurement, own presentation

	Dichotomous nominal	**Nominal**	**Ordinal**	**Metric**
Dichoto-mous nominal	Chi² Test / Fishers Test *Cramer's V*	Chi² Test / Fishers Test *Cramer's V*	Chi² Test / Fishers Test *Cramer's V*	(Point)biserial Correlation r_{ps} / r_{bs}
Nominal		Chi² Test / Fishers Test *Cramer's V*	Chi² Test / Fishers Test *Cramer's V*	ANOVA *Eta²*
Ordinal			Spearman's R *rho*	Spearman's R *rho*
Metric				Pearson's / Spearman's R *r / rho*

For correlations with **one (dichotomous and non-dichotomous) nominal variable and one (dichotomous and non-dichotomous) nominal or ordinal variable**, the Chi² test is used, whose calculations are based on a cross-tabulation (Field et al., 2013, p. 814). If the assumption of the Chi² test—that no more than 20% of the cells contain expected values of less than 5—is violated, Fisher's test is used. Fisher's test is specifically designed for smaller sample sizes (Field et al., 2013, p. 818). For both tests, the effect size *Cramer's V* is calculated. The values of *Cramer's V* lie between 0 and 1, while 1 indicates an almost perfect correlation (Gehrau et al., 2022, pp. 245–246). The *Cramer's V* is interpreted according to Cohen's approach. Thus, values between 0.1 and 0.3 indicate a low correlation, values between 0.3 and 0.5 indicate a medium correlation, and values above 0.5 indicate a strong correlation (Cohen, 1992).This categorisation also applies to the effect size of a calculation between **one dichotomous and one metric variable**. In that case the serial correlation is used. There you differ between the effect sizes point-biserial correlation coefficient (r_{pb}) and the biserial correlation coefficient (r_b). r_{pb} is used when one variable is a discrete dichotomy (for example pregnancy), whereas r_b is used when one variable is a continuous dichotomy (for example passing or failing an exam) (Field et al., 2013, p. 229). It has values between -1 and +1, with +-1 indicating the perfect correlation. A one-way analysis of variance (ANOVA) is used to analyse the correlation between **nominal (non-dichotomous) variables and metric variables**. The resulting effect measure is *Eta²* (Schendera, 2008, p. 6). *Eta²* can assume values between 0 and 1. The values are interpreted according to Cohen (1992) as well: Values from 0.10 to 0.15 are considered a small effect measure, values from 0.15 to 0.35 a medium effect measure and values above 0.35 a large effect measure (Cohen, 1992). In case of **two metric variables**, Pearson's correlation is the appropriate statistical tool. Pearson's correlation is employed for data that are normally distributed. In instances where this assumption is not met, the Spearman's correlation is utilised, in addition to its application for the correlation of two ordinal variables or one ordinal variable and one metric variable (Field et al., 2013, p. 223; Schendera, 2008, p. 6). The normality of the distribution is evaluated using the Shapiro-Wilk test, along with calculations of the mean, mode, and median. The effect size of the Pearson correlation is represented by the value of *r*, while the Spearman correlation is represented by *rho*. Both can assume values between -1 and +1, with 1 indicating a perfect positive (+) or negative (-) correlation. The interpretation of *r* and *rho* is also according to the approach proposed by Cohen (1992). Accordingly, correlations between 0.1 and 0.3 are indicative of a low level of correlation, correlations between 0.3 and 0.5 represent a medium level of correlation, and correlations above 0.5 are indicative of a strong level of correlation.

4.2.3 Multivariate logistic regression

In this thesis, a binary logistic regression with several predictors is applied. Firstly, the theory of binary logistic regression with its outcomes is described. The assumptions and quality criteria of the logistic regression model that are tested in this work are then presented. Possible data adjustments based on these are mentioned.

A logistic regression analyses the influence of one or more variables on a dichotomous outcome. The basic calculation behind this is that the logarithmic probability of the occurrence and non-occurrence of the outcome is determined for each of the predictors. The logarithmical probability is taken so that the assumption of a logistic regression that there is a linear relationship between predictor(s) and outcome can be fulfilled, in that the logarithmic values can take on values of infinity instead of 0 or 1 (Field et al., 2013, p. 315). The odds ratios are usually regarded as the value of the logistic regression to be interpreted. They are based on the odds, the calculation of which can be seen in the following simplified formula (Field et al., 2013, p. 320).

$$Odds = \frac{Logarithmised\ probability\ of\ occurrence\ of\ the\ outcome}{Logarithmised\ probability\ of\ non-occurrence\ of\ the\ outcome}$$

The odds ratios then result from the division of the odds after one unit change and the odds of the reference unit, as can be seen in the following simplified formula.

$$Odds\ ratios = \frac{Odds\ after\ one\ unit\ change\ in\ the\ predictor}{Odds\ of\ the\ reference\ unit}$$

The result of the odds ratios assumes values from 0 up. If it is below 1, it means that if the predictor adds a unit, the probability that the outcome of interest will occur decreases. A value of over 1 indicates the opposite, namely that if the predictor adds a unit, the probability that the outcome will occur increases (Field et al., 2013, p. 320). The odds ratios are generally given together with a 95%-CI, which indicates whether there is a significance or not. A significance is reached if the 95%-CI is not including the value 1. In this work, the odds ratio of rs12785878 is discussed to analyse the influence of a genetic variant.

A binary logistic regression includes several assumptions that the data set should fulfil. Some of them are addressed before conducting the analyses, some of them within the analyses.

Firstly, starting with the assumptions, which should be addressed before conducting any analysis, there is the need that the dependent variable is dichotomous binary (Schendera, 2008, p. 188). With the outcome's characteristics of 'vitamin D deficiency' and 'no vitamin D deficiency', it fits this criterion. Care should be taken to ensure that the non-occurrence of the analysed phenomenon is coded to 0 and the occurrence is coded to 1 (Schendera, 2008, p. 169). In the dataset used, this coding is adjusted so that 'no vitamin D deficiency' is coded as 0 and 'vitamin D deficiency' as 1. Moreover reference categories have to be defined for the independent nominal or ordinal variables depending on the change of interest (Field et al., 2013, p. 306). For the independent variable 'rs12785878', the major allele is defined as the reference to be able to represent the influence of a minor allele. For skin type, 'light skin type' is defined as the reference, as darker skin is associated with a higher risk of vitamin D deficiency (Bowyer et al., 2009; Brian-D Adinma et al., 2022; Christoph et al., 2020). Even though Krieger et al. (2018) found contrary, this thesis assumes in view of the literature described in subchapter 2.2.2 and the physiology described in chapter 2.2 and 2.2.1, that dark skin type is an influencing factor on vitamin D deficiency. Country of origin has the country group Germany/Switzerland as reference, as studies have shown that the risk of vitamin D deficiency is higher in non-European countries or ethnicities comprising people who are generally dark skinned (Christoph et al., 2020; Wuertz et al., 2013). The reference for supplementation with vitamin D preparations is the presence of supplementation, as this represents an advantage with regard to a possible vitamin D deficiency and a lack of supplementation represents a risk factor (Dovnik et al., 2017; Huang et al., 2023). The reference for the season of birth is summer, as the highest 25(OH)D_3 levels are measured at this time of year (Chen et al., 2024; Christoph et al., 2020).

Furthermore, since independent variables in a logistic regression need to be either metric or dichotomous nominal, dummy variables need to be created out of variables with more than two characteristics (here: Season of birth and country of origin). As part of the glm function in R, dichotomous dummy variables are automatically created with the reference category as one expression and the other ones combined to the other expression (Field et al., 2013, p. 306). The independent variables and the outcome should have a causal relationship with each other (Schendera, 2008, p. 167). The fact that these relationships exist for each of the predictors selected here has been demonstrated in chapter 2.2. With that literature research it is assumed that all relevant factors regarding vitamin D deficiency are considered. It is also important that a data point in each variable can be assigned to a specific participant (Schendera,

2008, p. 167). Each participant has been assigned their own ID, which means that this criterion has been met.

Coming to the four assumptions, which are addressed while doing the analyses: the variance, the necessary number of cases, multicollinearity and the presence of linearity. Within the univariate analysis it is examined, if the variables show a variance of zero, which should not be the case (Schendera, 2008, p. 170). Furthermore, a necessary number of cases must be achieved: At least 25 cases for each characteristic of the dependent variable should be available (Schendera, 2008, p. 170). Following that, in this case there should be 50 cases in total. It should also be ensured that the rarer of the two characteristics of the dependent variable occurs sufficiently frequently with respect to the number of predictors. This is calculated by dividing the frequency of the rarer expression by 10. The result is the maximum number of predictors that should be used in the model (Schendera, 2008, p. 170). Within the bivariate analysis multicollinearity among the independent variables is considered. All independent variables are tested for correlation with each other in order to rule out multicollinearity. A correlation of >0.70 is an indication of multicollinearity (Schendera, 2008, p. 169). In addition, the variance inflation factor (VIF) is calculated for all predictors (Field et al., 2013, p. 276). Values under 10 are considered as unproblematic (Field et al., 2013, p. 276; Schendera, 2008, p. 105). Lastly, the linearity needs to be checked. This is relevant for metric variables, since the assumption for logistic regression models says, that there has to be a linear correlation between those and the logarithm of the outcome. This inspection is done with the metric variable (current BMI) being multiplied with its logarithm. This calculation is saved in a new variable. With the new variable, the origin variable and the outcome a logistic regression model is conducted. If the new variable has a significant influence on the outcome, the assumption of linearity is not met (Schendera, 2008, p. 168).

With the named variables (rs12785878, current BMI, skin type, country of origin, season of birth and supplementation of vitamin D) the logistic model will be created with a hierarchical attempt. That means that the predictors are selected by the researcher based on the current state of research and they decide in which order the predictors are added to the model (Field et al., 2013, p. 264). The predictors may be inserted in the model in order of their importance in predicting the outcome. The resulting different models, variating with the predictors, are viewed with the goal of finding the best fitting model for the data. For that purpose, different approaches to assess the models' quality are used. Those are explained in the following.

Firstly, the Akaike information criterion (AIC) will be considered. A higher value of the AIC compared to the other models within the same data set means a poorer fitting of the model (Field et al., 2013, p. 263, 318). Furthermore, the Nagelkerke R^2 will be considered, which can be interpreted as the R^2 in linear regressions. This means it indicates how well the independent variables are suited to explain the outcome (Field et al., 2013, p. 318). The Nagelkerke R^2 can reach values between 0 and 1, with 1 being the best value (Schendera, 2008, p. 155). The regression model should be able to correctly reproduce most of the observed events using optimal estimates. In order to ascertain this, the Hosmer-Lemeshow test is initially employed. This test is a modified Pearson Chi-squared test based on expected frequencies distributed in ten equal groups. The test identifies whether there are differences between the expected and observed values. If these do not deviate significantly from each other, this speaks in favour of the quality of the model (Schendera, 2008, p. 170). In addition, for the purpose of comparing the expected and the observed values the 1) correct data overall in %, 2) the specify and 3) the sensitivity will be calculated. The expected values are being compared to the observed ones and with a cut-off value (defined as the median of the predicted values) the named calculations are done (Schendera, 2008, p. 173). Finally, the deviance statistic is calculated. In this approach, two models are considered: A baseline model with a single variable (the constant) and a model with new predictors. The deviance of the new model is subtracted from the deviance of the baseline model to obtain the likelihood ratio, which has a chi-square distribution. Therefore, this analysis allows for the determination of whether there is a significant difference in the models (Field et al., 2013, p. 316)

If the logistic regression includes more than 5 % of excluded cases in relation to the total sample size, the approach of multiple imputation chained equations (MICE) is performed using chained equations. Otherwise, the cases remain excluded and a complete case analysis is conducted (Azur et al., 2011).

5 Results

This chapter presents the results of the univariate and bivariate analyses as well as the logistic regression. The univariate analysis examines the general characteristics of the sample, including the selected predictors for the logistic regression. The bivariate analysis investigates the relationships between the predictors and the outcome 'vitamin D deficiency', as well as the relationships between the predictors themselves. Finally, the results of the selected logistic model are presented and categorised in accordance with the research question regarding the influence of the SNP rs12785878 on vitamin D deficiency.

5.1 Univariate analysis

The data determined in the univariate analysis is summarised in Table 5. First, the sociographic information of the sample is presented. Then, information describing (I) the pregnancy, (II) the vitamin D situation, and (III) the genetic status is discussed in more detail. If there are no missing values mentioned, the respective variable has none.

138 pregnant women met the criteria of the study. As mentioned in chapter 4.1 all of them had a c-section at the University Spital of Zurich. The mean age in the sample is 33.49 years (SD: 4.96). The youngest participant is 22 years old, the oldest 45 years old. The median age is 34. The majority (36.23%) of the women originate from Germany or Switzerland, followed by 'North America, Northern Europe, Caucasus, Central Asia, New Zealand' with 27.54%. There is one missing value for this variable. In relation to pregnancy and childbirth, the current BMI as well as the season of birth are reported. Most of the women gave birth in spring (43.75%), while the fewest gave birth in summer (7.25%). Every season is represented. There is one missing value (0.72%). The current BMI has a mean value of 28.83 (SD: 4.79). The values vary between 21.91 and 45.71 with a median of 27.81. Data is missing from two participants (1.44%). The outcome of interest of this thesis, vitamin D deficiency, shows that over half (52.17%) of the women giving birth are vitamin D deficient. 81% of the women reported taking vitamin D supplements, while the remaining 19% did not. Looking at the genetic information of DHCR7 and its SNP rs12785878, around 70% of the women have at least one minor allele. About 30% are homozygous for the major allele.

Table 5: Univariate analysis

	Population, n = 138
Sociodemographic	
Age (SD)	
Mean (SD)	33.49 years (4.67)
Median	34 years
Range	22 – 45 years
Origin country[1]	
Switzerland, Germany	50 women (36.23%)
North America, Northern Europe, Caucasus, Central Asia, New Zealand	38 women (27.54%)
South Europe, Australia, Latin America, Caribbean	20 women (14.49%)
South-East-Asia and Pacific	12 women (8.70%)
Africa and Middle East	17 women (12.32%)
NA	1 woman (0.72%)
Pregnancy and Birth	
Season of birth date[2]	
Spring	59 women (42.75%)
Summer	10 women (7.25%)
Fall	34 women (24.64%)
Winter	34 women (24.64%)
NA	1 woman (0.72%)
Current BMI	
Mean (SD)	28.83 (4.79)
Median	27.81
Range	21.91 - 45.71
NA	2 women (1.44%)
Vitamin D	
Vitamin D deficiency	
Yes (< 50 nmol/L 25(OH)D_3)	72 women (52.17%)
No (≥ 50 nmol/L 25(OH)D_3)	66 women (47.82%)
NA	NA
Skin type[3]	
Light	115 women (83.33%)
Dark	22 women (15.94%)
NA	1 woman (0.72%)
Vitamin D Supplementation	
Yes	112 women (81.16%)
No	26 women (18.84%)
NA	NA
Genetic Information	
Allele characterisation at rs12785878	
Major allele homozygote (G:G)	32 women (30.19%)
Minor allele homozygote (T :T) or heterozygote (G:T)	106 women (69.81%)
NA	NA

NA not available **SD** Standard deviation
[1] based on regional groupings of the World Bank Map (The World Bank, 2018) [2] is equivalent to the season of the blood collection [3] based on adaption of categorisation of Fitzpatrick (1988)

5.2 Bivariate analysis

The bivariate analysis and its results are presented in the following. First, the correlations of the independent variables with the outcome of vitamin D deficiency are considered. The results of the correlations among the predictors are then presented from a low correlation effect strength, according to the definitions of Cohen (1992).

Table 6 displays the effect sizes of several predictors related to vitamin D deficiency. The strongest correlations are found for country of origin (V = 0.28), SNP rs12785878 (V = 0.26), and current BMI (r_{bs}=0.25). The other variables, such as skin type, season of birth, and vitamin D supplementation, show very low associations with vitamin D deficiency, with effect sizes below 0.1. The interpretation follows Cohen's (1992) threshold, where values below 0.1 in r_{bs} and V are considered minimal.

Table 6: Bivariate analyses of predictors and vitamin D deficiency, own presentation

	Vitamin D deficiency
SNP rs12785878	V = 0.26
Skin type	V = 0.06
Current BMI	r_{bs} = 0.25
Country of origin	V = 0.28
Season of birth	V = 0.08
Supplementation of vitamin D	V = 0.09
light red low correlation ***white*** effect size below the low correlation threshold by Cohen (1992) $\boldsymbol{r_{bs}}$ biserial correlation ***V*** Cramer's V	

Table 7 presents the bivariate relationships between the predictors of the logistic regression. Country of origin shows a strong correlation with skin type (V = 0.63) and a moderate correlation with rs12785878 (V = 0.39). rs12785878 also has a moderate correlation with skin type (V = 0.23), while all other predictors, including BMI, season of birth, and vitamin D supplementation, show weak or minimal relationships.

Table 7: Bivariate analyses within the predictors, own presentation

	Country of origin	Skin type	Current BMI	Season	Supple-mentation
rs12785878	*V* = 0.39	*V* = 0.23	r_{bs} = - 0.18	*V* = 0.13	*V* = 0.13
Country of origin		*V* = 0.63	*No effect*[1]	*V* = 0.22	*V* = 0.10
Skin type			r_{bs} = 0.23	*V* = 0.11	*No effect*[1]
Current BMI				*No effect*[1]	*No effect*[1]
Season					*No effect*[1]
Supplementation					

light red low correlation ***medium red*** medium correlation
dark red strong correlation r_{bs} biserial correlation ***V*** Cramer's V
[1] Effect size below the low correlation thresholds by Cohen (1992)

In summary, the bivariate analyses show that certain predictors exhibit notable effect sizes in the correlation with vitamin D deficiency as well as among each other. Particularly noteworthy is the moderate correlation between the SNP rs12785878 and vitamin D deficiency, as well as the strong correlation between country of origin and skin type. These results emphasise that genetic factors such as the SNP rs12785878 and demographic factors such as country of origin play an important role in vitamin D status. Other variables, such as BMI, season of birth, and vitamin D supplementation, show weak or no associations with vitamin D deficiency and with each other. It is important to think about all effect sizes within the predictors, from small to large, in terms of the possibility of multicollinearity in the multiple logistic regression.

5.3 Multivariate logistic regression

Finally, the results of the logistic regression are presented. While parts of the assumptions have already been addressed in the methodology (formal requirements, defining reference categories, causal correlation of predictors and outcome), the assumptions considered in the analyses are described in this chapter. The final model selected is then presented with its results regarding the research question and the examination of its quality criteria.

The following assumptions were made and examined as part of the analyses:

- No variance of 0,
- a minimum of cases,
- no multicollinearity as well as
- presence of linearity.

Within the univariate analysis (see chapter 5.1), it becomes clear that no variable has a variance of 0, since there is a range of data values in the metric variable and different frequencies of the values within the nominal and ordinal variables. Furthermore, the sample consists of 138 people, thereby fulfilling the requirement of a minimum of 50 cases. The further assumption that the rarer of the two characteristics of the dependent variable needs to occur sufficiently frequently with respect to the number of predictors is calculated by dividing the frequency of the rarer expression by 10. This calculation yields the maximum number of predictors suitable for the model. So, in this case the rarer of the two characteristics of the dependent variable is a sufficient vitamin D status with 66 cases. That means that in the logistic regression a maximum of six predictor variables should be included. As mentioned in the bivariate analysis (chapter 6.2) there are some low and medium effect sizes between the predictors themselves. Particular attention should be paid to the correlations between the country of origin and skin type (V = 0.63) and the country of origin and rs12785878 (V = 0.39). The VIF is therefore calculated for all predictors to rule out if there may be multicollinearity. Every variable's VIF falls below 10, thereby rejecting the hypothesis of multicollinearity in the data set. Lastly the linearity between the metric variables and the outcome is calculated within a logistic regression. The only metric variable in the used data set is the current BMI. The current BMI does not occur significantly within the calculation. Following that, there is a linearity between the metric variable and the outcome.

As discussed in the methodology, the hierarchical approach initially resulted in several logistic models, each of which was analysed for their quality criteria. Step by step, the following variables were added to the simple model (rs12785878 and vitamin D deficiency) in the listed order: (1) skin type, (2) country of origin, (3) supplementation of vitamin D, (4) current BMI, and (5) season of birth. The overview of the models and the respective list of quality criteria is found in the Appendix. The final model was determined on the basis of the comparison of the quality criteria. The final model is now explained in more detail below.

In the final model, the influence of the SNP rs12785878 is adjusted for the variables current BMI, supplementation of vitamin D, country of origin, and skin type, so the season of birth is left out. An overview of the values determined within the model is found in Table 8. The model incorporates three missing values, which are not imputed, according to the beforehand defined cut-off value. Regarding the influence of the SNP rs12785878 in the DHCR7 gene on vitamin D deficiency, an odds ratio of 0.39 [95%-CI: 0.14 - 1.02, p-value 0.06] can be determined. The result is therefore not significant. For the sample itself, it can be stated that the probability of vitamin D deficiency decreases with the presence of at least one minor allele. The model has the best AIC value (184.48) in comparison to the AIC values of every other model (>188). In the deviance statistic, calculations (1) vs. basis model are significant, which indicates that there is a difference between the models. The comparison with the model is not significant. In the Hosmer-Lemeshow test, the calculation did not occur significantly, so there is no difference in the prediction and the observations according to this test. Looking at the comparisons of expected and observed values, the model gets 51.11% of all values right. The model correctly classifies the values of participants without a vitamin D deficiency with a specificity of 36.36%. Furthermore, it has a sensitivity of 65.22%, meaning that this percentage of all participants with a vitamin D deficiency are correctly classified. The model has a Nagelkerke R^2 value of 0.189, so 18.9% of the variance of vitamin D deficiency can be explained with those predictors.

Table 8: Odds ratios and quality criteria of the final model (adjusted for current BMI, supplementation of vitamin D, country of origin and skin type), own presentation

	Values final model
Missing values	3
OR (95%-CI, p-value) of rs12785878	0.39 (0.14 – 1.02, 0.06)
AIC	184.48
Hosmer-Lemeshow test	Not significant
Correct data in %	51.11%
Specificity in %	36.36%
Sensitivity in %	65.22%
Nagelkerke R^2	0.189

The logistic regression shows that although the influence of the SNP rs12785878 on vitamin D deficiency is not significant, there is a tendency to reduce the probability of deficiency in the presence of a T allele. However, the final model shows moderate

prediction accuracy. The sensitivity is higher than the specificity, and the explainable variance remains limited.

6 Discussion

The aim of this thesis was to determine the effect of the SNP rs12785878 on vitamin D deficiency in Swiss pregnant women. The results of this work are first summarised (chapter 6.1) and then discussed (chapter 6.2). Chapter 6.3 deals with the limitations of this work.

6.1 Key results

The research question "*How does the SNP rs12785878 in the gene DHCR7 influences vitamin D deficiency in pregnant women shortly before delivery*?" is answered in this thesis using a logistic regression with multiple predictors. The final regression model was adjusted for the variables current BMI, supplementation of vitamin D, country of origin and skin type, which were determined from literature. Data from 138 pregnant women who had a c-section at Zurich University Hospital in 2014-2016 are analysed. 52.17% of those pregnant women have a vitamin D deficiency. Approximately one third of the women is homozygous for the major allele (G:G) in rs12785878, while two thirds have a T:T or G:T expression. The logistic regression shows that having at least one T allele in rs12785878 lowers the chance of not having enough vitamin D [OR: 0.39]. However, the result is not significant with a 95%-CI of 0.14 - 1.02. This model explains 18.9% of the variance in vitamin D deficiency.

6.2 Interpretation of the results

The prevalence of vitamin D deficiency (52.17%) found in this thesis regarding pregnant women in the third trimester seems low compared to similar studies in Germany. Gellert et al. (2017) and Wuertz et al. (2013) show prevalence rates of 78.1% and 77.0%. Only Krieger et al. (2018) show a similar prevalence of 53.4%. However, it should be noted that the data used in this thesis comes from a subsample of the participants from Krieger et al., so the similarities in prevalence are not unexpected. To the best of the author's knowledge, the most recent study in Switzerland on vitamin D deficiency in pregnant women, conducted by Christoph et al. (2020) and focusing on the first and second trimesters, revealed that 73.23% of 1,382 women at the University Hospital Bern were vitamin D deficient. This is well aligned with the other reported prevalences from Germany. Based on systematic reviews and meta-analyses from Europe, this thesis's prevalence falls within a well-established range. For

example, Palacios and Gonzales (2014) showed that vitamin D deficiency in pregnant and breastfeeding women ranges from 20% (Spain), to 31/35% (England/United Kingdom), to 44/45% (Netherlands, Belgium) and up to 90% in Turkey. Karras et al. (2016) showed a similar picture when looking at vitamin D deficiency in the Mediterranean region (including studies from Greece, Turkey, Spain, Italy and Tunisia) with a range of prevalences from 22.7 to 90.3%. Considering the potential consequences of vitamin D deficiency, it is crucial to raise awareness about these reported prevalences, as all women who are deficient may be susceptible to the reported consequences.

Regarding rs12785878, the allele frequencies found in this work are also in line with the existing literature. G:G is the less common expression in pregnant women in the UK (Moon et al., 2017), Norway (Størdal et al., 2017) as well as in China (Shao et al., 2018). The relative frequency of G:G is between 7.1% and 27.6%, so that the 30.19% of this thesis' results fits the picture. However, in general, the regional peculiarities must be taken into account, as the G allele is found more frequently in other regions, such as Africa and South Asia (Harrison et al., 2024a). Nevertheless, there is consistency regarding the effect of a T-allele in rs12785878. It leads to a lower probability of vitamin D deficiency shown in this work (without significancy), which can also be found in the international research about women in their third trimester by Aji et al. (2020) and Moon et al. (2017). Shao et al. (2020) showed it for women without vitamin D supplementation in their first trimester. This is also consistent with the majority of findings in relation to the general population (Kuan et al., 2013; Kühn et al., 2014; Zhang et al., 2012). It should be noted that other studies do not define what was set as the major or minor allele in the analyses, making it difficult to interpret the results, as a major/minor allele can refer to the sample itself, the region or globally, depending on the individual interpretation (Harrison et al., 2024a; Schaaf and Zschocke, 2013, p. 32). Further, there are different thresholds referred to as vitamin D deficiency, such as Shao et. al (2018), where vitamin D deficiency was defined as a value of < 36 nmol/L or in a German study of Gellert et al. (2017), where a threshold of < 25.0 nmol/L is used. This confirms the fact that there is yet no uniformly defined value that is assessed as vitamin D deficiency (Cashman, 2020), which makes it difficult to achieve clear results.

There is 18.9% of variation in vitamin D deficiency that can be explained by the final model of this thesis, as shown by the Nagelkerke R^2. According to Zittermann (2022, p. 25), the totality of genes that influence vitamin D levels accounts for approximately 5% of the total variation. A Nagelkerke R^2 of 0.063 was found for the univariate logistic model for rs12785878 and its link to vitamin D deficiency in this study. However, it is

important to note that, as further explained in the limitations, there may be missing variables in the model, including interaction variables of rs12785878, which could lead to a larger univariate effect than what is actually observed.

6.3 Limitations

The limitations of this thesis arise in several areas. Firstly, the transferability of the results to other regions or to the entire group of pregnant women at Zurich University Hospital is only possible to a limited extent. The sample was neither randomised nor representative, as only women at Zurich University Hospital were examined. As Krieger et al. (2018) show, regional differences within Switzerland can already have a significant influence on vitamin D status: $25(OH)D_3$ levels in Bellinzona were significantly higher than in Zurich and Samedan. Krieger et al. attribute these differences to the duration of sunshine and the outside temperature, which significantly influence the UVB exposure of the skin in the different regions. The regional diversity of Switzerland makes it clear that participants from different regions should be considered and investigated in further studies in order to be able to make a statement for the entire Swiss population.

In terms of model quality, the statistical model has some limitations that influence the interpretation of the results. Although the model is the best model compared to the others with respect to the AIC value, there are limitations in other quality criteria. The Hosmer-Lemeshow test did not find any significant differences between the predicted and observed values. This means that there are no statistically significant differences in how well the model fits the predictions to the actual results. Nevertheless, only 51.11% of the observed values were predicted correctly, which indicates a limited prediction accuracy. The specificity of the model is 36.36 %, which means that only approx. one third of the participants without vitamin D deficiency were correctly classified as unaffected. This contrasts with a sensitivity of 65.22%, which shows that the model was able to correctly identify vitamin D deficiency in approximately two thirds of affected participants. These values indicate a disparity in model performance, as the model performs better in recognising cases of vitamin D deficiency but is less effective in accurately classifying unaffected participants. A reason for that may be, that the Nagelkerke R^2 value of 0.189 suggests that the model's predictors can only explain 18.9% of the variance in vitamin D deficiency. This suggests that factors not included in the model influence a significant portion of the variance. The bivariate analysis provided an initial indication of this, as three of the six selected predictors showed a minimal effect size with the outcome. This suggests that other factors may

moderate or lessen the influence of these variables on vitamin D status in this sample. Despite a thorough literature search, potentially relevant variables influencing vitamin D deficiency may not be included. Some of these influencing factors may not have been yet specifically investigated for pregnant women, although their importance in the general population, such as age, clothing, diet and time in the sun, has already been demonstrated (Zittermann, 2022). Another reason for that may be, that while the bivariate analysis showed significant correlations between skin type and country of origin, the interaction of these variables was not considered in the final model. Taking such interactions into account could have improved the variance explanation of the model. Furthermore, the potential impact of cholesterol metabolism, including the consideration of cholesterol levels, was not addressed due to the absence of a biomarker variable in the data set, despite the involvement of DHCR7 in this metabolic process. It should be noted that the original study was not focused on rs12785878, rather, it collected several SNP from multiple genes relevant to vitamin D metabolism. Consequently, further studies investigating the role of rs12785878 in vitamin D deficiency may provide additional insights into the potential involvement of interactions between cholesterol metabolism during pregnancy and vitamin D metabolism. Overall, the model provides valuable insights but leaves room for further investigation to clarify the exact relationships.

7 Conclusion

This thesis suggests that there may be a link between the T allele in rs12785878 and a lower risk of vitamin D deficiency in pregnant women in Switzerland. Although the result lacks statistical significance, it offers valuable insights suggesting that genetic factors, particularly rs12785878, may play a role in the target group. Further research, focussing on larger, representative cohorts or studies specifically targeting this variant, is needed. Notably, the exploration of genetic influences on vitamin D deficiency in pregnant women is a relatively young field, with a growing number of publications emerging in the last decade. Future studies examining the association of rs12785878 and vitamin D status should consider additional variables such as the influence of cholesterol metabolism, including cholesterol values, to improve model quality. To obtain representative results for Swiss pregnant women, particular attention should be given to addressing regional differences, as these have been shown to result in different vitamin D levels. Generally, the different definitions of vitamin D deficiency in the literature make direct comparisons of study results difficult. Therefore, standardised threshold values for the analysis of vitamin D deficiency are recommended to further improve the comparability of studies and thus ensure a well-founded investigation of influencing factors.

References

Abu El Maaty, M. A., Hassanein, S. I., Sleem, H. M. and Gad, M. Z. (2013) 'Effect of polymorphisms in the NADSYN1/DHCR7 locus (rs12785878 and rs1790349) on plasma 25-hydroxyvitamin D levels and coronary artery disease incidence', *Journal of nutrigenetics and nutrigenomics*, vol. 6, no. 6, pp. 327–335 [Online]. DOI: 10.1159/000360422.

Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K. C., McCullough, M. L., Gallicchio, L., Jacobs, E. J., Ascherio, A., Helzlsouer, K., Jacobs, K. B., Li, Q., Weinstein, S. J., Purdue, M., Virtamo, J., Horst, R., Wheeler, W., Chanock, S., Hunter, D. J., Hayes, R. B., Kraft, P. and Albanes, D. (2010) 'Genome-wide association study of circulating vitamin D levels', *Human molecular genetics*, vol. 19, no. 13, pp. 2739–2745 [Online]. DOI: 10.1093/hmg/ddq155.

Aji, A. S., Yusrawati, Y., Malik, S. G. and Lipoeto, N. I. (2020) 'The Association between Vitamin D-Related Gene Polymorphisms and Serum 25-Hydroxyvitamin D Concentration: A Prospective Cohort Study in Pregnant Minangkabau Women, Indonesia', *Journal of nutritional science and vitaminology*, vol. 66, Supplement, pp. 295–303 [Online]. DOI: 10.3177/jnsv.66.S295.

Alzohily, B., AlMenhali, A., Gariballa, S., Munawar, N., Yasin, J. and Shah, I. (2024) 'Unraveling the complex interplay between obesity and vitamin D metabolism', *Scientific reports*, vol. 14, no. 1 [Online]. DOI: 10.1038/s41598-024-58154-z.

Arshad, R., Sameen, A., Murtaza, M. A., Sharif, H. R., Iahtisham-Ul-Haq, Dawood, S., Ahmed, Z., Nemat, A. and Manzoor, M. F. (2022) 'Impact of vitamin D on maternal and fetal health: A review', *Food science & nutrition*, vol. 10, no. 10, pp. 3230–3240 [Online]. DOI: 10.1002/fsn3.2948.

Azur, M. J., Stuart, E. A., Frangakis, C. and Leaf, P. J. (2011) 'Multiple imputation by chained equations: what is it and how does it work?', *International Journal of Methods in Psychiatric Research*, vol. 20, no. 1, pp. 40–49 [Online]. DOI: 10.1002/mpr.329.

Bischoff-Ferrari, H. A., Keller, U., Burckhardt, P., Quack Lötscher, K. C., Gerber, B., l'Allemand, D., Laimbacher, J. and Rizzoli, R. (2012) 'Empfehlungen der Eidgenössischen Ernährungskommission zur Vitamin-D-Zufuhr für die Schweizer Bevölkerung', pp. 775–778 [Online]. DOI: 10.5167/UZH-73000.

Bowyer, L., Catling-Paull, C., Diamond, T., Homer, C., Davis, G. and Craig, M. E. (2009) 'Vitamin D, PTH and calcium levels in pregnant women and their

neonates', *Clinical endocrinology*, vol. 70, no. 3, pp. 372–377 [Online]. DOI: 10.1111/j.1365-2265.2008.03316.x.

Brian-D Adinma, J. I., Ahaneku, J. E., Adinma, E. D., Ugboaja, J. O., Okolie, V., Adinma-Obiajulu, N. D. and Edet, M. M. (2022) 'Vitamin D and associated factors, among pregnant women in southeastern Nigeria', *Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology*, vol. 42, no. 4, pp. 580–586 [Online]. DOI: 10.1080/01443615.2021.1931068.

Brouwer-Brolsma, E. M., Vaes, A. M. M., van der Zwaluw, N. L., van Wijngaarden, J. P., Swart, K. M. A., Ham, A. C., van Dijk, S. C., Enneman, A. W., Sohl, E., van Schoor, N. M., van der Velde, N., Uitterlinden, A. G., Lips, P., Feskens, E. J. M., Dhonukshe-Rutten, R. A. M. and Groot, L. C. P. G. M. de (2016) 'Relative importance of summer sun exposure, vitamin D intake, and genes to vitamin D status in Dutch older adults: The B-PROOF study', *The Journal of steroid biochemistry and molecular biology*, vol. 164, pp. 168–176 [Online]. DOI: 10.1016/j.jsbmb.2015.08.008.

Bu, F.-X., Armas, L., Lappe, J., Zhou, Y., Gao, G., Wang, H.-W., Recker, R. and Zhao, L.-J. (2010) 'Comprehensive association analysis of nine candidate genes with serum 25-hydroxy vitamin D levels among healthy Caucasian subjects', *Human genetics*, vol. 128, no. 5, pp. 549–556 [Online]. DOI: 10.1007/s00439-010-0881-9.

Bundesamt für Statistik (2024) *Geburten* [Online]. Available at https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/geburten-todesfaelle/geburten.html (Accessed 15 September 2024).

Cabaset, S., Krieger, J.-P., Richard, A., Elgizouli, M., Nieters, A., Rohrmann, S. and Quack Lötscher, K. C. (2019) 'Vitamin D status and its determinants in healthy pregnant women living in Switzerland in the first trimester of pregnancy', *BMC pregnancy and childbirth*, vol. 19, no. 1 [Online]. DOI: 10.1186/s12884-018-2150-1.

Cashman, K. D. (2020) 'Vitamin D Deficiency: Defining, Prevalence, Causes, and Strategies of Addressing', *Calcified tissue international*, vol. 106, no. 1, pp. 14–29 [Online]. DOI: 10.1007/s00223-019-00559-4.

Chen, B., Ji, P., Wang, Q., Qin, W. and Li, Z. (2024) 'Vitamin D levels and its influencing factors in pregnant women in mainland China: A systematic review and meta-analysis', *PLoS ONE*, vol. 19, no. 5 [Online]. DOI: 10.1371/journal.pone.0297613.

Christoph, P., Challande, P., Raio, L. and Surbek, D. (2020) 'High prevalence of severe vitamin D deficiency during the first trimester in pregnant women in Switzerland and its potential contributions to adverse outcomes in the pregnancy', *Swiss medical weekly*, vol. 150 [Online]. DOI: 10.4414/smw.2020.20238.

Cohen, J. (1992) 'A power primer', *Psychological bulletin*, vol. 112, no. 1, pp. 155–159 [Online]. DOI: 10.1037//0033-2909.112.1.155.

Dastani, Z., Li, R. and Richards, B. (2013) 'Genetic regulation of vitamin D levels', *Calcified tissue international*, vol. 92, no. 2, pp. 106–117 [Online]. DOI: 10.1007/s00223-012-9660-z.

Dawson-Hughes, B., Mithal, A., Bonjour, J.-P., Boonen, S., Burckhardt, P., Fuleihan, G. E.-H., Josse, R. G., Lips, P., Morales-Torres, J. and Yoshimura, N. (2010) 'IOF position statement: vitamin D recommendations for older adults', *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, vol. 21, no. 7, pp. 1151–1154 [Online]. DOI: 10.1007/s00198-010-1285-3.

Demay, M. B., Pittas, A. G., Bikle, D. D., Diab, D. L., Kiely, M. E., Lazaretti-Castro, M., Lips, P., Mitchell, D. M., Murad, M. H., Powers, S., Rao, S. D., Scragg, R., Tayek, J. A., Valent, A. M., Walsh, J. M. E. and McCartney, C. R. (2024) 'Vitamin D for the Prevention of Disease: An Endocrine Society Clinical Practice Guideline', *The Journal of clinical endocrinology and metabolism*, vol. 109, no. 8, pp. 1907–1947 [Online]. DOI: 10.1210/clinem/dgae290.

Dovnik, A., Mujezinović, F., Treiber, M., Pečovnik Balon, B., Gorenjak, M., Maver, U. and Takač, I. (2017) 'Determinants of maternal vitamin D concentrations in Slovenia : A prospective observational study', *Wiener klinische Wochenschrift*, no. 129, pp. 21–28 [Online]. DOI: 10.1007/s00508-016-1142-2.

Elm, E. von, Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. and Vandenbroucke, J. P. (2008) 'The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies', *Journal of clinical epidemiology*, vol. 61, no. 4, pp. 344–349 [Online]. DOI: 10.1016/j.jclinepi.2007.11.008.

Field, A., Miles, J. and Field, Z. (2013) *Discovering statistics using R*, Los Angeles, Sage.

Fitzpatrick, T. B. (1988) 'The Validity and Practicality of Sun-Reactive Skin Types I Through VI', *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871 [Online]. DOI: 10.1001/archderm.1988.01670060015008.

Föhl-Kuse, S. (2019) 'Präeklampsie und HELLP-Syndrom', *Ernährung & Medizin*, vol. 34, no. 01, pp. 33–36 [Online]. DOI: 10.1055/a-0831-3585.

Gehrau, V., Maubach, K. and Fujarski, S. (2022) *Einfache Datenauswertung mit R*, Wiesbaden, Springer Fachmedien Wiesbaden.

Gellert, S., Ströhle, A., Bitterlich, N. and Hahn, A. (2017) 'Higher prevalence of vitamin D deficiency in German pregnant women compared to non-pregnant women', *Archives of gynecology and obstetrics*, vol. 296, no. 1, pp. 43–51 [Online]. DOI: 10.1007/s00404-017-4398-5.

Ghio, A., Bertolotto, A., Resi, V., Volpe, L. and Di Cianni, G. (2011) 'Triglyceride metabolism in pregnancy', *Advances in clinical chemistry*, vol. 55, pp. 133–153 [Online]. DOI: 10.1016/B978-0-12-387042-1.00007-1.

Hanley, D. A., Cranney, A., Jones, G., Whiting, S. J., Leslie, W. D., Cole, D. E. C., Atkinson, S. A., Josse, R. G., Feldman, S., Kline, G. A. and Rosen, C. (2010) 'Vitamin D in adult health and disease: a review and guideline statement from Osteoporosis Canada', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, vol. 182, no. 12, 610 - 618 [Online]. DOI: 10.1503/cmaj.080663.

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K., Cockburn, A., Davidson, C., Silva, N. H. de, Dodiya, K., Donaldson, S., El Houdaigui, B., Naboulsi, T. E., Fatima, R., Giron, C. G., Genez, T., Grigoriadis, D., Ghattaoraya, G. S., Martinez, J. G., Gurbich, T. A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Kay, M., Kaykala, V., Le, T., Lemos, D., Lodha, D., Marques-Coelho, D., Maslen, G., Merino, G. A., Mirabueno, L. P., Mushtaq, A., Hossain, S. N., Ogeh, D. N., Sakthivel, M. P., Parker, A., Perry, M., Piližota, I., Poppleton, D., Prosovetskaia, I., Raj, S., Pérez-Silva, J. G., Salam, A. I. A., Saraf, S., Saraiva-Agostinho, N., Sheppard, D., Sinha, S., Sipos, B., Sitnik, V., Stark, W., Steed, E., Suner, M.-M., Surapaneni, L., Sutinen, K., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, T. A., Ware, D., Wass, E., Willhoft, N. L., Allen, J., Alvarez-Jarreta, J., Chakiachvili, M., Flint, B., Giorgetti, S., Haggerty, L., Ilsley, G. R., Keatley, J., Loveland, J. E., Moore, B., Mudge, J. M., Naamati, G., Tate, J., Trevanion, S. J., Winterbottom, A., Frankish,

A., Hunt, S. E., Cunningham, F., Dyer, S., Finn, R. D., Martin, F. J. and Yates, A. D. (2024a) 'Ensembl 2024', *Nucleic acids research*, vol. 52, D1, 891-899 [Online]. DOI: 10.1093/nar/gkad1049.

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K., Cockburn, A., Davidson, C., Silva, N. H. de, Dodiya, K., Donaldson, S., El Houdaigui, B., Naboulsi, T. E., Fatima, R., Giron, C. G., Genez, T., Grigoriadis, D., Ghattaoraya, G. S., Martinez, J. G., Gurbich, T. A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Kay, M., Kaykala, V., Le, T., Lemos, D., Lodha, D., Marques-Coelho, D., Maslen, G., Merino, G. A., Mira-bueno, L. P., Mushtaq, A., Hossain, S. N., Ogeh, D. N., Sakthivel, M. P., Parker, A., Perry, M., Piližota, I., Poppleton, D., Prosovetskaia, I., Raj, S., Pérez-Silva, J. G., Salam, A. I. A., Saraf, S., Saraiva-Agostinho, N., Sheppard, D., Sinha, S., Sipos, B., Sitnik, V., Stark, W., Steed, E., Suner, M.-M., Surapaneni, L., Sutinen, K., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, T. A., Ware, D., Wass, E., Willhoft, N. L., Allen, J., Alvarez-Jarreta, J., Chakiachvili, M., Flint, B., Giorgetti, S., Haggerty, L., Ilsley, G. R., Keatley, J., Loveland, J. E., Moore, B., Mudge, J. M., Naamati, G., Tate, J., Trevanion, S. J., Winterbottom, A., Frankish, A., Hunt, S. E., Cunningham, F., Dyer, S., Finn, R. D., Martin, F. J. and Yates, A. D. (2024b) *Ensembl 2024 - rs12785878* [Online]. Available at https://www.ensembl.org/Homo_sapiens/Variation/Explore?r=11:71455903-71456903;v=rs12785878;vdb=variation;vf=706853291 (Accessed 23 September 2024).

Holick, M. F., Binkley, N. C., Bischoff-Ferrari, H. A., Gordon, C. M., Hanley, D. A., Heaney, R. P., Murad, M. H. and Weaver, C. M. (2011) 'Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline', *The Journal of clinical endocrinology and metabolism*, vol. 96, no. 7, pp. 1911–1930 [Online]. DOI: 10.1210/jc.2011-0385.

Huang, Y.-L., Pham, T. T. M., Chen, Y.-C., Chang, J.-S., Chao, J. C.-J. and Bai, C.-H. (2023) 'Effects of Climate, Sun Exposure, and Dietary Intake on Vitamin D Concentrations in Pregnant Women: A Population-Based Study', *Nutrients*, vol. 15, no. 5 [Online]. DOI: 10.3390/nu15051182.

Kanaka, K. K., Sukhija, N., Goli, R. C., Singh, S., Ganguly, I., Dixit, S. P., Dash, A. and Malik, A. A. (2023) 'On the concepts and measures of diversity in the genomics era', *Current Plant Biology*, vol. 33 [Online]. DOI: 10.1016/j.cpb.2023.100278.

Karras, S., Paschou, S. A., Kandaraki, E., Anagnostis, P., Annweiler, C., Tarlatzis, B. C., Hollis, B. W., Grant, W. B. and Goulis, D. G. (2016) 'Hypovitaminosis D in pregnancy in the Mediterranean region: a systematic review', *European journal of clinical nutrition*, vol. 70, no. 9, pp. 979–986 [Online]. DOI: 10.1038/ejcn.2016.12.

Krieger, J.-P., Cabaset, S., Canonica, C., Christoffel, L., Richard, A., Schröder, T., Wattenwyl, B. L. von, Rohrmann, S. and Lötscher, K. Q. (2018) 'Prevalence and determinants of vitamin D deficiency in the third trimester of pregnancy: a multi-centre study in Switzerland', *The British journal of nutrition*, vol. 119, no. 3, pp. 299–309 [Online]. DOI: 10.1017/S0007114517003634.

Kuan, V., Martineau, A. R., Griffiths, C. J., Hyppönen, E. and Walton, R. (2013) 'DHCR7 mutations linked to higher vitamin D status allowed early human migration to northern latitudes', *BMC evolutionary biology*, vol. 13 [Online]. DOI: 10.1186/1471-2148-13-144.

Kühn, T., Kaaks, R., Teucher, B., Hirche, F., Dierkes, J., Weikert, C., Katzke, V., Boeing, H., Stangl, G. I. and Buijsse, B. (2014) 'Dietary, lifestyle, and genetic determinants of vitamin D status: a cross-sectional analysis from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Germany study', *European journal of nutrition*, vol. 53, no. 3, pp. 731–741 [Online]. DOI: 10.1007/s00394-013-0577-8.

McGrath, J. J., Saha, S., Burne, T. H. J. and Eyles, D. W. (2010) 'A systematic review of the association between common single nucleotide polymorphisms and 25-hydroxyvitamin D concentrations', *The Journal of steroid biochemistry and molecular biology*, vol. 121, 1-2, pp. 471–477 [Online]. DOI: 10.1016/j.jsbmb.2010.03.073.

Moon, R. J., Harvey, N. C., Cooper, C., D'Angelo, S., Curtis, E. M., Crozier, S. R., Barton, S. J., Robinson, S. M., Godfrey, K. M., Graham, N. J., Holloway, J. W., Bishop, N. J., Kennedy, S., Papageorghiou, A. T., Schoenmakers, I., Fraser, R., Gandhi, S. V., Prentice, A., Inskip, H. M. and Javaid, M. K. (2017) 'Response to Antenatal Cholecalciferol Supplementation Is Associated With Common Vitamin D-Related Genetic Variants', *The Journal of clinical endocrinology and metabolism*, vol. 102, no. 8, pp. 2941–2949 [Online]. DOI: 10.1210/jc.2017-00682.

Palacios, C. and Gonzalez, L. (2014) 'Is vitamin D deficiency a major global public health problem?', *The Journal of steroid biochemistry and molecular biology*, 144 Pt A, pp. 138–145 [Online]. DOI: 10.1016/j.jsbmb.2013.11.003.

Pereira-Santos, M., Costa, P. R. F., Assis, A. M. O., Santos, C. A. S. T. and Santos, D. B. (2015) 'Obesity and vitamin D deficiency: a systematic review and meta-analysis', *Obesity reviews : an official journal of the International Association for the Study of Obesity*, vol. 16, no. 4, pp. 341–349 [Online]. DOI: 10.1111/obr.12239.

Pilz, S., Zittermann, A., Obeid, R., Hahn, A., Pludowski, P., Trummer, C., Lerchbaum, E., Pérez-López, F. R., Karras, S. N. and März, W. (2018) 'The Role of Vitamin D in Fertility and during Pregnancy and Lactation: A Review of Clinical Data', *International journal of environmental research and public health*, vol. 15, no. 10 [Online]. DOI: 10.3390/ijerph15102241.

Posit team (2024) *RStudio: Integrated Development Environment for R* [Online], Boston. Available at http://www.posit.co/ (Accessed 1 September 2024).

Prabhu, A. V., Luu, W., Li, D., Sharpe, L. J. and Brown, A. J. (2016) 'DHCR7: A vital enzyme switch between cholesterol and vitamin D production', *Progress in Lipid Research*, vol. 64, pp. 138–151 [Online]. DOI: 10.1016/j.plipres.2016.09.003.

Qiu, X., Chen, X., Zuo, S., Ji, Y., Wen, Z., Wei, L., Wu, S., Le Diao, Li, B., Zhao, J. and Chen, T. (2020) 'Assessing vitamin D related genetic variants, status, and influence factors in pregnant women in Eastern and Central China', *Food science & nutrition*, vol. 8, no. 8, pp. 4078–4085 [Online]. DOI: 10.1002/fsn3.1674.

R Core Team (2024) *R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing* [Online], Vienna, Austria. Available at https://www.r-project.org/ (Accessed 1 September 2024).

Rebelos, E., Tentolouris, N. and Jude, E. (2023) 'The Role of Vitamin D in Health and Disease: A Narrative Review on the Mechanisms Linking Vitamin D with Disease and the Effects of Supplementation', *Drugs*, vol. 83, no. 8, pp. 665–685 [Online]. DOI: 10.1007/s40265-023-01875-8.

Schaaf, C. P. and Zschocke, J. (2013) *Basiswissen Humangenetik*, Berlin, Heidelberg, Springer Berlin Heidelberg.

Schendera, C. (2008) *Regressionsanalyse mit SPSS*, München, Oldenbourg Wissenschaftsverlag GmbH.

Schlaudecker, E. P., Munoz, F. M., Bardají, A., Boghossian, N. S., Khalil, A., Mousa, H., Nesin, M., Nisar, M. I., Pool, V., Spiegel, H. M. L., Tapia, M. D., Kochhar, S. and Black, S. (2017) 'Small for gestational age: Case definition & guidelines for data collection, analysis, and presentation of maternal immunisation safety data',

Vaccine, vol. 35, no. 48, pp. 6518–6528 [Online]. DOI: 10.1016/j.vaccine.2017.01.040.

Schweizerische Gesellschaft für Ernährung (2019) *Schweizer Referenzwerte - Schweizerische Gesellschaft für Ernährung* [Online]. Available at https://www.sge-ssn.ch/grundlagen/lebensmittel-und-naehrstoffe/naehrstoffempfehlungen/empfehlungen-blv/ (Accessed 1 September 2024).

Schweizerischer Nationalfond (2014) *Evaluation of Vitamin D Status and Its Determinants in Switzerland and of Possibilities to Improve Vitamin D Status in the Swiss Population* [Online]. Available at https://data.snf.ch/grants/grant/145194 (Accessed 1 September 2024).

Shao, B., Jiang, S., Muyiduli, X., Wang, S., Mo, M., Li, M., Wang, Z. and Yu, Y. (2018) 'Vitamin D pathway gene polymorphisms influenced vitamin D level among pregnant women', *Clinical nutrition (Edinburgh, Scotland)*, vol. 37, 6 Pt A, pp. 2230–2237 [Online]. DOI: 10.1016/j.clnu.2017.10.024.

Størdal, K., Mårild, K., Tapia, G., Haugen, M., Cohen, A. S., Lie, B. A. and Stene, L. C. (2017) 'Fetal and Maternal Genetic Variants Influencing Neonatal Vitamin D Status', *The Journal of clinical endocrinology and metabolism*, vol. 102, no. 11, pp. 4072–4079 [Online]. DOI: 10.1210/jc.2017-00827.

Ströhle, A. and Hahn, A. (2018) *Vitamin D in der Schwangerschaft - Ein zweischneidiges Schwert?: Reihe "Ernährung im Fokus" des Bundeszentrum für Ernährung* [Online]. Available at https://www.bzfe.de/fileadmin/resources/import/pdf/eifonline_vitamin_d_in_der_schwangerschaft_final.pdf (Accessed 12 September 2024).

The World Bank (2018) *World by region map* [Online]. Available at https://datatopics.worldbank.org/world-development-indicators/images/figures-png/world-by-region-map.pdf (Accessed 12 September 2024).

Wang, T. J., Zhang, F., Richards, J. B., Kestenbaum, B., van Meurs, J. B., Berry, D., Kiel, D. P., Streeten, E. A., Ohlsson, C., Koller, D. L., Peltonen, L., Cooper, J. D., O'Reilly, P. F., Houston, D. K., Glazer, N. L., Vandenput, L., Peacock, M., Shi, J., Rivadeneira, F., McCarthy, M. I., Anneli, P., Boer, I. H. de, Mangino, M., Kato, B., Smyth, D. J., Booth, S. L., Jacques, P. F., Burke, G. L., Goodarzi, M., Cheung, C.-L., Wolf, M., Rice, K., Goltzman, D., Hidiroglou, N., Ladouceur, M., Wareham, N. J., Hocking, L. J., Hart, D., Arden, N. K., Cooper, C., Malik, S., Fraser, W. D., Hartikainen, A.-L., Zhai, G., Macdonald, H. M., Forouhi, N. G., Loos, R. J. F., Reid, D. M., Hakim, A., Dennison, E., Liu, Y., Power, C., Stevens, H. E., Jaana, L., Vasan, R. S., Soranzo, N., Bojunga, J., Psaty, B. M., Lorentzon, M.,

Foroud, T., Harris, T. B., Hofman, A., Jansson, J.-O., Cauley, J. A., Uitterlinden, A. G., Gibson, Q., Järvelin, M.-R., Karasik, D., Siscovick, D. S., Econs, M. J., Kritchevsky, S. B., Florez, J. C., Todd, J. A., Dupuis, J., Hyppönen, E. and Spector, T. D. (2010) 'Common genetic determinants of vitamin D insufficiency: a genome-wide association study', *Lancet (London, England)*, vol. 376, no. 9736, pp. 180–188 [Online]. DOI: 10.1016/S0140-6736(10)60588-0.

Wiznitzer, A., Mayer, A., Novack, V., Sheiner, E., Gilutz, H., Malhotra, A. and Novack, L. (2009) 'Association of lipid levels during gestation with preeclampsia and gestational diabetes mellitus: a population-based study', *American journal of obstetrics and gynecology*, vol. 201, no. 5, 1-8 [Online]. DOI: 10.1016/j.ajog.2009.05.032.

Woolcott, C. G., Giguère, Y., Weiler, H. A., Spencer, A., Forest, J.-C., Armson, B. A. and Dodds, L. (2016) 'Determinants of vitamin D status in pregnant women and neonates', *Canadian journal of public health = Revue canadienne de sante publique*, vol. 107, 4-5, 410-416 [Online]. DOI: 10.17269/cjph.107.5629.

Wuertz, C., Gilbert, P., Baier, W. and Kunz, C. (2013) 'Cross-sectional study of factors that influence the 25-hydroxyvitamin D status in pregnant women and in cord blood in Germany', *The British journal of nutrition*, vol. 110, no. 10, pp. 1895–1902 [Online]. DOI: 10.1017/S0007114513001438.

Zhang, Y., Wang, X., Liu, Y., Qu, H., Qu, S., Wang, W. and Ren, L. (2012) 'The GC, CYP2R1 and DHCR7 genes are associated with vitamin D levels in northeastern Han Chinese children', *Swiss medical weekly*, vol. 142 [Online]. DOI: 10.4414/smw.2012.13636.

Zittermann, A. (2022) *Vitamin D im Überblick: Wissenschaftlich fundierte Grundlagen und aktuelle Forschungsergebnisse*, Berlin, Heidelberg, Springer Berlin Heidelberg.

Appendix

Strobe Checklist

Table 9: Checklist of STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) for cross-sectional studies (Elm et al., 2008)

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses
Results		
Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
Outcome data	15	Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorised
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		

Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

Code manual

Table 10: Variables of the used data set with their characteristics

Variable name in data set	Variable name in thesis	Level of measurement	Definition	Values
rs12785878DICH	rs12785878	Dichotom nominal		1 = G:G (major allele homozygous) 2 = G:T & T:T (min. one minor allele)
age	age	metric	Mother's age in years	discret metric
mut25ohgroup	Vitamin D deficiency	Dichotom nominal	Categorization of Mother's 25-(OH) vitamin D3 in nmol/L → deficiency: below 50 nmol/L → sufficiency: 50+ nmol/L	1 = deficient 0 = sufficient
skincolor	Skin type	Dichotom nominal	light skin type (Fitzpatrick levels I–III) and dark skin type (Fitzpatrick levels IV and V). (Wille, 2022)	1 = light skin 2 = dark skin
BMIakt	Current BMI	Metric	BMI right before delivery	continuous metric

Variable name in data set	Variable name in the-sis	Level of measurement	Definition	Values
helandgroup	country of origin	nominal	Categorization Country of origin	1= Switzerland/Germany 2 = North America/Northern Europe/ Caucasus/ Central Asia / New Zealand 3 = South Europe/Australia/Latin America/Carib-bean 4 = South- East-Asia and Pacific 5 = Africa and Middle East
season.y	season	ordinal	Season at time of birth	Spring Summer Fall Winter
supvitdtot	vitamin D supplementa-tion	dichotom nominal	Vitamin D supplementation	1 = yes 2 = no

R Packages

Table 11: Overview of R packages used in the statistical analyses of this thesis

Package	Used for..
car	Logistic regression (incl. testing assumptions and quality criteria)
carData	Logistic regression (incl. testing assumptions and quality criteria)
caret	Logistic regression (incl. testing assumptions and quality criteria)
Dpylr	Descriptive analyses, formatting data
effects	Logistic regression (incl. testing assumptions and quality criteria)
ggplot2	Logistic regression (incl. testing assumptions and quality criteria)
grid	Bivariate analyses, logistic regression (incl. testing assumptions and quality criteria)
lattice	Logistic regression (incl. testing assumptions and quality criteria)
lsr	Bivariate analyses
polycor	Bivariate analyses, polyserial correlation
Psych	Descriptive analyses
rcompanion	Bivariate analyses, logistic regression (incl. testing assumptions and quality criteria)
readxl	Loading data set
ResourceSelection	Logistic regression (incl. testing assumptions and quality criteria)
stats	Bivariate analyses, logistic regression (incl. testing assumptions and quality criteria)
tidyverse	Descriptive analyses
vcd	Bivariate analyses, logistic regression (incl. testing assumptions and quality criteria

Bivariate analysis

Table 12: Complete bivariate analysis within all predictors

	rs12785878	Skin type	Current BMI	Country of origin	Season	Supplementation
rs12785878		p-value = 0.02 *V* = 0.23	r_{bs} = -0.18	p-value = 0.001 *V* = 0.39	p-value = 0.52 *V* = 0.12	p-value = 0.19 *V* = 0.13
Skin type			r_{bs} = 0.23	p-value = 3.51e-11 *V* = 0.63	p-value = 0.54 *V* = 0.11	p-value = 0.76 *V* = 0.05
Current BMI				*eta²* = 0.04	p-value = 0.41 *rho* = 0.07	r_{bs} = - 0.03
Country of origin					p-value = 0.09 *V* = 0.22	p-value = 0.82 *V* = 0.10
Season						p-value = 0.80 *V* = 0.09
Supplementation						

white Effect size below the low correlation threshold by Cohen (1992) ***light red*** low correlation ***medium red*** medium correlation
r_{bs} biserial correlation ***V*** Cramer's V ***eta²*** eta quadrat

Overview of all logistic regression models and their quality criteria

Table 13: Overview of all logistic regression models considered and their quality criteria

		Model 1 rs12785878	**Model 2** Model 1 + skin type	**Model 3** Model 2 + country of origin	**Model 4** Model 3 + supplementa-tion	**Model 5 (final)** Model 4 + current BMI	**Model 6** Model 5 + season of birth
NAs		0	0	1	3	3	3
Odds ratios min. 1 variant allele (2) vs. wild type (1)		0.34	0.34	0.40	0.37	0.39	0.42
Confidence Intervalls (95%)		0.14 - 0.78	0.13 - 0.79	0.15 - 1.01	0.14 - 0.94	0.14 -1.02	0.15 - 1.14
AIC		188.37	188.79	188.07	188.44	184.48	189.08
Deviance test (p-values)	*vs. basic model*	0.009747359	0.03	0.02	0.02	0.01	0.02
	vs. model with only rs12785878	NA	0.45	0.12	0.10	0.06	0.11

		Model 1 rs12785878	**Model 2** Model 1 + skin type	**Model 3** Model 2 + country of origin	**Model 4** Model 3 + supplementa-tion	**Model 5 (final)** Model 4 + current BMI	**Model 6** Model 5 + season of birth
Variance clarification	*Hosmer Lemeshow R2*	0.035	0.037	0.083	0.091	0.114	0.118
	Cox and Snell R2	0.047	0.049	0.108	0.119	0.142	0.153
	Nagelkerke R2	0.063	0.066	0.144	0.158	0.189	0.201
Hosmer-Lemeshow test (p-values)		Can not be cal-culated be-cause the num-ber of bins (from 2 up-wards) led to negative df	2.2e-16	2.2e-16	0.86	0.94	0.96

		Model 1 rs12785878	**Model 2** Model 1 + skin type	**Model 3** Model 2 + country of origin	**Model 4** Model 3 + supplementa-tion	**Model 5 (final)** Model 4 + current BMI	**Model 6** Model 5 + season of birth
Correct data within the model	*Correct data in %*	100	100	56.20	40.15	51.11	50.37
	Specificity in %	100	100	42.42	25.76	36.36	34.84
	Sensitivity in %	100	100	69.01	53,52	65.22	56.22

Calculation of the variance inflation factor

Variable	**VIF**	**Df**
rs12785878	1.18	1
Skin type	1.74	1
Country of origin	1.93	4
Supplementation of vitamin D	1.04	1
Current BMI	1.04	1

Declaration of originality

I hereby formally declare that I have written the submitted work independently without any external support, except for the cited literature and other sources mentioned in this thesis. Any passages that have been taken verbatim or in spirit from other works are marked with the source.

Hamburg, 01.10.2024